Improved teleconnection between Arctic sea ice and the North Atlantic Oscillation through stochastic process representation

Kristian Strommen[1], Stephan Juricke[2,3], and Fenwick Cooper[1]

[1]University of Oxford, Oxford, United Kingdom
[2]Mathematics and Logistics, Jacobs University, Bremen, Germany
[3]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Correspondence: Kristian Strommen (kristian.strommen@physics.ox.ac.uk)

Abstract. The extent to which interannual variability in Arctic sea ice influences the midlatitude circulation has been extensively debated. While observational data supports the existence of a teleconnection between November sea ice in the Barents-Kara region and the subsequent winter North Atlantic Oscillation, climate models do not consistently reproduce such a link, with only very weak inter-model consensus. We show, using the EC-Earth3 climate model, that while an ensemble of coupled EC-Earth3 simulations shows no evidence of such a teleconnection, the inclusion of stochastic parameterizations to the ocean and sea ice component results in the emergence of a robust teleconnection comparable in magnitude to that observed. We show that this can be accounted for by an improved ice-ocean-atmosphere coupling due to the stochastic perturbations, which aim to represent the effect of unresolved ice and ocean variability. In particular, the weak inter-model consensus may to a large extent be due to model biases in surface coupling, with stochastic parameterizations being one possible remedy.

1 Introduction

Over the last several decades, Arctic sea ice has been undergoing a precipitous decline (Stroeve and Notz, 2018). Since this loss is unanimously projected to continue as long as greenhouse gas concentrations keep increasing (Notz and Community, 2020), a considerable number of studies have been devoted to understanding what influence this may or may not have on mid-latitude weather and climate (see, e.g., Barnes and Screen (2015) for a good overview). The role of sea ice variability in driving mid-latitude weather has also been extensively examined on interannual timescales, where it has, in particular, been suggested as a key source of predictability in seasonal forecasts of Euro-Atlantic boreal winter (García-Serrano et al., 2015; Dunstone et al., 2016; Kretschmer et al., 2016; Wang et al., 2017). Of central importance on both climate and seasonal timescales is a proposed teleconnection between November sea ice concentration in the Barents-Kara (BKS) region and the December-January-February (DJF) North Atlantic Oscillation (NAO), where a negative sea ice anomaly forces a negative NAO anomaly (Deser et al., 2007; Sun et al., 2015). Because the NAO is the dominant mode of variability in the Euro-Atlantic sector (Hurrell et al., 2003), such a teleconnection would provide a direct pathway for sea ice variability to affect the mid-latitudes.

Both tropospheric and stratospheric mechanisms have been suggested for such a teleconnection. In the tropospheric pathway, localized heat flux anomalies triggered by the exposure of the relatively warm ocean to the cold Arctic atmosphere may produce stationary Rossby waves (Hoskins and Karoly, 1981) which can subsequently grow to a large-scale NAO response (García-

Serrano et al., 2015). The stratospheric pathway posits that these waves may penetrate all the way up to the stratosphere, where they break and weaken the polar vortex: this would be expected to result in a negative NAO response at the surface in late winter (Peings and Magnusdottir, 2014; Kim et al., 2014). The importance of a favorable North Atlantic storm track has been emphasized for both mechanisms (Deser et al., 2007; Strong and Magnusdottir, 2010; Siew et al., 2020).

However, there is currently no consensus in the literature on whether this teleconnection actually exists at all. While studies looking for predictors of the winter NAO in reanalysis data frequently identify a robust BKS-NAO teleconnection as the largest source of predictability (Kretschmer et al., 2016; Wang et al., 2017), this is not straightforwardly reproduced using climate model experiments. Indeed, while models show a weak NAO response to imposed sea ice anomalies when integrated over a sufficiently long time period, the signal appears to be both much smaller and less consistent over shorter time periods, with only a weak inter-model consensus (see, e.g., Screen et al. (2018) for an overview). In addition, long climate simulations of just a single model with fixed forcing exhibit decade-long periods where the correlation between BKS sea ice and the NAO can be both positive, negative or zero (Koenigk and Brodeau, 2017). By considering observational data covering the entire 20th century, Kolstad and Screen (2019) suggested that such decadal variability is also visible in the real world, though such analysis is seriously confounded by the reduced quality of sea ice data prior to 1979 when the satellites came online[1]. All this has led some to suggest that the apparently significant positive correlation seen in the recent observational record may be purely a result of atmospheric internal variability (Koenigk and Brodeau, 2017; Warner et al., 2020; Blackport and Screen, 2021). Some studies have also hypothesised that the correlations arise due to the existence of a common atmospheric driver of both sea ice and the NAO (Peings, 2019; Siew et al., 2021).

The situation is further complicated by the discovery in recent years of a so-called 'signal-to-noise paradox' in seasonal forecasts of the winter NAO (Scaife and Smith, 2018). This paradox, in essence, is the fact that while forecast models initialised in November can now produce DJF NAO predictions that correlate remarkably well with the observed NAO, the signal is extremely weak in the forecast models and high correlations can only be realised by taking the mean across a large ensemble forecast (Eade et al., 2014; Dunstone et al., 2016). In particular, the size of the signal compared to the level of skill (i.e., the magnitude of the correlation) implies that the real world may be significantly more predictable than the models think it is. One possible explanation for this is that models have systematically under-persistent circulation anomalies (Strommen and Palmer, 2019), but another is that models fail to capture real-world teleconnections adequately (Siegert et al., 2016). This raises the possibility that the Arctic-NAO teleconnection is real, and that the weak and inconsistent signals seen in climate model simulations are a manifestation, or possibly even the cause, of the signal-to-noise paradox. While some studies, such as Blackport and Screen (2021), argue that the overall signal is too small to be robust even when using large ensembles of climate simulations, it has been noted in Baker et al. (2018) that not all seasonal forecast models exhibit skillful winter NAO forecasts. It is therefore not obvious, a priori, that all the climate models considered in studies such as Blackport and Screen (2021) represent the relevant processes correctly. This naturally begs the question: what processes might not be represented correctly in state-of-the-art climate models that could inhibit a realistic Arctic-NAO teleconnection?

[1]For example, the widely used HadISST sea ice data is "mostly climatologies before the 1950s" (Chapman, William and National Center for Atmospheric Research Staff (Eds)., 2013), and for Barents-Kara this can be seen to be the case for later years as well by inspection.

Many model errors have their origin in processes that are unresolved, either because they occur below the model grid-scale, or because they are not represented in the model physics. One increasingly widespread approach to address this is the idea of stochastic parameterization schemes (Berner et al., 2017). Stochastic schemes aim to represent the influence of uncertain processes, such as unresolved sub grid-scale physics, using carefully calibrated random noise. The potential for such schemes to radically improve weather forecasts is well known (Palmer et al., 2009; Sanchez et al., 2015; Berner et al., 2017), and a growing body of literature now suggests that such schemes can also have a beneficial impact on climate model simulations as well (see, e.g., Dawson and Palmer (2015); Watson et al. (2017); Christensen et al. (2017); Juricke et al. (2017); Strommen et al. (2019); Vidale et al. (2021) for some examples). Of particular relevance is the suggestion of Strommen et al. (2017) that stochasticity can in some cases make teleconnections more realistic.

In this paper, we will study the impact of including the stochastic parameterization schemes of Juricke et al. (2013); Juricke and Jung (2014) to the sea ice component and of Juricke et al. (2017) to the ocean component of the fully coupled EC-Earth3 climate model. We compare an ensemble of six deterministic (i.e., non-stochastic) simulations (labelled CTRL) spanning 1950-2015, with an equivalent ensemble (labelled OCE) with stochastic sea ice and ocean parameterizations active: see Data and Methods for details of the stochastic schemes and the model. We will show that while the CTRL simulations do not exhibit a systematic relationship between BKS sea ice and the NAO, the OCE simulations appear to systematically recover a significant teleconnection roughly comparable to that seen in the reanalysis product ERA5. Analysis using a linear inverse model (LIM), along with comparisons against simulations with prescribed sea-surface temperatures (SSTs) and sea ice, suggests that the stochastic schemes are primarily acting to improve the coupling between the atmosphere, ocean, and sea ice. The ability of stochastic schemes to profoundly impact atmosphere-ocean coupling has already been observed in, e.g., Christensen et al. (2017). Our results therefore support the hypothesis that inadequate ice-ocean-atmosphere coupling may be a key bias contributing to the weak and inconsistent ice-NAO teleconnection in climate models. A similar hypothesis has also been emphasised in Mori et al. (2019a) and Mori et al. (2019b) in the context of surface level teleconnections, and we add to this by showing that stochastic schemes aimed at improving ice and ocean variability may alleviate such coupling biases.

The structure of the paper is as follows. In Section 2 we describe the model and simulation setup, the stochastic ocean and sea ice schemes and the diagnostic methods used in this study. Section 3 discusses the effects of the stochastic schemes on the climatology of the model, while Section 4 focuses specifically on the impact of the schemes on the ice-NAO teleconnections in EC-Earth3. In Section 5 we test and discuss the extent to which improved teleconnections with the stochastic sea ice and ocean schemes are due to improved ice-ocean-atmosphere coupling. Finally, Section 6 consists of a discussion of the results and some final conclusions.

2 Data and Methods

2.1 The EC-Earth3 model and description of stochastic schemes

The model used for the coupled climate simulations in this study is a version of EC-Earth3. Specifically, we use the EC-Earth3P configuration developed for the HighResMIP protocol, as described in Haarsma et al. (2020). It is very closely related to the

version that was used for the introduction of the probabilistic Earth system model in Strommen et al. (2019). In their study, the focus was on stochasticity in the atmospheric and land surface component of uncoupled atmosphere-only simulations, while in this study the stochasticity is placed in the ocean and sea ice model component as discussed below.

The atmospheric model component of EC-Earth3 consists of a modified version of the Integrated Forecast System (IFS) developed and used at the European Centre for Medium Range Weather Forecasts (ECMWF). It includes the the land surface model Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (H-TESSEL) (Balsamo et al., 2009). The ocean model component is represented by the NEMO model version 3.6 (Madec and the NEMO team, 2016) which includes the LIM3 sea ice model (Vancoppenolle et al., 2012). The atmospheric model is run at a spectral resolution of T255, which corresponds to an approximate grid spacing of 80km at the equator, with 91 vertical layers. This model produces a reasonable looking quasi-biennial oscillation (Strommen and Palmer, 2019), suggesting that the stratosphere is well resolved. NEMO is run at a resolution of around 1° with 75 vertical layers. Note that, as discussed in Haarsma et al. (2020), the original NEMO configuration for EC-Earth3P produced an Atlantic Meridional Overturning Circulation (AMOC) with unrealistically low values: the configuration used in these experiments corresponds to the $p2$ configuration discussed in ibid, and therefore does have a realistic AMOC. We also note that the EC-Earth3P model was tuned to represent a realistic top of the atmosphere energy budget compared to observational estimates for the time period 1990-2010.

The two types of coupled simulations carried out in this study differ only by the use of stochastic parametrizations in the ocean and sea ice model component. The control simulation CTRL runs without any stochastic parametrizations turned on. The stochastic simulation OCE on the other hand has three stochastic ocean schemes (Juricke et al., 2017, 2018) and one stochastic sea ice scheme (Juricke et al., 2013; Juricke and Jung, 2014; Juricke et al., 2014) switched on.

The stochastic ocean schemes are based on perturbations to

1. the classical Gent–McWilliams parametrization for eddy induced advection (Gent and Mcwilliams, 1990) used in coarse resolution, non eddy-resolving ocean simulations (henceforth StoGM);

2. the enhanced vertical diffusion parametrization which is used for unstable stratification (henceforth StoDV);

3. the turbulent kinetic energy (TKE) parametrization through which the amplitude of vertical diffusivity and viscosity are obtained (henceforth StoTKE);

For the deterministic control simulation, all above mentioned parametrization schemes are used in their default, non-stochastic form.

The stochastic ocean schemes have been explained in detail and tested in long ocean-only simulations by Juricke et al. (2017) and have also been tested in coupled seasonal forecasts by Juricke et al. (2018). These studies showed that the StoGM and StoTKE schemes in particular can have a considerable impact on near surface variability in regions of strong horizontal gradients (StoGM) or strong atmosphere-ocean interactions (StoTKE). The StoGM scheme showed the strongest impact on variability in western boundary currents such as the Kuroshio and the Southern Ocean, as it varies the effective amplitude of

eddy induced temperature and salinity advection. The StoTKE scheme on the other hand had a pronounced impact on variability and mean state in the tropics and also in mid-latitudes, as it can affect the response of the mixed layer to atmospheric forcing. The StoDV scheme showed only a very limited response in these previous studies, as its variations only matter in areas of deep convection in the high latitudes and even there only at times of strong and deep convective activity. However, the effect of these schemes has so far not been tested in long (multidecadal) coupled simulations.

For the sea ice, the stochastic scheme implemented is the stochastic sea ice strength parametrization (henceforth StoSIS) by Juricke et al. (2013), which has so far been tested in ocean-only simulations with the Alfred Wegener Institute (AWI) sea ice-ocean model FESOM (Juricke et al., 2013); with NEMO (Brankart et al., 2015); in annual coupled seasonal ensemble simulations (Juricke et al., 2014); and finally in coupled climate simulations (Juricke and Jung, 2014) with the AWI-CM climate model. The scheme perturbs the resistance of sea ice to convergent motion, which can lead to plastic deformation in the sea ice model. This corresponds to the ridging of sea ice. Ridging can create sea ice of thicknesses beyond the thermodynamic equilibrium thickness of around 1-3 m (depending on local conditions and hemispheric differences between the Arctic and Antarctic) which can already be achieved by purely thermodynamic freezing of sea water. Especially for older multiyear ice in the Arctic, dynamic processes driven by convergent/divergent motion due to atmospheric or oceanic currents are dominating contributors when it comes to sea ice thickness distributions (Juricke et al., 2013). The StoSIS scheme simulates the variations and uncertainties in the local ice strength and can lead to either faster or slower ice convergence, where the non-linearity in the process leads to a stronger response for weak ice compared to strong ice (Juricke et al., 2013). Consequentially, ice tends to move faster under stochastic ice strength perturbations until the effect is balanced by thicker sea ice that also acts to strengthen the resistance towards plastic deformation. Furthermore, ridging tends to be stronger with StoSIS, especially along coastlines if the ice is moved towards the coast (Juricke et al., 2013). However, due to the strongly coupled system consisting of sea ice, atmosphere and ocean in the high latitudes, the climatological response both with respect to mean changes in the sea ice as well as surface flux variability varies between uncoupled (large increase in sea ice volume) and coupled (balancing increase of thick ice vs decrease of thinner ice) simulations (Juricke and Jung, 2014). The sensitive response of StoSIS and, consequently, sea ice dynamics to the atmospheric coupling is one of the main foci of this study.

In summary, we will be considering two configurations of EC-Earth3, one being a default control configuration, referred to as CTRL, and one which differs only from the default by including the stochastic schemes described, referred to as OCE.

2.2 Model simulations and observational data

For each of the two configurations CTRL and OCE, described in the previous section, six ensemble members were generated according to the *hist-1950* experimental protocol (Haarsma et al., 2020). Each member is therefore initialised on January 1st 1950, and run with observed anthropogenic forcing until January 1st 2015: these simulations thereby span 65 years. Different ensemble members are created by slightly perturbing the upper air temperatures at 5 randomly chosen grid points: the atmospheric variability of the members are found to be effectively uncorrelated after just a few months. Hereafter, we will use the terms CTRL and OCE to refer interchangeably to both the configuration of EC-Earth3 and the associated ensemble of model simulations.

To estimate the impact of mean state biases, we will also consider simulations with prescribed SSTs and sea-ice. A set of three deterministic ensemble members were generated by initial condition perturbations, as above: each member then simulates the period 1950-2015. This ensemble, and its experimental configuration, will be referred to as AMIP. Note that in accordance with the HighResMIP Protocol (Haarsma et al., 2016), the prescribed forcing uses daily HadISST2 data, as opposed to the more common monthly forcing. This in principle allows the simulations to simulate more sub-seasonal variability.

To place the correlations in context, we also estimate teleconnections in the historical coupled simulations for 31 single-member CMIP6 models (Eyring et al., 2016), detailed in Table B1, and for an additional 39 simulations from HighResMIP (Haarsma et al., 2016) obtained from six models run at different resolutions with multiple ensemble members, as detailed in Table B2. Only data covering 1980-2015 is used for these simulations.

In order to further assess the statistical significance of the teleconnection, we also made use of additional deterministic coupled simulations using the slightly earlier version EC-Earth3.1, as generated for the Climate SPHINX Project (Davini et al., 2017). This data consists of three ensemble members spanning the period 1900-2015 using historical forcings. We found that the teleconnection behaviour of these ensemble members was consistent with that of the CTRL ensemble: to keep the presentation streamlined, these are not included in the analysis shown.

Finally, for observational data, we make use of the ERA5 data set (Hersbach et al., 2020). The OSI450 sea ice data set (Lavergne et al., 2019) is also used in Section 4 to compare the teleconnection region identified in ERA5.

2.3 Methods

The NAO is defined as the leading principal component (PC) of daily, detrended geopotential height anomalies at 500hPa (zg500). The sign of the principal component is imposed to make the corresponding empirical orthogonal function match the standard NAO pattern. This is first carried out using DJF data only, and then extended to November by projecting the November zg500 anomalies onto the NAO pattern obtained from DJF. A daily climatology is then directly fitted to the NAO time series and removed from it: virtually identical results are obtained if a seasonal cycle is instead removed from each gridpoint in the domain prior to carrying out the computation of the PC. This procedure is carried out separately for ERA5 and each of the two ensembles CTRL and OCE; in the case of the model ensembles, all ensemble members are used to define the NAO pattern itself, while climatologies are computed separately for each member to allow for differences in the mean. This deliberate choice of methodology means the NAO patterns will differ slightly between ERA5, CTRL and OCE. Teleconnections typically project onto the dominant modes of variability (Shepherd, 2014), and since these will almost always differ somewhat in models versus observations, allowing for such differences (rather than enforcing identical patterns) is a way to not overly penalise model performance. However, in our case, the NAO patterns are extremely similar in all data sets (pattern correlations of around 0.97) so the results are unlikely to be sensitive to this choice. All available data is used in these computations, so that for ERA5 the period 1980-2015 is used, while for CTRL and OCE the period 1950-2015 is used. We obtained virtually identical results if using only 1980-2015 for each data set.

Time series of daily sea-ice concentration (siconc) anomalies in a given region are computed by averaging over grid points, detrending and removing a directly fitted daily climatology. There is little material difference to the resulting time series if the

detrending, or removal of the seasonal cycle, is done at each gridpoint before averaging instead. As with the NAO, this is done using all available data, before restricting to specific time periods. The region considered here is Barents-Kara (BKS), defined by the box 67N-80N, 10E-75E.

Correlations are computed as Pearson correlation coefficients. In plots showing correlations at multiple grid points, statistical significance at the 95% confidence interval is computed using the standard error. In plots showing changes to the mean state at grid points, significance is estimated with a two-tailed t-test, with no assumption of equal variance. In plots showing changes in the standard deviation at grid points, significance is estimated using the Bartlett test for SSTs. For sea ice, the distributions show larger deviations from normality, so we use the Levene test here. When computing correlations between November sea ice and DJF NAO time series (as in Table 1), a null hypothesis of each time series as independent AR1 processes is assumed, where the lag is 1 year. By fitting such a process to each time series and simulating 10,000 random draws, we obtain confidence intervals for the correlations expected by chance due to interannual autocorrelation. For ERA5 and the individual ensemble members of CTRL and OCE, which have a sample size of 35 (covering the 35 complete DJF seasons between January 1980 and December 2015), a 95% confidence interval is approximately given by ± 0.35. For the concatenated time series of CTRL and OCE, which have a larger sample size of $35 \cdot 6 = 210$, the standard deviation is ≈ 0.066, with a 95% confidence interval of approximately ± 0.13 and a 99% interval of ± 0.16.

When evaluating differences in time series between the CTRL and OCE ensembles (of the NAO, sea ice, or some grid point), the time series of each ensemble member are typically concatenated back to back prior to comparison. Note that in a standard forecasting context, where one is comparing an ensemble of forecasts x_i for $i = 1, \ldots, N$, against a fixed observational time series y, the correlation between the ensemble mean and y equals the correlation between the time series of the concatenated members x_i and the time series obtained by concatenating y with itself N times. Therefore, the concatenated time series can be thought of as a natural extension of the ensemble mean which makes sense even when the 'observed' time series y is no longer fixed (as is the case when comparing CTRL and OCE).

Finally, in this paper we will generally limit our focus to the 35 winters between 1980 and 2015, where observational estimates are particularly trustworthy. Earlier time periods will be considered briefly to assess decadal variability and the role of trends.

3 Impact on the climate mean state

We first show the impact of the stochastic schemes on the long-term mean and variability of the model, noting that the impact of the stochastic ocean schemes in longer coupled simulations with EC-Earth has not previously been documented in the literature.

Figure 1 shows the differences in the mean and standard deviation of sea ice concentration (siconc) for CTRL versus ERA5, as well as the impact on these differences in OCE. Note that this latter impact is visualised as the difference OCE minus CTRL to more clearly highlight the changes. The differences are computed across 35 November months from the period 01-01-1980 to 31-12-2015: note that data is only drawn from the 35 winters spanning a full November-February season. A significant bias

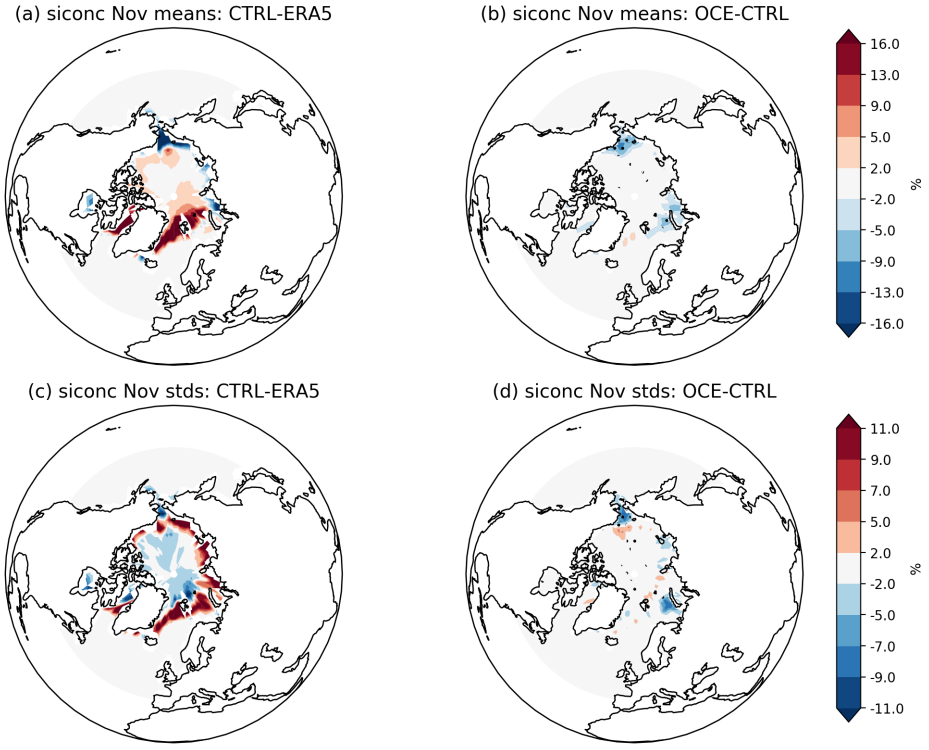

Figure 1. Sea ice concentration (mean and standard deviations) in November. Mean quantities in (a) CTRL-ERA5 and (b) OCE-CTRL. Standard deviations in (c) CTRL-ERA5 and (d) OCE-CTRL. Stippling in (b) and (d) highlights gridpoints where the change is statistically significant ($p<0.05$); in (a) and (c) almost every gridpoint outside the zero contour is significant so stippling is not included for visual ease. The period considered is 1980-2015.

around the Greenland and Barents seas can be seen in the CTRL experiments, with the model producing too much sea ice with excessive interannual variability. The spatial pattern of the bias in the standard deviation, Figure 1(c), implies that this excess variability is due to a tendency for the ice edge to extend out much further in EC-Earth3 compared to ERA5. The main impact of OCE is to reduce the sea ice concentration along the ice edge and mitigate the excessive ice edge protrusion in the Barents sea. This is consistent with the impact of the StoSIS in the coupled AWI-CM model (Juricke and Jung, 2014), suggesting that the changes here are mainly due to the sea ice perturbations promoting stronger ice motion towards the coast of Greenland and the Canadian Arctic Archipelago. This result is therefore consistent with their physical explanation of accelerated ice transport caused by an effective (non-linear) weakening of the sea ice by the random ice strength perturbations. Comparing Figures 1(b) and (d) with (a) and (c) suggests that this is reducing the CTRL bias by around 10% in the Labrador, Barents and Kara seas, but increasing the bias in the Bering sea. All these impacts were found to be qualitatively similar when considering the DJF period instead.

In summary, the OCE simulations appear to exhibit stochastically induced changes to ice transport which change the position, extension and seasonal variability of the ice edge relative to CTRL. Such systematic changes to the variability are further supported by changes to the leading EOFs of sea ice concentration in OCE (not shown).

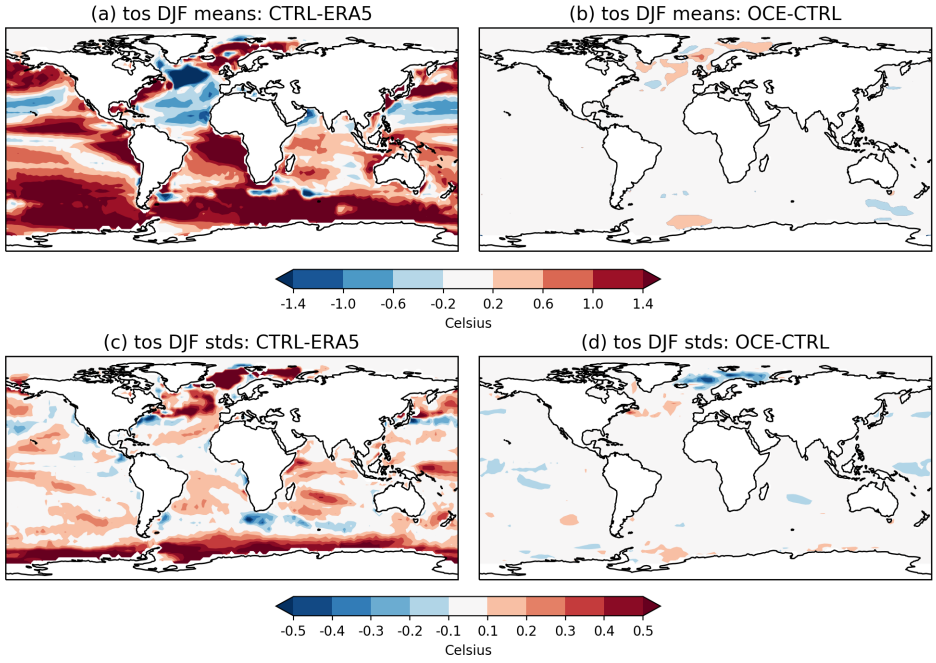

Figure 2. Sea surface temperature (mean and standard deviations) in DJF. Mean quantities in (a) CTRL-ERA5 and (b) OCE-CTRL. Standard deviations in (c) CTRL-ERA5 and (d) OCE-CTRL. In each plot, every gridpoint outside the zero contour is significant ($p<0.05$). The period considered is 1980-2015.

Figure 2 shows analogous differences in the mean and standard deviation of SSTs for CTRL versus ERA5 and OCE. The only places where OCE appears to notably alter the mean state are the North Atlantic and Barents-Kara region. EC-Earth3 exhibits the common model bias of a cold North Atlantic (Wang et al., 2014), known to be associated with biases in the variability and positioning of the Gulf Stream. The inclusion of stochasticity improves (i.e., increases) variability in the Gulf Stream and this is a likely source of the differences in the North Atlantic mean state between CTRL and OCE, though it is clear that these changes do not generally constitute a reduction in the mean SST bias. Because there is a link between North Atlantic SSTs and the North Atlantic jet (Brayshaw et al., 2011; Keeley et al., 2012), we also show the impact on 850hPa zonal winds in Figure B1 of Appendix B. Interestingly, the changes to the jet more clearly constitute a bias reduction, which may suggest that these changes are not purely SST driven but rather related to the altered teleconnection discussed in the next section.

Examination of other variables supports an overall conclusion that the stochastic sea ice and ocean schemes are having only a modest impact on the climate mean state, and do not, for example, alter the net surface global energy (not shown). We also found no evidence of significant changes to direct ocean-atmosphere coupling, as measured using the methods of Frankignoul

	BKS-NAO correlation
ERA5	**0.39**
CTRL	-0.04
$CTRL_1$	-0.14
$CTRL_2$	0.02
$CTRL_3$	-0.20
$CTRL_4$	-0.01
$CTRL_5$	-0.13
$CTRL_6$	0.10
OCE	**0.26**
OCE_1	0.13
OCE_2	0.17
OCE_3	**0.36**
OCE_4	0.08
OCE_5	**0.54**
OCE_6	0.26

Table 1. Correlations between November sea ice anomalies in the November Barents-Kara sea ice concentration and the DJF NAO mean, over the period 1980-2015, for each data set. Subscript labels $CTRL_n$ and OCE_n (N=35) denote ensemble members 1-6, while the entry for CTRL and OCE (N=210) uses the concatenated time series. Entries that are significant ($p<0.05$) are marked in bold. Significance is measured against a null hypothesis of uncorrelated, random AR1 processes.

et al. (1998) and Von Storch (2000) (not shown); as will be seen later, changes to surface coupling do manifest in OCE but only in the presence of sea ice. It is likely that this limited impact on the climate mean state is due to the fact that the 1° ocean does not permit eddies and is strongly damped by viscosity. Consequently, the stochastic ocean schemes, which are primarily attempting to perturb the variability of turbulent processes, cannot achieve much without making the perturbations extreme in magnitude (Juricke et al., 2017). This viewpoint is supported by the fact that in separate experiments using a 0.25° ocean, the impact of the same schemes on SST variability was much greater (not shown). The main exception to this is in the Arctic and regions such as the North Atlantic, where, firstly, the sea ice perturbations play an active role, and, secondly, where the interplay between Gulf stream variability, atmosphere-ocean coupling and vertical mixing in the ocean is large enough for the ocean perturbations to have a bigger impact.

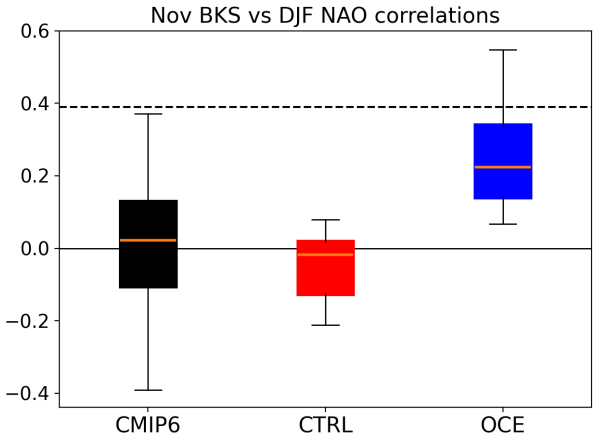

Figure 3. Boxplots of correlations between sea ice anomalies in the November Barents-Kara sea ice and the DJF NAO mean, over the period 1980-2015, for CMIP6 (black), CTRL (red), and OCE (blue). In each case, the box extends from the first to the third quartile of the data, with a line at the median. The whiskers extend from the box by 1.5 times the inter-quartile range. The stipled black line marks the correlation attained in ERA5.

4 Impact on the Arctic-NAO Teleconnection

4.1 Changes to correlations and assessment of significance

Table 1 summarises the magnitude of the teleconnection for each data set by showing the correlations between the November BKS ice concentration and the winter NAO over the period 1980-2015. Correlations significant to the 95% threshold are shown in bold. It can be seen that ERA5 exhibits a significant positive correlation of ≈ 0.39, consistent with that reported by many other studies using a variety of techniques and observational datasets. None of the six individual CTRL members, nor the concatenated data set, show significant correlations, exhibiting both small positive and negative values. By contrast, every ensemble member of OCE shows a positive correlation. While only two of these attain statistical significance, after concatenation, OCE has a correlation of 0.26, which sits comfortably outside the 4 sigma range of the AR1 null hypothesis. The exact likelihood of this occurring under the null hypothesis was estimated to be less than 1 in 10,000.

To visualise these changes, and place them in context, we show, in Figure 3, box-and-whisker plots of these correlations along with those from the CMIP6/HighResMIP simulations (referred to hereafter as just CMIP6 for simplicity). It is evident that the CMIP6 ensemble, consisting here of 70 model simulations covering the period 1980-2015, shows no consistent signal, with the distribution consistent with a null hypothesis that the correlations are just random draws from a mean zero normal distribution with a standard deviation of 0.17. The same conclusion holds if the HighResMIP simulations are excluded. The exact mean of 0.018 is slightly positive, consistent with earlier studies suggesting that models on the whole have a weak positive correlation, but the consensus is clearly weak, with around half the models showing negative correlations. We note that discrepancies with other studies showing a slightly stronger multi-model consensus may be due to the fact that we consider

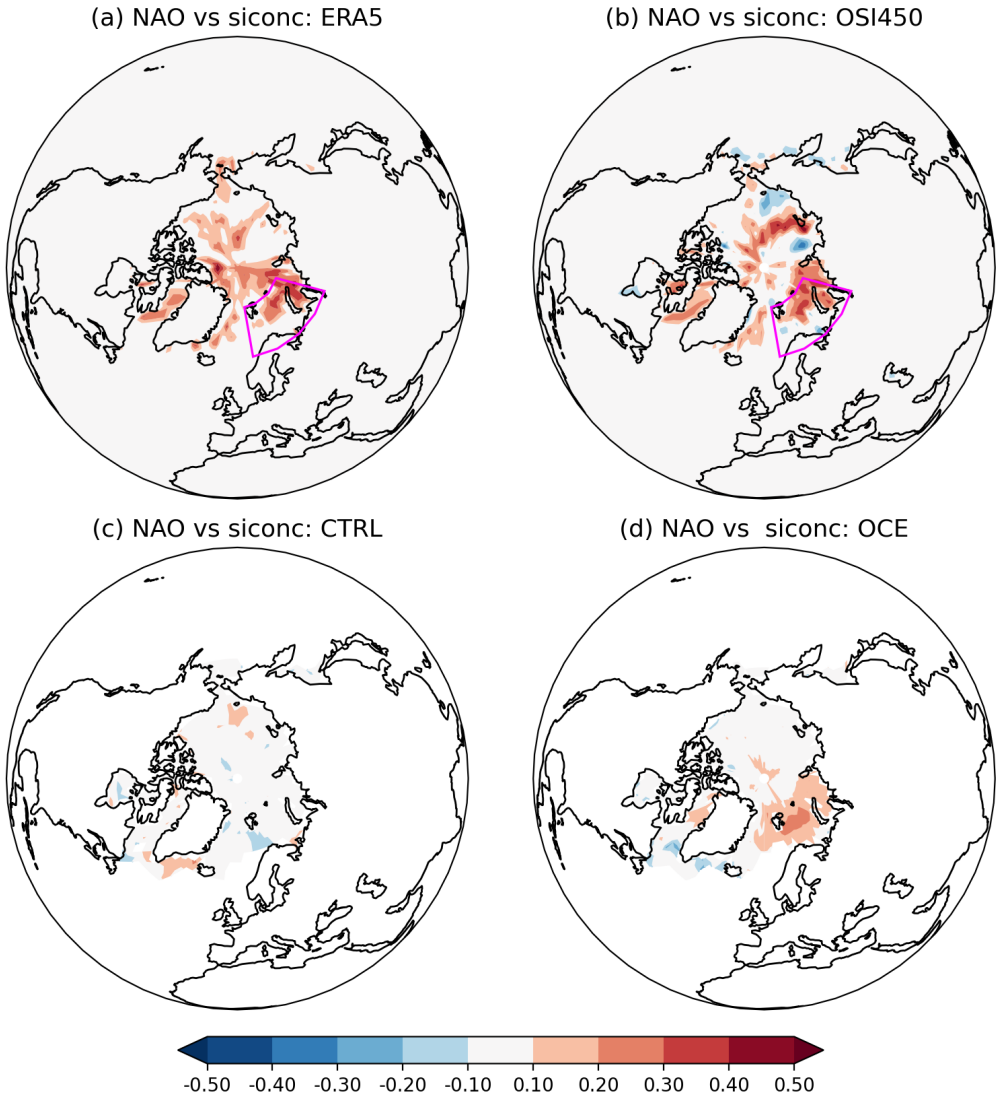

Figure 4. Correlations between the detrended November siconc anomalies at each gridpoint and the DJF NAO time series, over the period 1980-2015. In (a) ERA5, (b) OSI450, (c) CTRL and (d) OCE. The ensemble members of the CTRL and OCE experiments have all been concatenated together. For the CTRL and OCE plots (c) and (d), every gridpoint outside the zero contour is statistically significant ($p < 0.05$). For ERA5 and OSI450, only gridpoints in the Barents and Kara seas are significant. In (a) and (b) the Barents-Kara region has been marked by a purple box.

here a wider range of models than many studies, as well as the fact that different studies use differing experimental protocols that are not necessarily directly comparable (e.g., simulations using fixed anthropogenic forcing vs simulations using actual historical forcings). The choice of which months to consider also differs between many studies. Nevertheless, it is visually clear

that the deterministic CTRL correlations are consistent with random draws from the CMIP6 distribution and do not appear as unusually weak in the context we are interested in.

On the other hand, the OCE correlations are all positive with a mean value in the upper tercile of the CMIP6 distribution. The correlation of 0.54, obtained by one OCE ensemble member, is not obtained by any of the CMIP6 simulations we analysed. Under a null hypothesis that the correlations being considered are independent random draws around a Gaussian fit to the CMIP6 distribution, the chance of drawing 6 positive correlations is $0.5^6 \approx 0.016$, i.e. 1.6%. The chance of drawing 6 independent correlations that are not just positive but comparably large was estimated as being an order of magnitude less likely. In reality, the assumption that the 6 OCE ensemble members are strictly independent is likely false, since, for example, we cannot rule out the possibility that being initialised in the same ocean state has predisposed the ensemble towards a particular pattern of decadal variability. However, since the CTRL members also start from the same state, the influence of the initialisation on our results appears to be weak.

Figure 4 shows the spatial pattern of the correlations in each data set, by plotting correlations of the DJF NAO time series against November sea ice concentration at every gridpoint in the Arctic. As expected, CTRL shows only a few spurious looking correlations, with no significant correlations in the BKS region. For OCE, the largest correlations are found in the Barents sea, but significant correlations are also found in the Kara sea. The fact that the correlations in the Kara sea in OCE are weaker than those in ERA5 may be related to the sea ice biases shown in Figure 1.

In summary, we conclude that the correlations obtained in the OCE ensemble are significantly larger than both those in CMIP6 and CTRL. In particular, the circulation in OCE appears to exhibit a robust ice-NAO teleconnection.

4.2 Decadal variability, trends and climate change

While the focus of this paper is on the period 1980-2015, we make some remarks on decadal variability in the teleconnection and the possible influence of trends, which may be of interest for future work. The reader is warned that, in the absence of more detailed analysis, this discussion is necessarily somewhat speculative.

We find that all the simulations appear to be initialised into a period of weak, but broadly positive, BKS-NAO correlations, with no clear difference between CTRL and OCE (not shown). The two ensembles diverge from each other around 1975, so that by 1980 the correlations of OCE are robustly separated from those of CTRL, as discussed in the preceding section. It is likely that some of this variability is a result of random internal variability, but consideration of the spatial pattern of the correlations suggests a somewhat more nuanced picture. This can be seen by plotting the equivalent of Figure 4 for earlier periods: see Figure B2 for an example of this for the period 1950-1985. This shows that in this earlier period, the weak, positive BKS-NAO correlations in CTRL and OCE are driven by a weak signal from the Barents sea. More striking however is the presence of a much larger signal emanating from the Greenland and Labrador seas in both ensembles. Further analysis (not shown) suggests that the region of significant gridpoint correlations shrinks over time: in CTRL, the region basically shrinks to nothing by 1975, while in OCE it shrinks to Barents and Kara (Figure 4).

It seems likely that this behaviour is at least in part due to the sea ice loss that occurs in this time period. Both models lose a considerable amount of sea ice in the Greenland, Barents and Kara seas, with the OCE model losing somewhat less in

Barents and Greenland and somewhat more in Kara. The loss of sea ice in the Greenland sea in particular is associated with a permanent retreat of the sea ice edge, resulting in a collapse of interannual sea ice variability, and hence correlations, in this region.

Several studies have made the point that the strongest heatflux anomalies are associated with the variability of the ice edge. The strong temperature gradients between the relatively warm ocean and the cold atmosphere aloft mean that variations in the extent of the ice edge are associated with heat flux anomalies reaching as high as $500 W m^{-2}$: see, e.g., Koenigk et al. (2009) for a comprehensive overview. The fact that it is these heat flux anomalies that are hypothesised to influence the circulation, as opposed to the sea ice per se, suggests a plausible physical justification for the teleconnection being associated with the region of maximum November ice edge variability. For observational data in the modern period, this corresponds to the Barents and Kara seas, particularly the Barents sea (Deser et al., 2000; Vinje, 2001; Koenigk et al., 2009). However, given the significant changes to the ice edge over the 20th century, it may be that other regions, such as the Greenland sea, played an important role in the past.

To summarise, we find that the trend of a shrinking ice edge appears to influence the location of significant gridpoint correlations, and that much of the decadal variability in the BKS-NAO correlation prior to 1980 can be accounted for by the fact that other regions, particularly the Greenland sea, appear to be more important in earlier periods. It is noteworthy that even in these earlier periods, the OCE ensemble has significantly higher correlations than CTRL in the relevant regions (see Figure B2), suggesting that the stochastic perturbations are consistently enhancing ice-NAO teleconnections across the full period 1950-2015.

A final point of interest is that both CTRL and OCE show entirely comparable negative trends of the NAO over 1950-2015 (not shown). It is therefore possible that the stochastic schemes are having a negligible impact on the long term climate change response to sea ice loss in EC-Earth3, since differing responses might be expected to manifest as differing NAO trends. It would be interesting to run future projections of the CTRL and OCE configurations to assess this further. As far as future changes to interannual variability are concerned, the discussion on decadal variability above is consistent with the trivial observation that the teleconnection would be expected to slowly vanish as the ice edge retreats.

4.3 Tropospheric vs stratospheric pathways

Figure 5 shows the vertical structure of the teleconnection, measured by regressing November BKS ice against zonally averaged DJF zonal winds, at different latitudes and pressure levels. Two things are apparent from this. Firstly, the teleconnection in OCE, viewed here as a northward shift of the jet in response to a positive BKS anomaly, extends all the way up to the stratosphere. This means that both tropospheric and stratospheric pathways may be present in the teleconnection seen in OCE. We do not attempt to disentangle the two at present, but remind the reader that the stochastic perturbations in OCE are strictly limited to the ocean and ice at the surface, so it is unlikely that changes to the teleconnection are a result of direct changes to the stratospheric dynamics. Secondly, the teleconnection in OCE is still weaker than that of ERA5: while the correlations are comparable in magnitude, the actual regression coefficients are larger in ERA5, and are statistically significant across the entire vertical structure. The same picture holds if regressing against zonally averaged geopotential height anomalies instead.

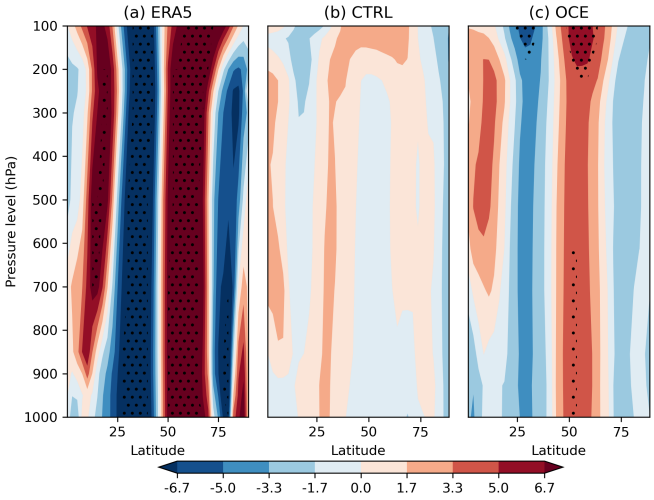

Figure 5. Regression coefficients of November BKS sea ice against zonally averaged DJF zonal winds at various latitudes (degrees north) and pressure levels (hPa): in (a) ERA5, (b) CTRL and (c) OCE. The period considered is 1980-2015. Stipling highlights gridpoints where the associated correlation coefficient is statistically significant ($p < 0.05$).

5 A Quantitative Explanation via Ice-Ocean-Atmosphere Coupling

5.1 Mean state changes alone do not suffice

In Section 3 it was shown that the stochastic schemes have notably changed the mean state of North Atlantic SSTs and Arctic sea ice. An obvious first hypothesis is therefore that the improved teleconnection is the result of these changes. To test this, we examined an equivalent set of three deterministic ensemble members using prescribed observational SST/ice boundary forcing from the HadISST data set, which we refer to as the AMIP ensemble. The equivalent version of Figure 4 for AMIP (Figure B3), obtained by correlating the model's NAO with the prescribed sea ice at each gridpoint, shows no indication of any ice-NAO link at any gridpoint. The correlations of the November BKS sea ice anomalies and the DJF NAO, for each of the three ensemble members, are 0.18, 0.01 and -0.05: the correlation using the concatenated time series is 0.06. None of these values are significantly different from 0, and considering other 35-year periods between 1950 and 2015 does not change this. We therefore conclude that the AMIP ensemble does not exhibit an ice-NAO teleconnection.

Note that the tendency for AMIP-style models to have weaker Arctic-midlatitude teleconnections than their coupled counterparts has been noted in both a multi-model context (Blackport and Screen, 2021), and for EC-Earth3 in particular (Caian et al., 2018). Consequently, having a perfect ice/ocean mean state does not necessarily produce a realistic teleconnection.

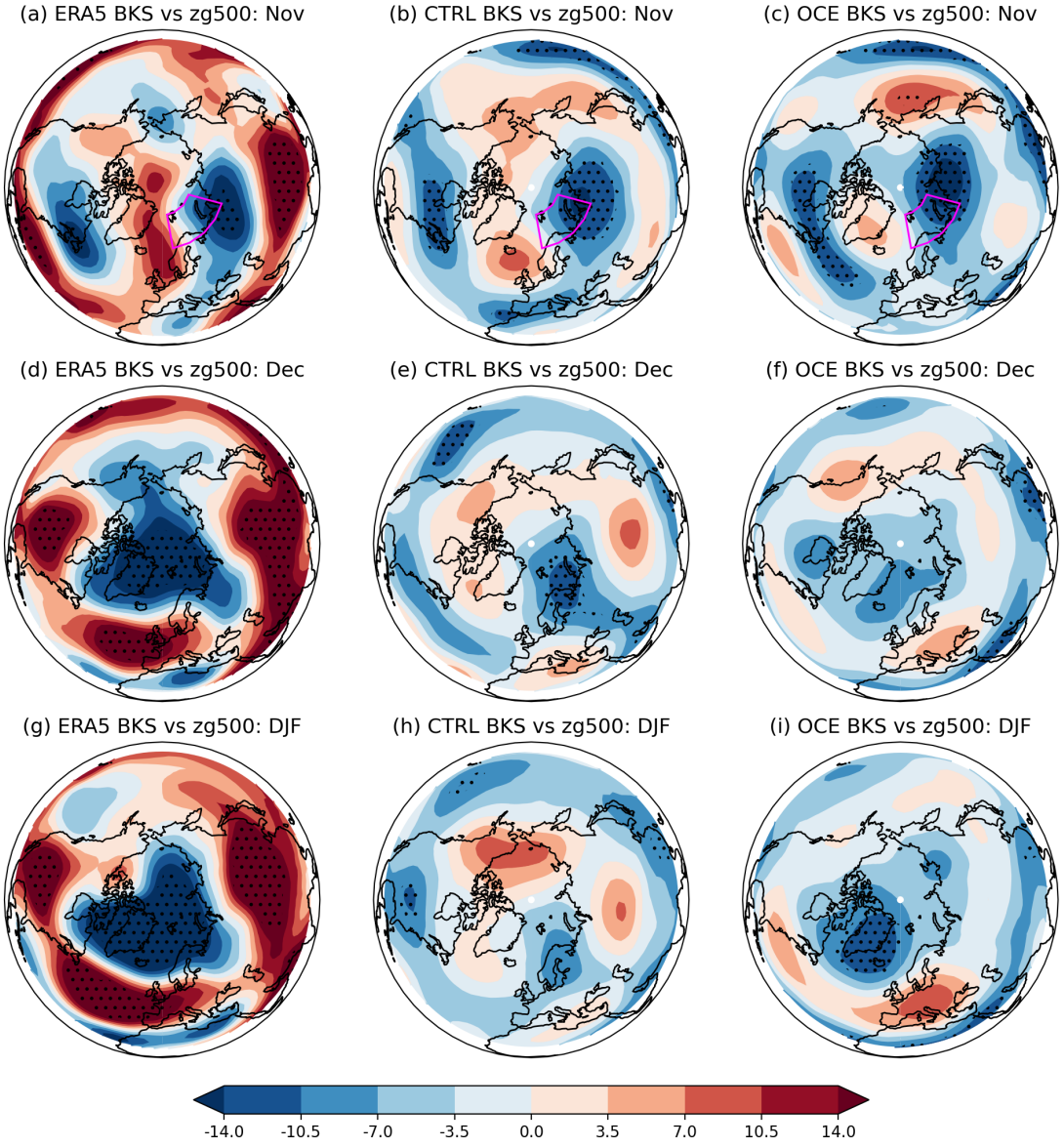

Figure 6. Regression coefficients of November BKS sea ice anomalies against monthly zg500 gridpoint anomalies in ERA5 for (a) November, (d) December and (g) DJF. The same for CTRL (resp. OCE) in (b), (e) and (h) (resp. (c), (f) and (i)). The period considered is 1980-2015. Stipling highlights gridpoints where the associated correlation coefficient is statistically significant ($p < 0.05$).

5.2 Coupled Ice-NAO dynamics and the LIM model

In the previous section we saw that prescribing observed ocean and sea ice forcing does not result in the deterministic model exhibiting a significant teleconnection. If internal atmospheric variability is playing a key role in generating the teleconnection,

then prescribing the sea ice may act to obfuscate this (Blackport and Screen, 2021). However, another possibility is that an important role is being played by the dynamic coupling of the atmosphere and the ocean/ice. The potential importance of coupling to simulate the circulation or surface response associated with Arctic sea ice anomalies was already emphasised in Strong et al. (2009) and Strong and Magnusdottir (2011), as well as more recently in Deser et al. (2016), Mori et al. (2019a) and Mori et al. (2019b).

To gain additional insight into this, we first examine the temporal evolution of the teleconnection. This is done in Figure 6, which shows the regression coefficients of November BKS sea ice anomalies against monthly zg500 anomalies in November, December and DJF, for ERA5, CTRL and OCE.

The instantaneous November regression coefficients, Figure 6(a), (b) and (c), are all fairly similar in all three data sets, with a tripole pattern formed by a low over the Kara sea, a high near Greenland and a low near the eastern coast of North America. The pattern projects weakly onto the negative NAO in the Euro-Atlantic sector but deviates from the canonical NAO pattern over Greenland and the Arctic region by comparison with Figure 6(g), which shows the full DJF NAO response in ERA5. This November tripole pattern will be capturing the combined effects of the atmospheric forcing on the ice as well as the forcing of the ice on the atmosphere, but appears to be most consistent with atmospheric forcing in the form a northerly flow into the Barents sea[2], promoting enhanced sea ice (Koenigk et al., 2009). In the following months, both ERA5 and OCE see this tripole evolve into a larger low centred on Greenland, with a broad high across the North Atlantic and Europe; in DJF this pattern clearly corresponds to a positive NAO for both. In CTRL, on the other hand, the same evolution never takes place, and the seasonal evolution simply consists of a slow dissipation of the initial anomaly. Similar plots showing the evolution of heatflux and temperature anomalies (Figures B4 and B5) corroborate this story, with a similar initial anomaly that evolves relatively realistically over time in OCE but simply peters out in CTRL.

The ability of OCE to evolve the same initial response better than CTRL might be plausibly attributed to changes in persistence, either of the NAO or the sea ice. However, since the atmospheric component is identical in OCE and CTRL, there is no obvious mechanism for how the persistence of the NAO might change, and the lack of a teleconnection when prescribing the sea ice rules out the persistence timescales of the sea ice alone from explaining the change. Figure 7 confirms that the daily autocorrelation of the NAO has not changed notably between OCE and CTRL. Both ensembles have much higher sea ice persistence than ERA5, but again, the difference between OCE and CTRL is small by comparison. By contrast, Figure 8 shows a marked difference between OCE and CTRL in the lag correlation. Most notably, the OCE model appears to be considerably more realistic when the sea ice leads the NAO, with CTRL showing virtually no impact of sea ice on the NAO for lags of up to 50 days ahead. When the NAO leads the ice, CTRL appears more realistic for small lags, but the difference has largely vanished by lags of a month or more.

Note that the opposite signs of the lag correlation based on whether the ice or the NAO is leading has been reported previously in both model data (Magnusdottir et al., 2004) and observational data (Strong et al., 2009), and suggests a natural physical interpretation of the coupling (ibid). A positive sea ice anomaly in the BKS region (i.e., an extension of the sea ice edge) leads to a reduced local heatflux into the atmosphere, which, via some combination of Rossby wave forcing, changes to

[2]We thank two anonymous reviewers for pointing this out.

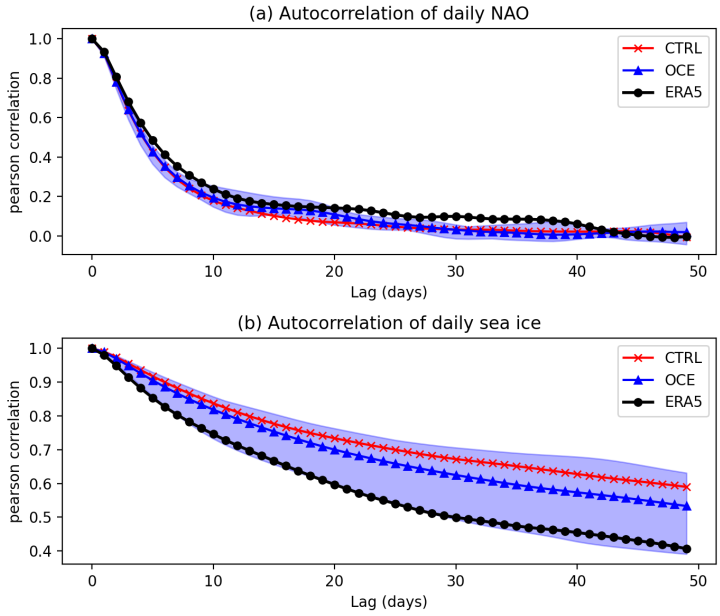

Figure 7. Autocorrelation of (a) the daily NAO and (b) the daily BKS sea ice for ERA5 (black dotted), CTRL (red crosses), and OCE (blue triangles). Data is restricted to days in November through February, 1980-2015. Shading indicates the ensemble spread of OCE: the CTRL spread (not shown) is similar in extent.

the meridional temperature gradient and stratospheric pathways, force the positive phase of the NAO. This corresponds to a northward shift of the eddy-driven jet (Woollings et al., 2010), which would lead to anomalous wind stress along the ice edge and a consequent reduction of the initial positive sea ice anomaly. The more northerly jet may also lead to shifts in the distribution of sea ice more broadly which are potentially important to support a realistic evolution of the initial atmospheric response.

In light of our results so far, it is natural to ask if the changes to short timescale coupling between the ice and the NAO can account for the changes in the seasonal timescale teleconnection. To test this, we model the ice-NAO system using the following pair of coupled ordinary differential equations:

$$\frac{d}{dt}\text{NAO} = a \cdot \text{NAO} + b \cdot \text{ICE} + \xi_{\text{NAO}}, \tag{1}$$

$$\frac{d}{dt}\text{ICE} = c \cdot \text{NAO} + d \cdot \text{ICE} + \xi_{\text{ICE}}. \tag{2}$$

Here NAO is just the daily NAO index, and ICE is the daily sea ice anomaly in the BKS region. The coefficients a and d are capturing the presence of autocorrelation, while the coefficients b and c capture the presence of coupling; the ξ terms in both equations represent the residual forcing on both quantities, and are assumed to be random Gaussian processes with no temporal autocorrelation and a mean of 0. This model, hereafter referred to as the LIM, has been used extensively in the literature to capture coupled variables in climate data, such as atmosphere-ocean coupling, and the coefficients and noise terms

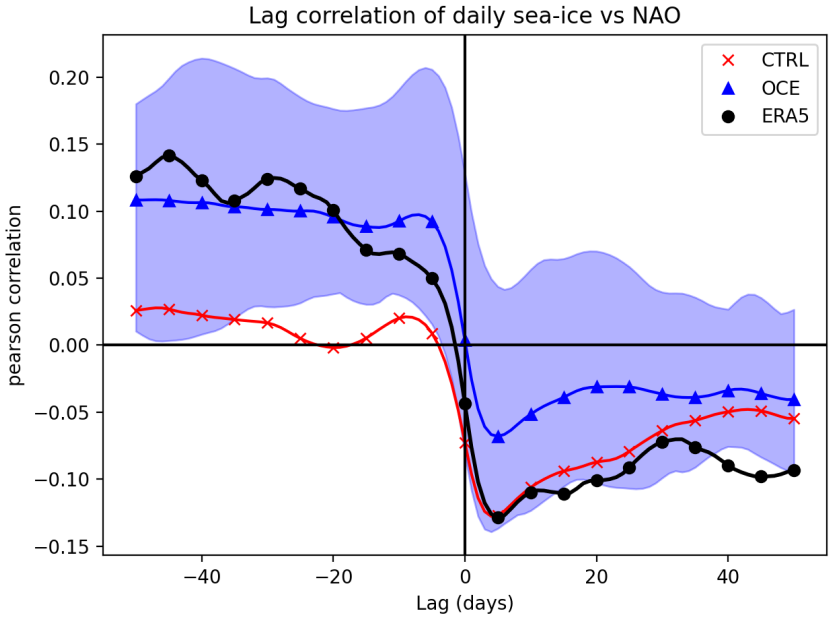

Figure 8. Lagged correlations of the daily NAO against the daily BKS sea ice for ERA5 (black dotted), CTRL (red crosses), and OCE (blue triangles). Negative lags correspond to sea ice forcing the NAO, and vice versa for positive lags. Data is restricted to days in November through February, 1980-2015. Shading indicates the ensemble spread of OCE: the CTRL spread (not shown) is similar in extent. Note that only every 5th point has been marked by a symbol for visual clarity.

can be estimated using Linear Inverse Modeling (see, e.g. Penland and Sardeshmukh (1995); Penland and Magorian (1993); Alexander et al. (2008); Hawkins and Sutton (2009); Newman et al. (2009) for some examples). A brief summary of how to do this is included in Appendix A: the reader may consult Penland and Sardeshmukh (1995) for more details. We are not aware of earlier examples in the literature applying the LIM framework to ice-atmosphere coupling, though the approach of Strong et al. (2009) and Strong and Magnusdottir (2011) is closely related.

5.3 Validation of the LIM hypothesis

Figure 9 shows the estimated LIM coefficients for ERA5 (black), CTRL (red) and OCE (blue), where daily data spanning November through February (NDJF) of 1980-2015 has been used. Fitting the LIM requires a choice of a lag, and we used the simplest possible choice of 1 day, which essentially amounts to fitting the full seasonal ice-NAO dynamics to the 1-day dynamics. Studies such as Deser et al. (2007) suggest that it may take up to a few weeks for sea ice anomalies to project onto the NAO, so it may in principle be better to use a longer lag. However, within the range of lags for which the LIM hypothesis is a good fit, the model parameters are independent of the specific choice of lag (Penland and Magorian, 1993). When carrying out the fitting to CTRL and OCE, all ensemble members were used, under the assumption that the extent of coupling should in theory be identical across different ensemble members. Because the estimated coefficients are inevitably influenced by the

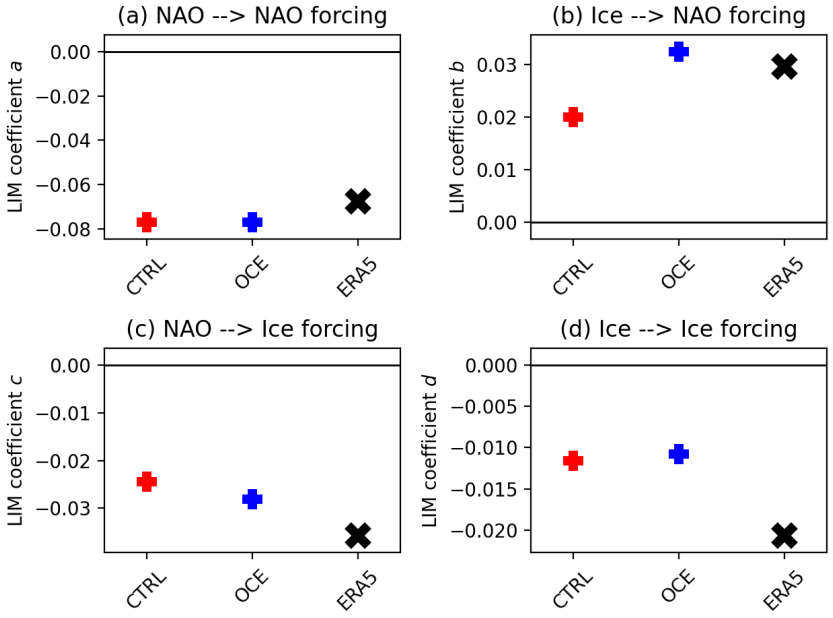

Figure 9. Estimated coefficients of the LIM model, as described by equations (1) and (2). In (a): the coefficient a, describing the forcing of the NAO on itself one day later; (b) the coefficient b, describing the forcing of sea ice on the NAO one day later; (c) the coefficient c, describing the forcing of the NAO on sea ice one day later; (d) the coefficient d, describing the forcing of sea ice on itself one day later. CTRL (resp. OCE) data is marked in red (resp. blue). ERA5 is shown with a black cross. Daily data covering November-February, 1980-2015, is used.

chaotic internal variability, using all ensemble members (rather than fitting to each member separately) also gives more robust estimates.

Once these coefficients and the November initial conditions for each data set are obtained, it is straight forward to generate random reconstructions of the daily NDJF data using the LIM, by feeding in the ice and NAO anomalies on the 1st of November for each year as initial conditions and integrating the system (see Appendix A for details). Examples of LIM-reconstructed DJF NAO time series can be seen in Figure B6, which also illustrates that the reconstructed time series have, on average, the same variance as the true time series. These reconstrutions can then be used to validate the LIM hypothesis for ERA5, CTRL and OCE, by testing if the ice-NAO correlations of each data set are indistinguishable from the ice-NAO correlations of randomly drawn LIM reconstructions. To test this, we generated, for each data set, 1000 random reconstructions of the corresponding LIM, and for each of these correlated the LIM November ice with the LIM DJF NAO. For ERA5, this produced a mean correlation of 0.27 with a 95% confidence interval of -0.05 to 0.56; for OCE, a mean correlation of 0.36 and a confidence interval of 0.06 to 0.62; and for CTRL, a mean correlation of 0.26 and a confidence interval of -0.08 to 0.55. Three things can be inferred from these distributions. Firstly, by comparison with Table 1, the teleconnection correlations in ERA5 and those from the OCE ensemble are consistent with the distribution obtained from their respective LIM, suggesting that the

behaviour of ERA5 and OCE is consistent with the LIM hypothesis. Secondly, the correlations in the actual CTRL ensemble are *not* consistent with the LIM distribution, being clustered down at the lower end of the distribution with 3 of the 6 members falling outside the 95% confidence interval (cf. Table 1). In other words, the behaviour of CTRL is not consistent with the LIM hypothesis. Thirdly, the variation across the LIM coefficients (Figure 9) is largely immaterial for generating significant positive correlations. In particular, the reason the CTRL model fails to have a significant ice-NAO teleconnection cannot be accounted for by, e.g., the smaller b coefficient compared to ERA5 and OCE. Rather, it is that the LIM hypothesis as a whole fails in the case of CTRL.

By more systematically varying the LIM coefficients (not shown), we found that significant positive correlations are generated in the LIM primarily from the non-zero forcing of the ice on the NAO and the long persistence of sea ice anomalies. The forcing of the NAO on the ice has a mostly negligible effect on the magnitude of the correlation. The exact values of the parameters, including the forcing from the NAO, are on the other hand crucial for obtaining reconstructions with the correct variance. If the NAO forcing term is suppressed then the LIM DJF NAO is considerably larger in magnitude than the true DJF NAO of the given data set. In other words, the sea ice forcing being continuously damped by the NAO is what allows the LIM reconstructions to have a variance closely matching that observed, but because correlations are insensitive to magnitudes this does not impact the ice-NAO correlations generated by the LIM.

Finally, we note that no meaningful difference was found between the noise terms of the LIM for CTRL and OCE.

5.4 Why has the coupling improved in the stochastic ensemble?

The conclusion of the previous section was that the LIM hypothesis is reasonable for both ERA5 and OCE, but fails in the case of CTRL. This suggests that the relevant ice-ocean-atmosphere coupling is not simulated correctly in CTRL, and is improved by the inclusion of stochasticity. Why is this the case? The LIM hypothesis is, in essence, that the full winter ice/NAO dynamics can be inferred from the initial conditions on November 1st and how the ice and NAO are linked on daily time scales. There are three obvious ways this might fail for CTRL: 1) the initial conditions are unrealistic in CTRL; 2) the daily time-scale link itself is unrealistic in CTRL; 3) the dynamical response to the initial conditions, as it evolves over the winter season according to the daily time-scale link, is being systematically disrupted somehow in CTRL.

Concerning the first possibility, we find no meaningful difference between the initial ice and NAO conditions of CTRL vs OCE: see Figure B7. While Figure 1 shows that there are some localised differences in the variability across the BKS region, the total variance of the November 1st ice across the region as a whole is almost identical for both ensembles. There is no significant difference between the NAO variances of the two ensembles either. It does not therefore seem likely that differences in the initial conditions are playing an important role.

Concerning the second possibility, since the ice-NAO link is mediated via heatfluxes, it is natural to consider if the heatflux response in CTRL is somehow deficient compared to ERA5 and OCE. As a first test, we find no evidence that the BKS-averaged heatflux response to a sea ice anomaly on November 1st is significantly different in the CTRL or OCE ensemble means, including the vertical extent of the response. This is already apparent from Figures 6, B4 and B5, which show that the mean November signal looks roughly similar in CTRL and OCE. However, more subtle differences between the two ensembles

are apparent in Figure 10, which shows lagged correlations between daily heatfluxes and daily BKS sea ice. While the ensemble mean correlations are similar for CTRL and OCE at all lags, the CTRL ensemble shows a much wider ensemble spread than OCE. This is especially notable for negative lags, i.e., when heatfluxes lead the ice. For lags of 5 days or less, ERA5 shows a clear positive correlation: an increase in upwards heatflux[3] removes heat from the surface, leading to an increase in sea ice. In OCE, all ensemble members show positive correlations for these lags, but for CTRL the response is almost equally likely to have the wrong sign. This suggests that, counter-intuitively, the inclusion of stochastic perturbations causes the sea ice response to be more tightly constrained. In fact, this increased stability of the ice/heatflux relationship is apparent at all lags in Figure 10, pointing towards a generally improved surface coupling in OCE that might be expected to improve the ice-NAO teleconnection.

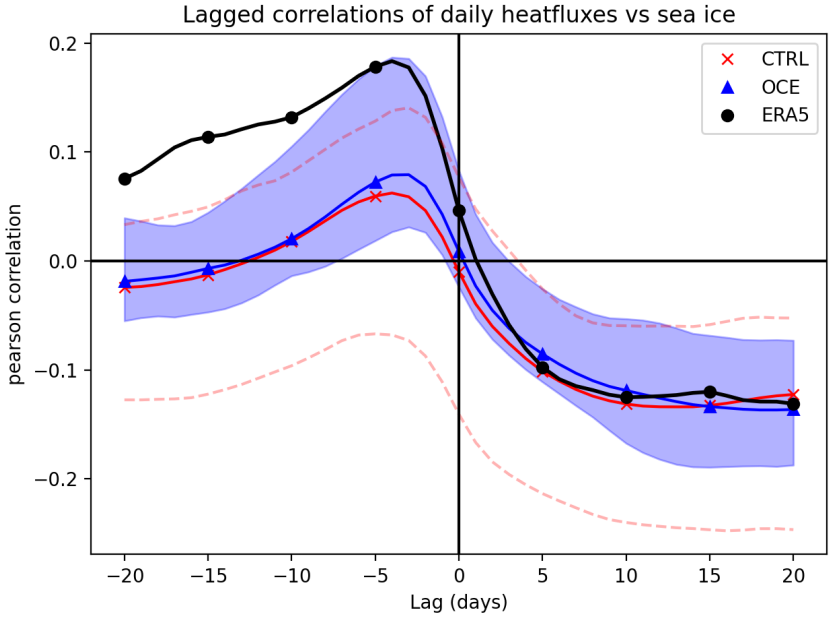

Figure 10. Lagged correlations of the daily BKS heatfluxes against daily BKS sea ice for ERA5, black dotted (resp. CTRL, red crosses, and OCE, blue triangles). Negative lags correspond to heat fluxes forcing the ice, and vice versa for positive lags. Note that positive heatflux values correspond to an upwards flux. Data is restricted to days in November through February, 1980-2015. Blue shading indicates the ensemble spread of OCE; the CTRL spread is indicated by red stipled lines. Note that only every 5th point has been marked by a symbol for visual clarity.

Finally, concerning the third possibility, we suggest two ways in which the evolution of the initial pressure anomaly might be systematically disrupted in CTRL. Firstly, a northward shift of the jet (i.e., a positive NAO) will act to adjust the entire sea ice edge, not just that of Barents and Kara (Koenigk et al., 2009). The SSTs, crucial for generating heatflux anomalies and likely reinforcing the anomaly through ocean-atmosphere coupling (Deser et al., 2016), will also be adjusted by wind stress across the North Atlantic as a whole. These non-local responses may be crucial for the correct evolution of the initial pressure

[3]We remind the reader of the sign convention for fluxes: positive values denote an upwards flux.

anomaly: such adjustments would not be captured by the AMIP simulations, potentially also explaining why these fail to exhibit a teleconnection. As discussed in the preceeding paragraph, Figure 10 suggests that such remote adjustments might be better simulated in OCE. Figure B4 shows a difference in the heatflux response around the Labrador and Greenland seas in December between ERA5/OCE and CTRL, and we speculate that adjustments in these regions may be important. Secondly, it may be that some other unrealistic atmospheric forcing occurring later in the winter is destructively interfering with the ice-NAO link in CTRL. There is some indication that the CTRL ensemble exhibits an unrealistically strong teleconnection between ENSO and the NAO (see Figure B8) which is alleviated in OCE: because the sign of this ENSO link is the opposite of the ice-NAO link, this may be a source of such destructive interference. The fact that deterministic models can have overly dominant ENSO teleconnections that are alleviated with the inclusion of stochasticity was also noted in Strommen et al. (2017).

Other more complex ice-ocean-atmosphere interactions, such as those discussed in Caian et al. (2018), might also be done poorly in CTRL, and it would be interesting to apply their methods here. In addition, it cannot be ruled out that the mean position of the storm track needs to be well aligned with the ice edge in order to foster the growth of the initial anomaly (Deser et al., 2007), implying that the mean state changes between CTRL, OCE and AMIP may be important as well. Note that the AMIP simulations, while having perfect SSTs and ice, still have biases in their simulated storm track which could prevent a robust teleconnection from emerging.

It is clear that more work is required to understand the mechanisms at play, which we hope to address in the future. Nevertheless, our analysis suggests that the improved teleconnection in OCE may be a combination of improved ice-ocean-atmosphere coupling and mean state changes.

6 Discussion and Conclusions

We briefly summarise the results of our analysis.

- The inclusion of stochasticity to the ocean and sea ice components of EC-Earth3, as evaluated with the OCE ensemble experiments, leads to the emergence of a statistically significant teleconnection between November Barents-Kara sea ice and the DJF NAO, comparable to, but somewhat weaker than, that observed in ERA5. No such teleconnection is present in the deterministic EC-Earth3 (the CTRL ensemble; Table 1 and Figure 3). The teleconnection signal in OCE extends all the way up to the stratosphere (Figure 5).
- Comparison with the CMIP6 and CTRL ensemble suggests that the odds of generating six random OCE ensemble members which, by chance, all show a consistent ice-NAO teleconnection, is extremely low (Figure 3). This implies that the teleconnection is likely a real feature of the circulation in OCE.
- An AMIP-style ensemble of deterministic simulations with prescribed sea ice and SSTs is still not able to manifest an ice-NAO teleconnection, suggesting that the relevant model biases are related to coupling at the surface (Figure B3).
- Analysis using lag correlations and a simple LIM model suggests that the teleconnections in ERA5 and OCE can be accounted for by the hypothesis that the direct, local coupling between the sea ice and the NAO on daily timescales

completely captures the seasonal dynamics, but that this hypothesis fails in the case of CTRL (Figure 8 and Section 5.3). This suggests that the ice-ocean-atmosphere coupling of EC-Earth3 has been improved by the inclusion of the stochastic schemes.

- Two broad hypotheses are put forward for the improved ice-NAO coupling in OCE: 1) a more realistic relationship between heatfluxes and sea ice (Figure 10); and 2) a systematic disruption of the BKS-induced pressure anomaly in CTRL by some other unrealistic atmospheric variability, such as an overly dominant ENSO (Figure B8). Improvements to the mean ice state (Figure 1) and zonal winds (Figure B1) may also be playing a role.

These results have important implications for the study of Arctic-midlatitude teleconnections. The potential importance of realistic sea ice-ocean-atmosphere coupling in generating an ice-NAO teleconnection has previously been highlighted by Strong et al. (2009). The possible role of inadequate surface coupling in models has more recently been highlighted by Mori et al. (2019a) in the context of surface level teleconnections, who have also emphasised the role of poorly simulated sea ice variability in models (Mori et al., 2019b). Our work adds further evidence to these views by explicitly demonstrating how model development aimed at improving ice and ocean variability, here in the form of stochastic process representation, resulted in the emergence of a robust teleconnection in a model which did not previously exhibit one. Our analysis supports the hypothesis that improvements to ice-ocean-atmosphere coupling are playing an important role in this result, and furthermore makes it clear that one cannot expect, a priori, that a given climate model does have realistic coupling. In particular, the inconsistency across different models (Blackport and Screen, 2021), and within long simulations of a single model (Koenigk and Brodeau, 2017), appears consistent with a hypothesis that most models fail to simulate a teleconnection due to inadequate coupling. It is also possible that the signal-to-noise paradox in seasonal forecasting is partially due to such biases.

While the most direct interpretation of our analysis seems to be that stochasticity has altered the surface coupling, it is plausible that mean state changes are also playing a role. This is due to the potential sensitivity of the sea ice signal to a favorable storm track, which implies that getting the right signal may depend on the combined sea ice, ocean, and atmospheric mean state. As such, the lack of a teleconnection in the AMIP experiments does not conclusively rule out that the mean state changes in OCE (Section 3) are important. The analysis of lagged correlations and the LIM model lend support to the role of coupling, but it might be that the ice edge of OCE is more optimally aligned with the storm track compared to in CTRL, and that the changes seen in Figures 8 and 9 are simply reflecting this. Untangling this would require more careful analysis of targeted experiments which fall outside the scope of this paper.

If the stochastic schemes are, in fact, improving the coupling between sea ice, ocean, and the atmosphere, what is the precise mechanism which makes the perturbations have this impact? Here too, more careful analysis would be required which goes beyond the scope of this paper. However, a simple conceptual hypothesis might be the following. The evolution of the deterministic sea ice, including how it responds to atmospheric forcing, may be biased in various ways. Years where the sea ice forces the NAO may depend sensitively not just on the initial sea ice anomaly, but also on wider adjustments to Arctic sea ice and North Atlantic SSTs following the initial atmospheric response. The deterministic model may simply be doing these adjustments incorrectly a lot of the time (as hinted at by Figure 10), with the role of stochasticity being to force the model to

not always do the same thing. We note that this idea of stochasticity acting as 'damage mitigation' against biased deterministic models is ubiquitous in weather forecasting (Palmer et al., 2009; Berner et al., 2017). That the stochastic sea ice scheme can lead to very different mean responses for Arctic sea ice due to the coupling response with the atmosphere has been shown by Juricke and Jung (2014) with dedicated full and one-way coupling experiments. We consider it likely that the stochastic sea ice scheme is the dominant factor for changes to the sea ice-NAO teleconnection seen here, but note that the study of Juricke et al. (2018) has shown that the stochastic ocean schemes can also lead to improved skill over North America in seasonal forecasts, a teleconnection effect they relate to improved ocean-atmosphere coupling. It may also be that interactions between the ice and ocean schemes are essential. Further ensemble experiments would be required in order to test the impact of each scheme separately.

Our analysis has some important limitations. For example, we cannot exclude the possibility that the experiments we considered are biased in some way, e.g. due to the initial ocean state predisposing OCE towards certain patterns of decadal variability associated with higher ice-NAO correlations. Another clear limitation is that the impact of these stochastic schemes on Arctic teleconnections has only been tested using a single climate model, and may not produce similar effects in other models. Indeed, it is well known that the same stochastic scheme can have different impacts on different models (Strommen et al., 2019). There may also be important non-linear effects such as those discussed in Caian et al. (2018). Finally, we made no attempt to separate the tropospheric and stratospheric pathways. On the other hand, we note that the stochastic schemes were introduced precisely to improve variability and coupling (Juricke et al., 2013; Juricke and Jung, 2014; Juricke et al., 2014, 2017), so there is good reason a priori to expect such changes to manifest in experimental data.

Our results, which suggest that the inclusion of a stochastic component in the sea ice and ocean can alleviate model biases, both in coupling, the mean state and variability, adds to an increasingly large body of work showing that stochasticity can be beneficial in models across all timescales. It is especially noteworthy, that although changes to the mean and variability of the model due to stochasticity in this study may, in some regions, be rather moderate, other important physical mechanisms in the climate system, such as teleconnections, might be better represented with the stochastic schemes. While the use of stochasticity in the atmospheric component is becoming more widespread, its inclusion in other components of the model is still novel. The potential benefits of representing uncertainty in all major components of a climate model was first raised in Palmer (2012), and our paper adds further weight to this view.

Data availability. Data for the first three ensemble members of CTRL and OCE are publicly available on Zenodo, via the DOI https://doi.org/10.5281/zeno ERA5 data is publicly available via the Copernicus Data Store.

Appendix A: Linear Inverse Modelling

A system of coupled ordinary differential equations, defined by equations (1) and (2), is used to describe coupled ice-NAO interactions on daily timescales. The parameters are fitted using standard methods of linear inverse modelling. Detailed de-

scriptions of how to do this in full generality can be found in Penland and Magorian (1993) and Penland and Sardeshmukh (1995), but to aid the reader we briefly outline the computational steps. Let $B = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be the coefficient matrix, and let x be the vector made up of the daily time series of the NAO and ICE. To estimate B in terms of daily timescale coupling, we computed the lag-0 covariance matrix C_0 and the lag-1 covariance matrix C_1 of x. The matrix B was then estimated as the matrix logarithm $B = log(C_1 C_0^{-1})$. The noise terms can then be estimated by computing the eigenvectors and values of the 'noise matrix' $Q = -1 \cdot (BC_0 + C_0 B^T)$.

Once the parameters have been fitted in this way for a given data set, a reconstruction of a daily DJF NAO time series can then be created, for a given year, by initializing the coupled system with the NAO and ICE anomalies of November 1st of that year and integrating the defining equations forward in time. A single November through February is therefore just a 120 day integration from the initial conditions.

Appendix B

We include some Figures and Tables left out of the main text.

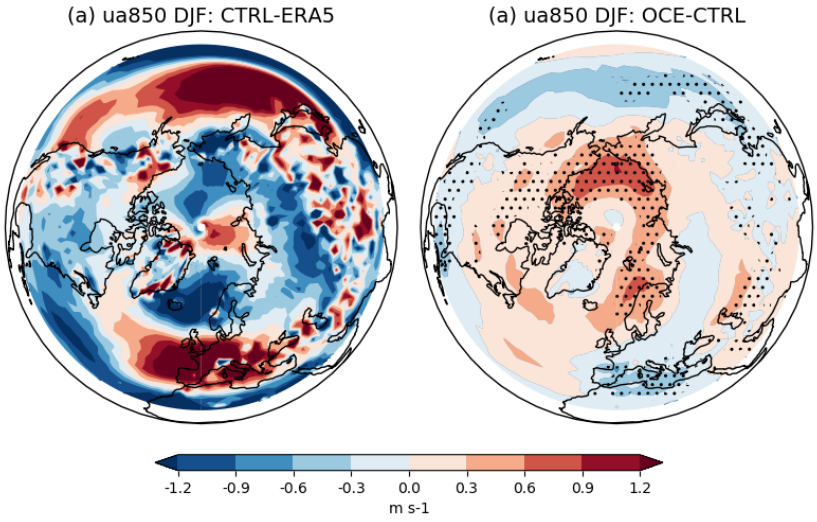

Figure B1. Mean zonal winds at 850hPa (ua850) over the DJF season. In (a) CTRL minus ERA5, (b) OCE minus CTRL. Stipling highlights gridpoints where the change is statistically significant ($p < 0.05$); in (a) most points outside the zero contour are significant so stipling is not included for visual ease. The period covered is 1980-2015.

Model name	Ensemble member
ACCESS-CM2	r1i1p1f1
AWI-ESM-1-1-LR	r1i1p1f1
BCC-CSM2	r1i1p1f1
BCC-ESM1	r1i1p1f1
CanESM5	r1i1p1f1
CESM2-FV2	r1i1p1f1
CESM2	r1i1p1f1
CESM2-WACCM-FV2	r1i1p1f1
CESM2-WACCM	r1i1p1f1
CNRM-CM6-1	r1i1p1f2
CNRM-CM6-1-HR	r1i1p1f2
CNRM-ESM2	r1i1p1f2
EC-Earth3	r1i1p1f1
FGOALS-f3	r1i1p1f1
FGOALS-g3	r1i1p1f1
GFDL-CM4	r1i1p1f1
GISS-E2-1-G	r1i1p1f1
HadGEM3-GC31-LL	r1i1p1f3
HadGEM3-GC31-MM	r1i1p1f3
INM-CM4-8	r1i1p1f1
INM-CM5-0	r1i1p1f1
IPSL-CM6A-LR	r1i1p1f1
MIROC6	r1i1p1f1
MPI-ESM1-2-HAM	r1i1p1f1
MPI-ESM1-2-HR	r1i1p1f1
MPI-ESM1-2-LR	r1i1p1f1
MRI-ESM2-0	r1i1p1f1
NorESM2-LM	r1i1p1f1
NorESM2-MM	r1i1p1f1
TaiESM1	r1i1p1f1
UKESM1-0-LL	r1i1p1f2

Table B1. CMIP6 models used in this paper.

Model name	Ensemble member
AWI-CM-1-1-LR	r1i1p1f002
AWI-CM-1-1-HR	r1i1p1f002
CMCC-CM2-HR4	r1i1p1f1
CMCC-CM2-VHR4	r1i1p1f2
CNRM-CM6-1	r[1,2]i1p1f2
CNRM-CM6-1-HR	r1i1p1f1
EC-Earth3P	r[1,2,3]i1p2f1
EC-Earth3P-HR	r[1,2,3]i1p2f1
ECMWF-IFS-LR	r[1,2,3,4,5,6,7,8]i1p1f1
ECMWF-IFS-MR	r[1,2,3]i1p1f1
ECMWF-IFS-HR	r[1,2,3,4,5,6]i1p1f1
HadGEM-GC31-LL	r1i[1,2,3,4,5,6,7]p1f1
HadGEM-GC31-MM	r1i1p1f1
HadGEM-GC31-HM	r1i[1,2,3]p1f1
HadGEM-GC31-HH	r1i1p1f1
MPI-ESM1-2-HR	r1i1p1f1
MPI-ESM1-2-XR	r1i1p1f1

Table B2. HighResMIP models used in this paper.

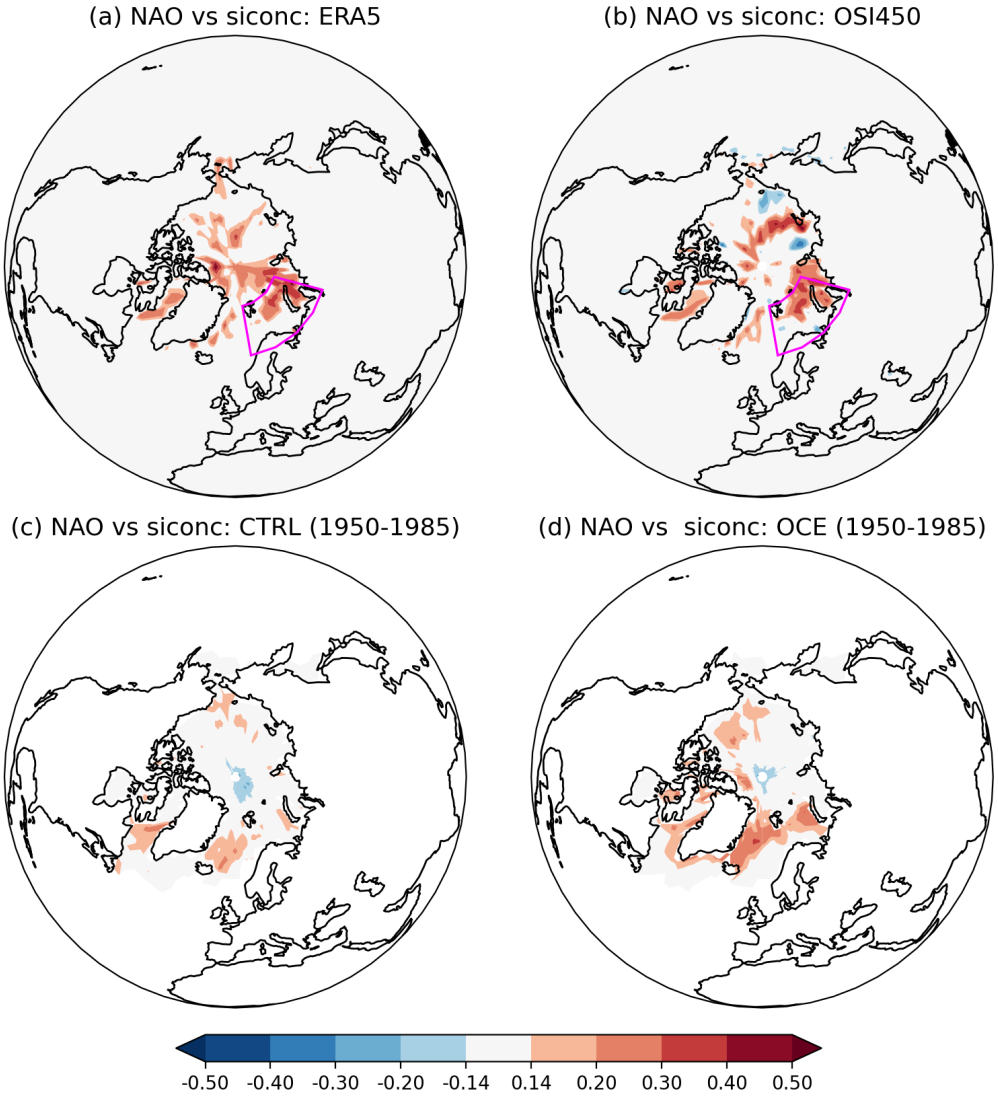

Figure B2. Correlations between the detrended November siconc anomalies at each gridpoint and the DJF NAO time series. In (a) ERA5 (1980-2015), (b) OSI450 (1980-2015), (c) CTRL (1950-1985) and (d) OCE (1950-1985). The ensemble members of the CTRL and OCE experiments have all been concatenated together. For the CTRL and OCE plots (c) and (d), every gridpoint outside the zero contour is statistically significant ($p < 0.05$). For ERA5 and OSI450, only gridpoints in the Barents and Kara seas are significant. In (a) and (b) the Barents-Kara region has been marked by a purple box.

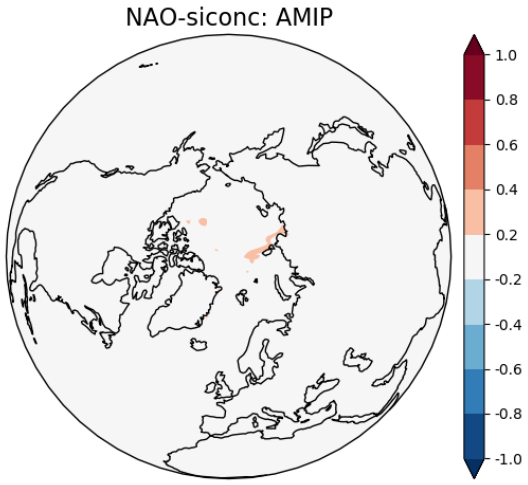

Figure B3. Correlations between the detrended November siconc anomalies at each gridpoint and the DJF NAO time series, using the AMIP ensemble. All three members have been concatenated. The period covered is 1980-2015.

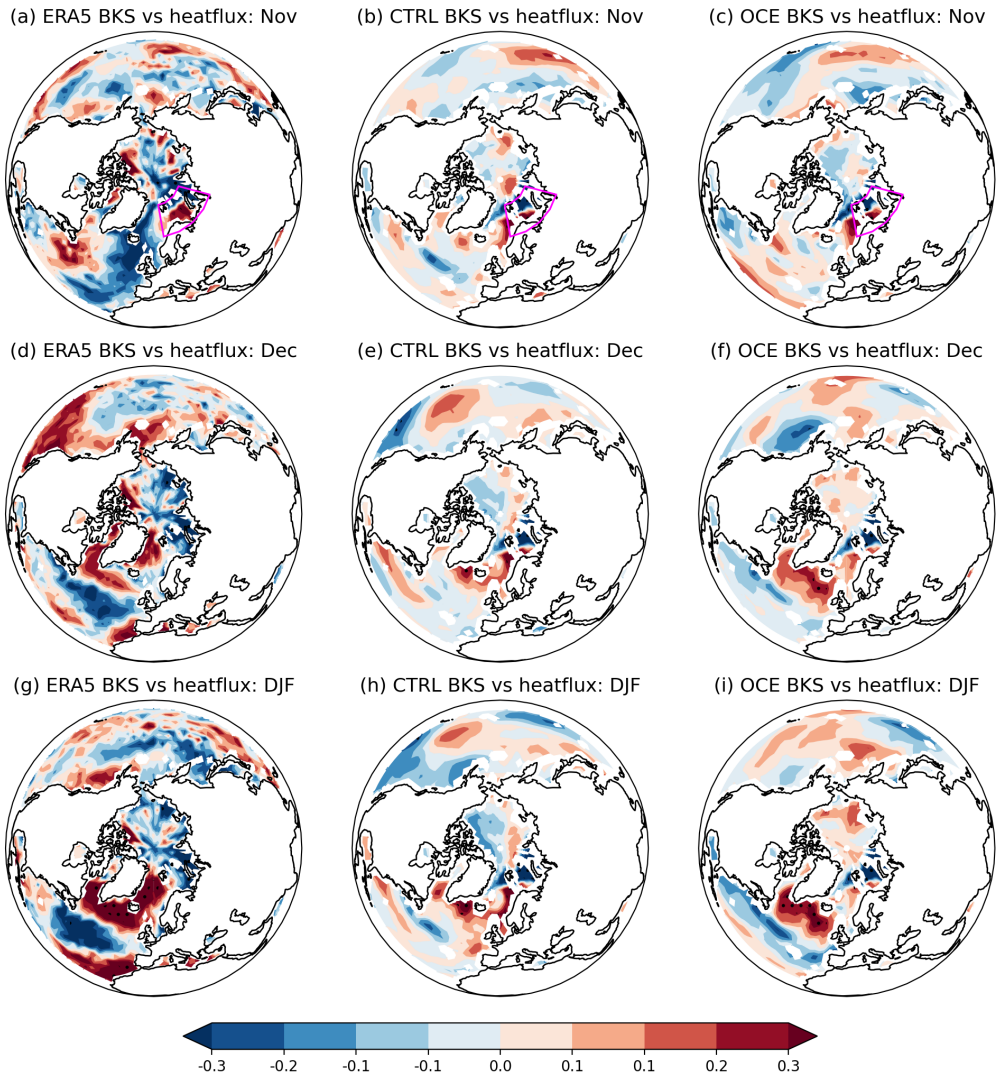

Figure B4. Regression coefficients of November BKS sea ice anomalies against monthly heatflux gridpoint anomalies in ERA5 for (a) November, (d) December and (g) DJF. The same for CTRL (resp. OCE) in (b), (e) and (h) (resp. (c), (f) and (i)). The period considered is 1980-2015. Stipling highlights gridpoints where the associated correlation coefficient is statistically significant ($p < 0.05$). The Barents-Kara region is shown with a purple box.

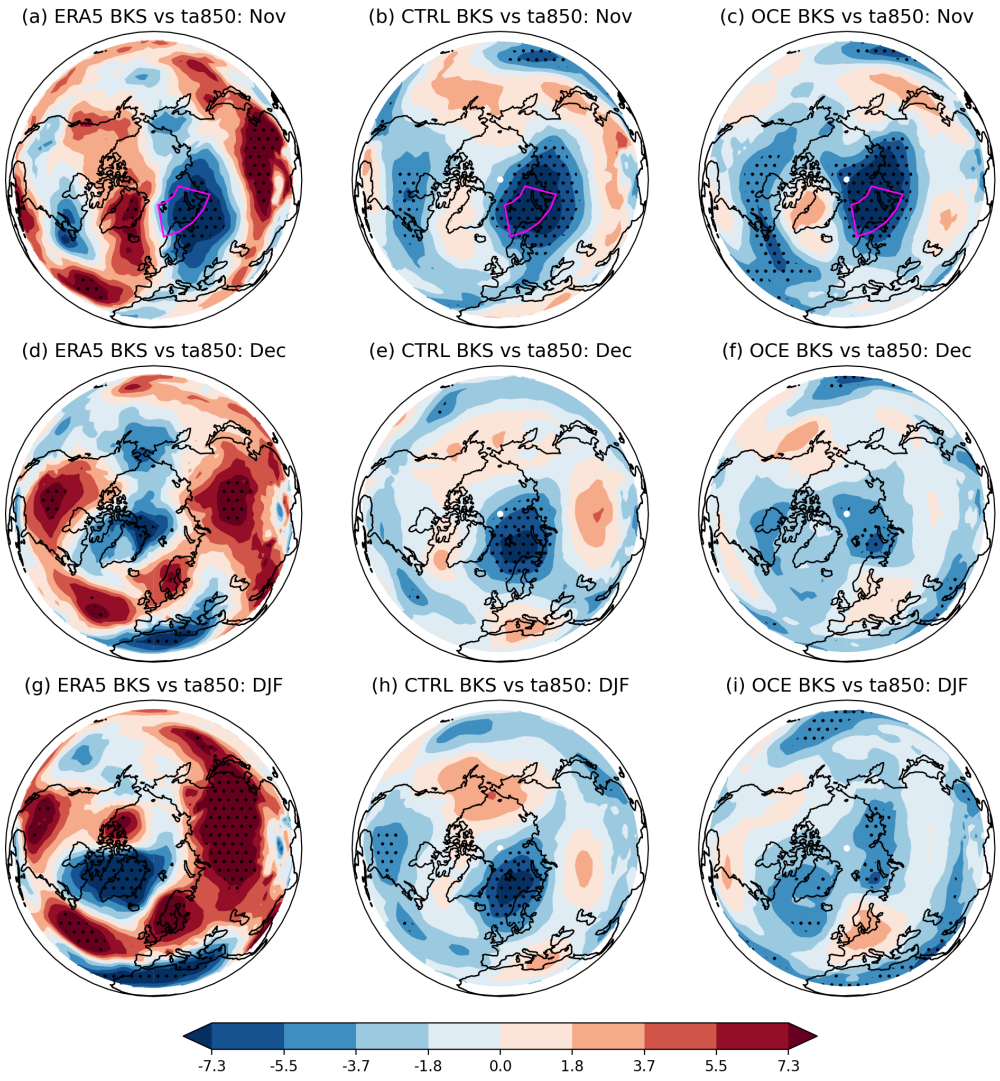

Figure B5. Regression coefficients of November BKS sea ice anomalies against monthly 850hPa temperature (ta850) anomalies at each gridpoint in ERA5 for (a) November, (d) December and (g) DJF. The same for CTRL (resp. OCE) in (b), (e) and (h) (resp. (c), (f) and (i)). The period considered is 1980-2015. Stipling highlights gridpoints where the associated correlation coefficient is statistically significant ($p < 0.05$). The Barents-Kara region is shown with a purple box.

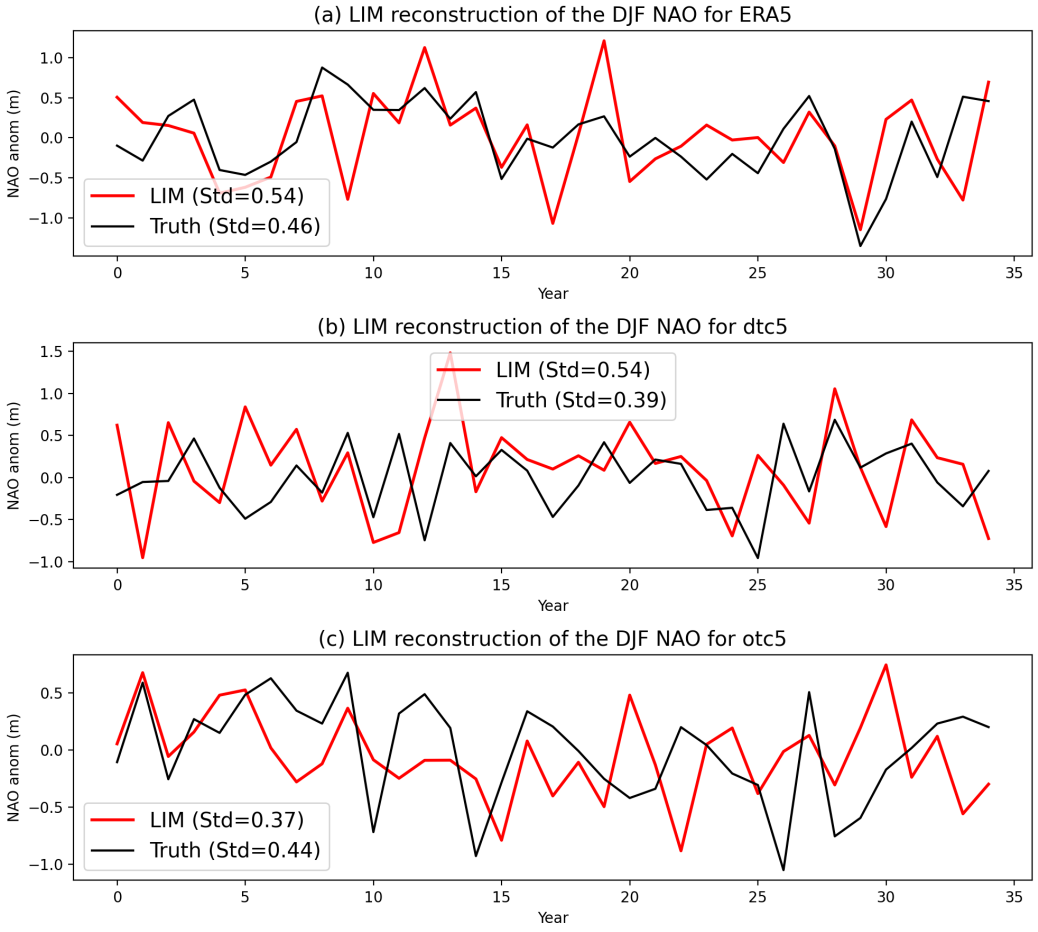

Figure B6. Randomly generated reconstructions of the DJF NAO time series using the LIM initialised each 1st of November for the years 1980 to 2015. For (a) ERA5, (b) the 5th ensemble member of CTRL and (c) the 5th member of OCE. In each plot, the LIM reconstruction is in red and the true time series in black. The standard deviations of each time series has been noted in the legends. Reconstructions of other ensemble members of CTRL and OCE are qualitatively similar.

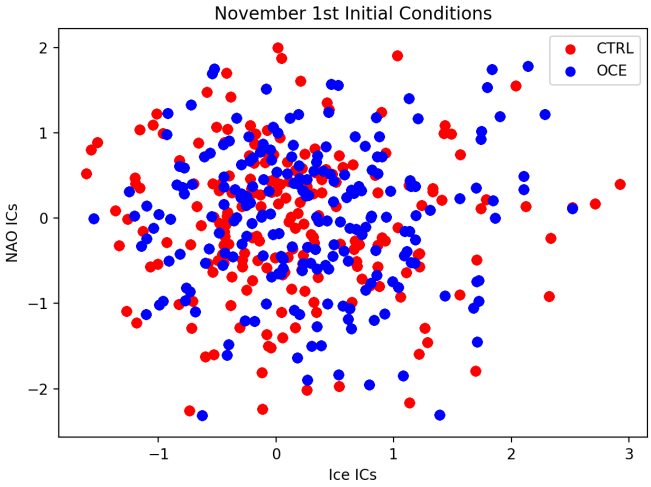

Figure B7. Scatter plot of the ice/NAO initial conditions, i.e. the values of the BKS sea ice (x-axis) and NAO anomalies (y-axis) on November 1st, for every year between 1980 and 2015. In red: CTRL; in blue: OCE.

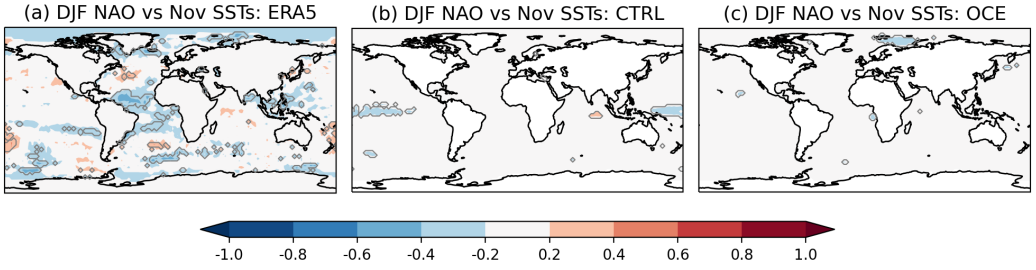

Figure B8. Correlations between detrended November SSTs at each gridpoint and the DJF NAO timeseries for (a) ERA5, (b) CTRL and (c) OCE. The ensemble members of the CTRL and OCE experiments have been concatenated together. The period covered is 1980-2015. The grey contour highlights statistically significant correlations ($p < 0.05$).

Author contributions. KS and SJ jointly generated the main model simulations used. KS performed the analysis, created the plots and led the writing of the manuscript. SJ developed and implemented the stochastic schemes used and assisted in the writing of the manuscript and interpretation of results. FC carried out the original LIM analysis of ERA5 and assisted in further analysis and interpretation of the LIM.

Competing interests. We declare there are no competing interests.

Acknowledgements. KS gratefully acknowledges funding from the Thomas Philips and Jocelyn Keene Junior Research Fellowship, Jesus College, Oxford. The experiments considered in this study were created as part of, and while supported by, the Horizon 2020 project PRIMAVERA (grant number 641727), funded by the European Research Council. SJ is contributing with this work to the projects M3 and L4 of the Collaborative Research Centre TRR 181 "Energy Transfer in Atmosphere and Ocean" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 274762653. The main experiments used in this study were generated using supercomputing units from an ECMWF Special Project: we thank ECMWF support staff for their patience and invaluable assistance. We also thank three anonymous reviewers and the co-editor for extensive feedback which greatly improved the quality of the paper.

References

Alexander, M. A., Matrosova, L., Penland, C., Scott, J. D., and Chang, P.: Forecasting Pacific SSTs: Linear inverse model predictions of the PDO, Journal of Climate, 21, 385–402, 2008.

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., and Scaife, A. A.: An Intercomparison of Skill and Overconfidence/Underconfidence of the Wintertime North Atlantic Oscillation in Multimodel Seasonal Forecasts, Geophysical Research Letters, https://doi.org/10.1029/2018GL078838, 2018.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, Journal of Hydrometeorology, 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

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

Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Cámara, A., Weisheimer, A., Weniger, M., Williams, P. D., and Yano, J.-I.: Stochastic parameterizations: Toward a New View of Weather and Climate Models, Bulletin of the American Meteorological Society, 98, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1, 2017.

Blackport, R. and Screen, J. A.: Observed statistical connections overestimate the causal effects of arctic sea ice changes on midlatitude winter climate, Journal of Climate, https://doi.org/10.1175/JCLI-D-20-0293.1, 2021.

Brankart, J.-M., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P.-A., Brasseur, P., and Verron, J.: A generic approach to explicit simulation of uncertainty in the NEMO ocean model, Geoscientific Model Development, 8, 1285–1297, https://doi.org/10.5194/gmd-8-1285-2015, 2015.

Brayshaw, D. J., Hoskins, B., and Blackburn, M.: The basic ingredients of the north atlantic storm track. Part II: Sea surface temperatures, Journal of the Atmospheric Sciences, https://doi.org/10.1175/2011JAS3674.1, 2011.

Caian, M., Koenigk, T., Döscher, R., and Devasthale, A.: An interannual link between Arctic sea-ice cover and the North Atlantic Oscillation, Climate Dynamics, 50, 423–441, https://doi.org/10.1007/s00382-017-3618-9, 2018.

Chapman, William and National Center for Atmospheric Research Staff (Eds).: MS Windows NT Kernel Description, https://climatedataguide.ucar.edu/climate-data/walsh-and-chapman-northern-hemisphere-sea-ice, 2013.

Christensen, H. M., Berner, J., Coleman, D. R., and Palmer, T. N.: Stochastic parameterization and El Niño-southern oscillation, Journal of Climate, 30, 17–38, https://doi.org/10.1175/JCLI-D-16-0122.1, 2017.

Davini, P., Von Hardenberg, J., Corti, S., Christensen, H. M., Juricke, S., Subramanian, A., Watson, P. A., Weisheimer, A., and Palmer, T. N.: Climate SPHINX: Evaluating the impact of resolution and stochastic physics parameterisations in the EC-Earth global climate model, Geoscientific Model Development, 10, 1383–1402, https://doi.org/10.5194/gmd-10-1383-2017, 2017.

Dawson, A. and Palmer, T. N.: Simulating weather regimes: impact of model resolution and stochastic parameterization, Climate Dynamics, 44, 2177–2193, https://doi.org/10.1007/s00382-014-2238-x, 2015.

Deser, C., Walsh, J. E., and Timlin, M. S.: Arctic sea ice variability in the context of recent atmospheric circulation trends, Journal of Climate, https://doi.org/10.1175/1520-0442(2000)013<0617:ASIVIT>2.0.CO;2, 2000.

Deser, C., Tomas, R. A., and Peng, S.: The transient atmospheric circulation response to North Atlantic SST and sea ice anomalies, Journal of Climate, https://doi.org/10.1175/JCLI4278.1, 2007.

Deser, C., Sun, L., Tomas, R. A., and Screen, J.: Does ocean coupling matter for the northern extratropical response to projected Arctic sea ice loss?, Geophysical Research Letters, 43, 2149–2157, https://doi.org/https://doi.org/10.1002/2016GL067792, 2016.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M., and Knight, J.: Skilful predictions of the winter North Atlantic Oscillation one year ahead, Nature Geoscience, 9, 809–814, https://doi.org/10.1038/ngeo2824, 2016.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N.: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world?, Geophysical Research Letters, 41, 5620–5628, https://doi.org/10.1002/2014GL061146, 2014.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/GMD-9-1937-2016, 2016.

Frankignoul, C., Czaja, A., and L'Heveder, B.: Air-sea feedback in the North Atlantic and surface boundary conditions for ocean models, Journal of Climate, https://doi.org/10.1175/1520-0442(1998)011<2310:ASFITN>2.0.CO;2, 1998.

García-Serrano, J., Frankignoul, C., Gastineau, G., and de la Cámara, A.: On the predictability of the winter Euro-Atlantic climate: Lagged influence of autumn Arctic sea ice, Journal of Climate, https://doi.org/10.1175/JCLI-D-14-00472.1, 2015.

Gent, P. R. and Mcwilliams, J. C.: Isopycnal Mixing in Ocean Circulation Models, Journal of Physical Oceanography, 20, 150 – 155, https://doi.org/10.1175/1520-0485(1990)020<0150:IMIOCM>2.0.CO;2, 1990.

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Corti, S., Davini, P., Exarchou, E., Fabiano, F., Fladrich, U., Fuentes Franco, R., García-Serrano, J., von Hardenberg, J., Koenigk, T., Levine, X., Meccia, V. L., van Noije, T., van den Oord, G., Palmeiro, F. M., Rodrigo, M., Ruprich-Robert, Y., Le Sager, P., Tourigny, E., Wang, S., van Weele, M., and Wyser, K.: HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – description, model computational performance and basic validation, Geoscientific Model Development, 13, 3507–3527, https://doi.org/10.5194/gmd-13-3507-2020, 2020.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geoscientific Model Development, 9, 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016, 2016.

Hawkins, E. and Sutton, R.: Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling, Journal of Climate, 22, 3960–3978, 2009.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hoskins, B. J. and Karoly, D. J.: The steady linear response of a spherical atmosphere to thermal and orographic forcing., Journal of the Atmospheric Sciences, https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2, 1981.

Hurrell, J. W., Kushnir, Y., Otterson, G., and Visbeck, M.: An Overview of the North Atlantic Oscillation, The North Atlantic Oscillation: Climatic Significance and Environmental Impact, 134, 263, https://doi.org/10.1029/GM134, 2003.

Juricke, S. and Jung, T.: Influence of stochastic sea ice parametrization on climate and the role of atmosphere-sea ice-ocean interaction, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372, https://doi.org/10.1098/rsta.2013.0283, 2014.

Juricke, S., Lemke, P., Timmermann, R., and Rackow, T.: Effects of Stochastic Ice Strength Perturbation on Arctic Finite Element Sea Ice Modeling, Journal of Climate, 26, 3785 – 3802, https://doi.org/10.1175/JCLI-D-12-00388.1, 2013.

Juricke, S., Goessling, H. F., and Jung, T.: Potential sea ice predictability and the role of stochastic sea ice strength perturbations, Geophysical Research Letters, 41, 8396–8403, https://doi.org/https://doi.org/10.1002/2014GL062081, 2014.

Juricke, S., Palmer, T. N., and Zanna, L.: Stochastic sub-grid scale ocean mixing: Impacts on low frequency variability, Journal of Climate, pp. JCLI–D–16–0539.1, https://doi.org/10.1175/JCLI-D-16-0539.1, 2017.

Juricke, S., MacLeod, D., Weisheimer, A., Zanna, L., and Palmer, T. N.: Seasonal to annual ocean forecasting skill and the role of model and observational uncertainty, Quarterly Journal of the Royal Meteorological Society, 144, 1947–1964, https://doi.org/https://doi.org/10.1002/qj.3394, 2018.

Keeley, S. P., Sutton, R. T., and Shaffrey, L. C.: The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.1912, 2012.

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

Koenigk, T. and Brodeau, L.: Arctic climate and its interaction with lower latitudes under different levels of anthropogenic warming in a global coupled climate model, Climate Dynamics, https://doi.org/10.1007/s00382-016-3354-6, 2017.

Koenigk, T., Mikolajewicz, U., Jungclaus, J. H., and Kroll, A.: Sea ice in the Barents Sea: Seasonal to interannual variability and climate feedbacks in a global coupled model, Climate Dynamics, https://doi.org/10.1007/s00382-008-0450-2, 2009.

Kolstad, E. W. and Screen, J. A.: Nonstationary Relationship Between Autumn Arctic Sea Ice and the Winter North Atlantic Oscillation, Geophysical Research Letters, 46, 7583–7591, https://doi.org/https://doi.org/10.1029/2019GL083059, 2019.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J.: Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation, Journal of Climate, https://doi.org/10.1175/JCLI-D-15-0654.1, 2016.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Madec, G. and the NEMO team: NEMO ocean engine version 3.6 stable, Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace, 27, 2016.

Magnusdottir, G., Deser, C., and Saravanan, R.: The Effects of North Atlantic SST and Sea Ice Anomalies on the Winter Circulation in CCM3. Part I: Main Features and Storm Track Characteristics of the Response, Journal of Climate, 17, 857 – 876, https://doi.org/10.1175/1520-0442(2004)017<0857:TEONAS>2.0.CO;2, 2004.

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

Mori, M., Kosaka, Y., Watanabe, M., Taguchi, B., Nakamura, H., and Kimoto, M.: Reply to: Is sea-ice-driven Eurasian cooling too weak in models?, Nature Climate Change, 9, 937–939, https://doi.org/10.1038/s41558-019-0636-0, 2019b.

Newman, M., Sardeshmukh, P. D., and Penland, C.: How important is air-sea coupling in ENSO and MJO evolution?, Journal of Climate, https://doi.org/10.1175/2008JCLI2659.1, 2009.

Notz, D. and Community, S.: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, e2019GL086 749, https://doi.org/https://doi.org/10.1029/2019GL086749, e2019GL086749 10.1029/2019GL086749, 2020.

Palmer, T., Buizza, R., Jung, T., Leutbecher, M., Shutts, G. J., Steinheimer, M., and Weisheimer, A.: Stochastic Parameterization and Model Uncertainty, vol. 598, http://www.ecmwf.int/publications/, 2009.

Palmer, T. N.: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction, Quarterly Journal of the Royal Meteorological Society, 138, 841–861, https://doi.org/10.1002/qj.1923, 2012.

Peings, Y.: Ural Blocking as a Driver of Early-Winter Stratospheric Warmings, Geophysical Research Letters, 46, 5460–5468, https://doi.org/https://doi.org/10.1029/2019GL082097, 2019.

Peings, Y. and Magnusdottir, G.: Response of the wintertime northern hemisphere atmospheric circulation to current and projected arctic sea ice decline: A numerical study with CAM5, Journal of Climate, https://doi.org/10.1175/JCLI-D-13-00272.1, 2014.

Penland, C. and Magorian, T.: Prediction of Niño 3 sea surface temperatures using linear inverse modeling, Journal of Climate, 6, 1067–1076, 1993.

Penland, C. and Sardeshmukh, P.: The optimal growth of tropical sea surface temperature anomalies, J. Climate, 8, 1999–2024, 1995.

Sanchez, C., Williams, K. D., and Collins, M.: Improved stochastic physics schemes for global weather and climate models, Quart. J. Roy. Meteor. Soc., 2015.

Scaife, A. A. and Smith, D.: A signal-to-noise paradox in climate science, npj Climate and Atmospheric Science, https://doi.org/10.1038/s41612-018-0038-4, 2018.

Screen, J. A., Deser, C., Smith, D. M., Zhang, X., Blackport, R., Kushner, P. J., Oudar, T., McCusker, K. E., and Sun, L.: Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models, Nature Geoscience, https://doi.org/10.1038/s41561-018-0059-y, 2018.

Shepherd, T. G.: Atmospheric circulation as a source of uncertainty in climate change projections, Nature Geoscience, 7, 703–708, https://doi.org/10.1038/ngeo2253, 2014.

Siegert, S., Stephenson, D. B., Sansom, P. G., Scaife, A. A., Eade, R., and Arribas, A.: A Bayesian framework for verification and recalibration of ensemble forecasts: How uncertain is NAO predictability?, Journal of Climate, 29, 995–1012, https://doi.org/10.1175/JCLI-D-15-0196.1, 2016.

Siew, P. Y. F., Li, C., Sobolowski, S. P., and King, M. P.: Intermittency of Arctic–mid-latitude teleconnections: stratospheric pathway between autumn sea ice and the winter North Atlantic Oscillation, Weather and Climate Dynamics, https://doi.org/10.5194/wcd-1-261-2020, 2020.

Siew, P. Y. F., Li, C., Ting, M., Sobolowski, S. P., Wu, Y., and Chen, X.: North Atlantic Oscillation in winter is largely insensitive to autumn Barents-Kara sea ice variability, Science Advances, 7, eabg4893, https://doi.org/10.1126/sciadv.abg4893, 2021.

Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, https://doi.org/10.1088/1748-9326/aade56, 2018.

Strommen, K. and Palmer, T. N.: Signal and noise in regime systems: A hypothesis on the predictability of the North Atlantic Oscillation, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.3414, 2019.

Strommen, K., Christensen, H. M., Berner, J., and Palmer, T. N.: The impact of stochastic parametrisations on the representation of the Asian summer monsoon, Climate Dynamics, 0, 1–14, https://doi.org/10.1007/s00382-017-3749-z, 2017.

Strommen, K., Christensen, H. M., Macleod, D., Juricke, S., and Palmer, T. N.: Progress towards a probabilistic Earth system model: Examining the impact of stochasticity in the atmosphere and land component of EC-Earth v3.2, Geoscientific Model Development, https://doi.org/10.5194/gmd-12-3099-2019, 2019.

Strong, C. and Magnusdottir, G.: The role of Rossby wave breaking in shaping the equilibrium atmospheric circulation response to North Atlantic boundary forcing, Journal of Climate, https://doi.org/10.1175/2009JCLI2676.1, 2010.

Strong, C. and Magnusdottir, G.: Dependence of NAO variability on coupling with sea ice, Climate Dynamics, https://doi.org/10.1007/s00382-010-0752-z, 2011.

Strong, C., Magnusdottir, G., and Stern, H.: Observed Feedback between Winter Sea Ice and the North Atlantic Oscillation, Journal of Climate, 22, 6021 – 6032, https://doi.org/10.1175/2009JCLI3100.1, 2009.

Sun, L., Deser, C., and Tomas, R. A.: Mechanisms of Stratospheric and Tropospheric Circulation Response to Projected Arctic Sea Ice Loss, Journal of Climate, 28, 7824 – 7845, https://doi.org/10.1175/JCLI-D-15-0169.1, 2015.

Vancoppenolle, M., Bouillon, S., Fichefet, T., Goosse, H., Lecomte, O., Morales Maqueda, M. A., and Madec, G.: The Louvain-la-Neuve sea ice model, Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace, 31, 2012.

Vidale, P. L., Hodges, K., Vannière, B., Davini, P., Roberts, M. J., Strommen, K., Weisheimer, A., Plesca, E., and Corti, S.: Impact of stochastic physics and model resolution on the simulation of tropical cyclones in climate GCMs, Journal of Climate, https://doi.org/10.1175/JCLI-D-20-0507.1, 2021.

Vinje, T.: Anomalies and trends of sea-ice extent and atmospheric circulation in the Nordic Seas during the period 1864-1998, Journal of Climate, https://doi.org/10.1175/1520-0442(2001)014<0255:AATOSI>2.0.CO;2, 2001.

Von Storch, J. S.: Signatures of air-sea interactions in a coupled atmosphere-ocean GCM, Journal of Climate, https://doi.org/10.1175/1520-0442(2000)013<3361:SOASII>2.0.CO;2, 2000.

Wang, C., Zhang, L., Lee, S.-K., Wu, L., and Mechoso, C. R.: A global perspective on CMIP5 climate model biases, Nature Climate Change, 4, 201–205, 2014.

Wang, L., Ting, M., and Kushner, P. J.: A robust empirical seasonal prediction of winter NAO and surface climate, Scientific Reports, 7, 279, https://doi.org/10.1038/s41598-017-00353-y, 2017.

Warner, J. L., Screen, J. A., and Scaife, A. A.: Links Between Barents-Kara Sea Ice and the Extratropical Atmospheric Circulation Explained by Internal Variability and Tropical Forcing, Geophysical Research Letters, https://doi.org/10.1029/2019GL085679, 2020.

Watson, P. A. G., Berner, J., Corti, S., Davini, P., von Hardenberg, J., Sanchez, C., Weisheimer, A., and Palmer, T. N.: The impact of stochastic physics on tropical rainfall variability in global climate models on daily to weekly timescales, Journal of Geophysical Research: Atmospheres, 122, 5738–5762, https://doi.org/10.1002/2016JD026386, 2017.

Woollings, T., Hannachi, A., and Hoskins, B.: Variability of the North Atlantic eddy-driven jet stream, Quarterly Journal of the Royal Meteorological Society, 136, 856–868, https://doi.org/10.1002/qj.625, 2010.