North Pacific zonal wind response to sea ice loss in the Polar Amplification Model Intercomparison Project and its downstream implications

Bryn Ronalds1 · Elizabeth A. Barnes1 · Rosie Eade2 · Yannick Peings3 · Michael Sigmond4

Received: 3 March 2020 / Accepted: 22 June 2020 / Published online: 9 July 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Recent studies suggest that the wintertime North Pacific eddy-driven jet stream will strengthen and extend eastward in response to Arctic sea ice loss. Using output from the Polar Amplification Model Intercomparison Project we examine the mean change of the North Pacific wintertime zonal winds, and use cluster analysis to explore the change in sub-seasonal, wintertime variability in zonal winds between experiments with future Arctic sea ice concentrations relative to a pre-industrial run. Further, given the relationship between the North Pacific jet stream and North American weather regimes, we also examine the changes in surface temperature variability over North America. The four climate models investigated here exhibit robust agreement in both sign and structure of the atmospheric responses, with a strengthened wintertime North Pacific jet, an increase in anomalously strong and extended jet events, and a decreased frequency of weakened and equatorward-shifted jet events in response to reduced Arctic sea ice. The models also show changes in wintertime, North American surface temperature patterns that are consistent with the zonal wind changes seen in the North Pacific. There is an increase in the frequency of occurrence of the North American temperature dipole pattern, defined as anomalously warm temperatures in the west or northwest and anomalously cold temperatures in the east or southeast, and a decrease in the frequency of anomalously cold temperatures over North America.

Keywords Atmospheric circulation · Arctic amplification · Jet stream variability

1 Introduction

Many recent studies have examined the impacts of Arctic warming and sea ice loss on both the Northern Hemisphere large scale circulation (e.g. Ronalds et al. 2018; Screen et al. 2018a, b; Zappa et al. 2018; Blackport and Screen 2019; Ronalds and Barnes 2019) and midlatitude weather regimes (e.g. Sellevold et al. 2016; Cohen et al. 2018; Overland and Wang 2018a, b; Li and Luo 2019). Much of the work looking at the midlatitude eddy-driven jet streams has shown that, both in the zonal mean and over the North Atlantic basin, the eddy-driven jet stream weakens and shifts equatorward in response to Arctic amplification (e.g. Screen et al. 2018b). Recent work by Ronalds and Barnes (2019), however,

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00382-020-05352-w) contains supplementary material, which is available to authorized users.

Bryn Ronalds bryn.ronalds@colostate.edu
Elizabeth A. Barnes elizabeth.barnes@colostate.edu
Rosie Eade rosie.eade@metoffice.gov.uk
Yannick Peings ypeings@uci.edu
Michael Sigmond michael.sigmond@canada.ca

1 Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA
2 MetOffice Hadley Centre, Exeter, Devon, UK
3 Department of Earth System Science, University of California, Irvine, Irvine, CA, USA
4 Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC, Canada
showed that the winter (January–March) mean North Pacific eddy-driven jet stream strengthens and extends in response to Arctic amplification in a fully coupled climate model, with no change in jet latitude. The exact mechanisms leading to this North Pacific response have not yet been determined, though changes in wavebreaking have been shown to cause a strengthened jet in idealized modeling experiments (Ronalds and Barnes 2019). If the wintertime mean North Pacific jet does indeed strengthen and extend in response to Arctic amplification, this could have downstream impacts on North American temperatures and precipitation (e.g. Strong and Davis 2008), particularly if any mean changes in the wintertime jet stream are also associated with changes in the storm track’s sub-seasonal variability.

One hypothesis concerning the possible impacts of Arctic sea ice loss on the atmospheric circulation is that it will lead to changes in Northern Hemisphere midlatitude weather regimes via alteration of the large-scale atmospheric circulation (e.g. Kug et al. 2015; Lee et al. 2015; Cvijanovic et al. 2017; Li and Luo 2019). Of growing interest is the role of Arctic sea ice loss on the North American wintertime temperature dipole, defined as anomalously warm temperatures in the west and severe cold to the east (e.g. Wang et al. 2015a, b; Lee et al. 2015; Cvijanovic et al. 2017; Chien et al. 2019). This temperature pattern occurs when the wintertime climatological North American geopotential height ridge/trough pattern becomes strongly amplified and persistent, leading to anomalously warm and dry weather conditions in the west and severe cold temperatures in the east (e.g. Wang et al. 2015a, b; Singh et al. 2016). While this ridge/trough pattern is associated with both topography and land-sea contrast, recent studies have argued that the North Pacific circulation is the dominant factor in its amplification and persistence (e.g. Teng and Branstator 2017; Swain et al. 2017), and others have showed an indirect link between Arctic sea ice loss and the amplified ridge/trough pattern (Lee et al. 2015; Cvijanovic et al. 2017). We therefore hypothesize that any changes to the North Pacific zonal wind field will have associated downstream impacts for North American weather. Specifically, changes in the North Pacific may lead to increased cold air outbreaks (e.g. Kug et al. 2015), or possibly increased frequency of occurrence of the warm west/cold east North American temperature dipole (e.g. Lee et al. 2015; Chien et al. 2019). Given the previous findings of a strengthened and extended North Pacific jet stream in response to Arctic sea ice loss (Ronalds et al. 2018), and the work linking the atmospheric circulation in the North Pacific to North American weather regimes (e.g. Jaffe et al. 2011; Lee et al. 2015; Griffin and Martin 2017; Swain et al. 2017; Chien et al. 2019), we explore the change in sub-seasonal variability of both the zonal winds over the North Pacific and the surface temperatures over North America in response to Arctic sea ice loss.

While there have been numerous studies examining the regional atmospheric impacts of Arctic sea ice loss in recent decades, there is still considerable uncertainty in the possible consequences of Arctic warming and sea ice loss (see Screen et al. 2018b; Smith et al. 2019; Cohen et al. 2020, and references therein). Much of the uncertainty derives from poor understanding of the particular physical mechanisms, the large internal variability associated with atmospheric circulations and regional weather, and differences between models and modelling experiments (Smith et al. 2019). This is addressed in the sixth Coupled Model Intercomparison Project (CMIP6; Eyring et al. 2016) by coordinating a Polar Amplification Model Intercomparison Project (PAMIP; Smith et al. 2019). The goal of the project is to coordinate a multi-model sea ice loss and Arctic warming set of experiments. Each modeling centre is given identical forcing files and follows the same experimental protocol, giving an unprecedented set of coordinated sea ice loss experiments across multiple climate models (see Smith et al. 2019).

This work aims to answer the following questions using output from PAMIP:

1. How does the wintertime North Pacific eddy-driven jet stream respond to the same Arctic sea ice loss across multiple models?
2. What changes in the internal variability make up the wintertime mean jet stream response?
3. Do we see consistent changes in downstream surface temperatures associated with the changes to the North Pacific eddy-driven jet?

## 2 Data and methods

### 2.1 Data

The Polar Amplification Model Intercomparison Project (PAMIP) is a subset of ongoing CMIP6 experiments (Eyring et al. 2016). The goal of the project is to identify the forced response of the global climate system to changes in polar sea ice, as well as differentiate the relative roles of local sea ice concentrations (SIC) versus remote sea surface temperatures (SST) in driving polar amplification. To this end, numerous experiments were constructed, with tier classifications. Tier 1, the first experiments to be run, consists of atmosphere-only time slice experiments. All experiments are initialized using conditions from April 1st, 2000 and are run for 14 months, discarding the first two months as spin-up. Each climate model produces at least 100 ensemble members per run, and the method of ensemble generation differs between each model. Each experiment has one forcing change, either sea surface temperature (SST) conditions or sea ice concentration (SIC) conditions. There are three categories for both...
the SST and SIC conditions: pre-industrial, present day or future. The pre-industrial conditions are based on the multi-model ensemble historical runs from CMIP5 (Taylor et al. 2012), the future ice conditions represent a multi-model mean under a two-degree global warming scenario, and present day conditions consist of the 1979–2008 climatology from the Hadley Centre Sea Ice and Sea Surface Temperature data set (Rayner et al. 2003). For more information on the experimental set-ups and the derivations of the forcing files see Smith et al. (2019).

In this work we compare results from two of the Tier 1 experiments (labeled 1.5 and 1.6). Experiment 1.5 consists of present day SST’s and pre-industrial Arctic sea ice, while experiment 1.6 also consists of present day SST’s but with future Arctic sea ice. For the remainder of this work these two experiments will be referred to as piArcSIC (1.5) and futArcSIC (1.6). Four models provide daily data for these two experiments: CESM2 (Danabasoglu et al. 2020), CanESM5 (Swart et al. 2019), HadGEM3 (Walters et al. 2019) and SC-WACCM4 (Smith et al. 2014). Because recent studies have shown that the stratosphere plays an important role in the circulation response to Arctic sea ice loss (e.g. Sun et al. 2015; Nakamura et al. 2016; Zhang et al. 2018; Romanowsky et al. 2019; Wu et al. 2019), it is important to note that only HadGEM3 and SC-WACCM4 have fully resolved stratospheres, with both having greater vertical resolution than CESM2 and CanESM5 (see Table 1 for more information on each model). CESM2 and SC-WACCM4 each contain 100 ensemble members per run, generated by perturbing the initial temperatures (perturbations are on the order of $10^{-14}$ K). The CanESM5 also produced 100 ensemble members by first branching off of 10 independent AMIP simulations, then perturbing each nine more times by changing a random seed parameter in the cloud scheme. HadGEM3 produced 150 members generated using the stochastic physics method (see Ciavarella et al. 2018, for more details).

We use daily zonal winds at 700 hPa (U700) over the North Pacific basin (5–85 °N, 120–240 °E) and daily surface temperatures ($T_s$) over North America (5–85 °N, 200–320 °E). Because our interest is in the sub-seasonal time scale, both the winds and surface temperatures are temporally smoothed using a 10-day low-pass Lanczos filter with 100 coefficients before we limit our data to midwinter only (January–February). The results are qualitatively similar when using either a 5-day or 7-day filter, although they are much noisier. We chose to omit December and March from the wintertime analysis as the monthly mean North Pacific zonal wind fields were significantly different from the January–February means and so focus only on the consistent midwinter response. All four models have daily U700 data, but at this time we only have daily surface temperatures for CESM2, CanESM5 and HadGEM3.

In order to test for significance when examining the change in North Pacific U700 we use the Wilks (2016) False Discovery Rate (FDR) method, which accounts for the large spatial autocorrelation in the data. To apply this method you must first define the $\alpha_{FDR}$, which depends on the scale of autocorrelation, and use it to calculate the FDR threshold (Eq. 1). In this case we used $\alpha_{FDR} = 2\alpha_{global}$, where $\alpha_{global}$ is the chosen significance level. The $FDR_{crit}$ acts as a threshold to find the “true” $p$-value cut-off, given the spatial autocorrelation present in the data. This is done by plotting the $FDR_{crit}$ against the calculated $p$ value for all gridpoints (size N) and finding the point of intersection (see Fig. 3 in Wilks 2016). Only those grid points with a $p$-value below the point of intersections are considered significant.

$$FDR_{crit} = \frac{i}{N} \alpha_{FDR}, \quad \text{where} \quad i = 1, \ldots, N \quad (1)$$

### 2.2 Cluster analysis

In order to examine the modes of internal variability within our data we perform a cluster analysis technique known as k-means clustering (Hartigan and Wong 1979). K-means cluster analysis categorizes the entirety of a data set into a user-specified number of clusters (or centroids). Thus, there is some subjectivity in the number of centroids chosen, which will be discussed further below. The algorithm is straightforward to apply to large datasets, and there are no orthogonality constraints as in Empirical Orthogonal Functions, though the classification is more rigid, with each day belonging to a single centroid. The algorithm operates

---

### Table 1  Model information

| Model Name | Institution | Horiz. resolution | Vert. resolution |
|------------|-------------|------------------|-----------------|
| CESM2      | National Center for Atmospheric Research | 0.9° × 1.25° | 33 levels (2 hPa) |
| CanESM5    | Canadian Centre for Climate Modelling and Analysis | 2.8° × 2.8° | 49 levels (1 hPa) |
| HadGEM3    | MetOffice UK Hadley Centre | 0.8° × 0.5° | 85 levels (85 km) |
| SC-WACCM4  | National Center for Atmospheric Research | 2.5° × 1.9° | 66 levels (140 km) |

Horizontal resolution is given in degrees longitude by degrees latitude. Vertical resolution includes the number of vertical levels and the model top.
iteratively, assigning individual data points to the closest centroid, defined as the centroid with the minimum squared Euclidean distance (Hartigan and Wong 1979). At each iteration, the centroids move to the middle of the data points they represent, and the new minimum distances are calculated. This process is repeated until the centroids remain nearly stationary, at which point the algorithm stores the centroids and the distances. In this study, we repeat this entire training process 500 times in order to ensure that a global, rather than local, minimum is found. The resulting centroids chosen from the 500 iterations are from the iteration with the minimum summed distance.

Before performing the cluster analysis we first calculated the daily anomalies of both the piArcSIC (pre-industrial Arctic sea ice) and futArcSIC (future Arctic sea ice) experiments. This was done by removing the ensemble mean for each day, i.e. remove the ensemble mean for January 1st from all ensemble members, then do the same for January 2nd, and so on, in each experiment separately (piArcSIC versus futArcSIC). By removing the daily means rather than a single January–February climatology we eliminate any seasonality from the data, ensuring that the resulting clusters are not skewed towards the beginning of January versus end of February. By doing this to both experiments separately and then combining the resulting daily anomalies, we have removed the ensemble-mean forced change and are left with the the daily variability about that mean. The cluster analysis identifies patterns applying to both experiments, and the frequencies of occurrence can be compared across the two experiments. For CESM2, CanESM5 and SC-WACCM4 this means we have two experiments, each with 100 ensemble members of 59 days (January–February, no leap years), giving us 11,800 daily anomalies. For HadGEM3, we have 150 ensemble members of 60 days (30-day months), giving us 18,000 days. The k-means cluster analysis is then applied to the full pool of daily anomalies for each model separately. In order to choose the number of clusters we tested a range from 4-9 using the daily U700 anomalies and compared the resulting centroids from each model. Ultimately, six centroids appeared to be sufficient to account for the typical variability within the data while also ensuring the centroids remained distinct.

Applying this process to the North Pacific U700 daily anomalies results in six centroids that represent the six maps of zonal wind anomalies that best describe the main patterns of variability about the mean. The algorithm also outputs the categorization of each input day into its respective cluster. Separating the input days back into the two experiments, piArcSIC and futArcSIC, we can calculate the frequencies of occurrence of each cluster: how many days from each experiment look like each centroid? This allows us to compare the frequencies from piArcSIC and futArcSIC and see if certain clusters, or zonal wind anomaly patterns, become more or less frequently visited with Arctic sea ice loss. The significance of these frequency changes is tested using a bootstrapping approach whereby we shuffle days from both experiments together and randomly split the data in half and recalculate the cluster frequency change between each half. This process is repeated 10,000 times, and the resulting distribution of frequency changes define the null distribution. We then choose as our cutoff the two-tailed 80% and 90% confidence regions.

The same k-means cluster analysis is also applied to the North American daily anomalous surface temperatures from both experiments. For consistency with our North Pacific zonal wind analysis six centroids were again chosen, and significance was tested using the same bootstrapping approach.

3 Results

3.1 North Pacific zonal wind variability

The change in the ensemble mean January–February mean zonal winds at 700 hPa (U700) across the North Pacific basin is shown in Fig. 1, represented by the shading. The contours in Fig. 1 represent the piArcSIC mean U700, and the black dots represent the FDR significance at the 99% confidence level. All four models show a significantly strengthened North Pacific jet stream (red shading) in the futArcSIC experiment, with strong easterly anomalies along the poleward flank (blue shading), indicating a narrowed jet. CESM2, CanESM5 and HadGEM3 (Fig. 1a–c) also show an extended jet, with strong westerly anomalies extending further east across the basin. CESM2, HadGEM3, and SC-WACCM4 (Fig. 1a, c, d) also exhibit easterly anomalies along the equatorward flank of the jet. None of the models show evidence of a shifted jet in the North Pacific.

Next, we decompose the mean change in U700 into the change in sub-seasonal variability using the k-means clustering analysis described above, and compare it to the k-means cluster analysis of the surface temperatures (see next section). This is done for all four models individually, and we find that all four models result in broadly similar cluster patterns in which we group the North Pacific U700 centroids. The k-means cluster analysis of North American $T_z$ also results in similar centroids for CESM2, CanESM5 and HadGEM3. The similarity across models allows us to broadly define the six main patterns of U700 and $T_z$ variability. Due to these strong similarities, we only show the results from CESM2 for the remainder of the paper as they are representative of all four models, although we include comparisons across models in the supplemental materials (Fig. SM1–SM8).

The six centroids representing the main patterns of variability in the North Pacific daily anomalous U700 in CESM2
Fig. 1  The change in January–February ensemble mean zonal wind at 700 hPa across the North Pacific basin between futArcSIC and piArcSIC for a CESM2 ($n = 5900$), b CanESM5 ($n = 5900$), c HadGEM ($n = 9000$), and d SC-WACCM4 ($n = 5900$). Shading denotes wind change and contours denote the piArcSIC mean. Black dots represent FDR significance at the 99% confidence level.

Fig. 2  The six main patterns of daily anomalous U700 variability for January–February (shading). Contours represent the piArcSIC January–February mean U700, and bold face on the $\% \Delta$ represents significance at 90% confidence. All results are for CESM2.
are shown in Fig. 2. Each pattern is given a descriptor which will be used throughout the remainder of this paper:

(a) Super Strengthen jet is extremely strong, extended eastward, and more zonal and/or shifted equatorward slightly.
(b) Strengthen/Extend jet is strong, extended eastward, and shifted slightly poleward, particularly in the exit region.
(c) Poleward Tilt jet is strong, and the exit region is dominantly shifted poleward.
(d) Weaken/Retract jet is retracted, confined mostly to the west Pacific, as well as broader and weakened.
(e) Equatorward Shift jet is shifted equatorward, and weakened in some models, particularly SC-WACCM4 (see Fig. SM2 in the online supplementary materials).
(f) Poleward Shift: jet is weakened and shifted poleward.

The percent change in frequency for each pattern is calculated and included in the panel titles in Fig. 2. In CESM2, the three strengthening patterns resulting from k-means cluster analysis all increase in frequency in futArcSIC relative to piArcSIC (Fig. 2a–c). Conversely, the three weakened and/or shifted jet patterns all decrease in frequency. While in CESM2 only the Super Strengthen, Equatorward Shift and Poleward Shift patterns (Fig. 2a, e, f) are considered significant at the 90% confidence level (indicated by the bold %Δ in the titles), we can compare the changes in frequency of the similar patterns across all four models. We show these in Fig. 3, where each bar indicates the percent change in frequency of each pattern in futArcSIC relative to piArcSIC for all four models. We have included two thresholds for significance calculated using a bootstrapping approach: 80% confidence, denoted by the darker colour bars and the single asterisk, and 90%, denoted by the double asterisk. While 80% confidence is a lower threshold than commonly used, any multi-model agreement further increases our confidence in the changes in frequency. While one model showing 80% confidence means there is a 20% chance of error, two models both passing the 80% threshold suggests an error of 0.2^2, i.e. a 4% chance of error. The three strengthened jet patterns are shown in Fig. 3a, while the three weakened jet patterns are shown in Fig. 3b.

All four models show sign agreement in the percent change of frequency for the Super Strengthen (increased, Fig. 3a), Poleward Tilt (increased, Fig. 3a), and Equatorward Shift (decrease, Fig. 3b) patterns. Further, the majority of the models show significance of at least 80% for these three patterns, increasing our confidence that this variability response is forced by the Arctic sea ice loss, rather than just noise. For the other three patterns, we consider the changes seen in Strengthen/Extend (Fig. 3a) and Poleward Shift (Fig. 3b) to likely be sampling noise since the changes in frequency are small and there is substantial model disagreement. The Weaken/Retract pattern (Fig. 3b), however, is complicated by the fact that in SC-WACCM4 there were two centroids that exhibited significant jet retraction and weakening, the one we labeled Weaken/Retract, which increased very slightly, and the centroid we labeled Equatorward Shift, which decreased significantly (see Fig. SM2 in supplemental materials). This, coupled with the strong decrease in frequency of the Weaken/Retract pattern for both CanESM5 and HadGEM3, at a 90% confidence level, suggests that the decreased frequency of this pattern may also be a robust forced variability response to Arctic sea ice loss.

Figures 2 and 3 suggest that there is an increase in strengthened and extended January-February North Pacific jet events (Super Strengthen and Poleward Tilt patterns, Fig. 2a, c), and a decrease in weakened and retracted jet events (Weaken/Retract and Equatorward Shift patterns, Fig. 2d, e) in response to Arctic sea ice loss, in addition to the mean strengthening and extension of the jet seen in Fig. 1. We can also use the output of the k-means cluster analysis to determine if, on any given day in futArcSIC, the winds are stronger than a similar day in piArcSIC. In other words, we want to know if the futArcSIC days assigned to each centroid have different full wind fields than the piArcSIC days, particularly in the jet region. To answer this, we use the classification of each day to a specific cluster to create composite maps of the full, North Pacific U700 field for each cluster for both piArcSIC and futArcSIC separately and take the difference. These are shown in Fig. 4, where shading represents the difference in wind field composites and contours represent the piArcSIC composites. For all six patterns in CESM2 there are stronger winds in the vicinity of the jet in futArcSIC (red shading), and generally weakened winds elsewhere (blue shading). Again, this is true for each of the six patterns across all four models (see Fig. SM3–SM4 in supplemental materials). Further, the biggest differences in full wind fields occur in the three weakened jet patterns: Weaken/Retract, Equatorward Shift, and Poleward Shift (Fig. 4d–f). This suggests that not only is the jet generally stronger on a day-to-day basis in the futArcSIC experiment in response to Arctic sea ice loss, but that days characterized as retracted and weak jet events, or shifted jet events, have much faster jets than their piArcSIC counterparts, contributing strongly to the mean strengthened jet response seen in Fig. 1.

In the January-February mean, the North Pacific jet is strengthened and extended in response to Arctic sea ice loss across all four models (Fig. 1). This increased jet strength represents not only a generally faster jet on a day-to-day basis (Fig. 4), but also an increase in individual strengthened jet events and a decrease in weakened and/or shifted jet events (Fig. 2). Based on the results shown in Fig. 3, and the level of model agreement, it appears that Arctic sea ice loss is leading to significant changes in frequency of certain
sub-seasonal variability patterns. There is an increased frequency in Super Strengthen and Poleward Tilt patterns, and a decreased frequency in Weaken/Retract and Equatorward Shift patterns. This is of interest in terms of downstream impacts.

Many studies have looked at the relationships between the North Pacific jet stream variability and North American weather regimes. Notably, Griffin and Martin (2017) found that a strengthened and extended North Pacific jet (similar to our Super Strengthen pattern) is associated with a strong ridge formation along the North American west coast, and the development of the North American temperature dipole pattern, with anomalously warm temperatures in the west and anomalously cold temperatures in the east. Weaken/Retract jet events were found to be associated with anomalously cold west coast and warm east coast, and the Equatorward Shift pattern was associated with anomalously cold air over the northern half of North America (Griffin and Martin 2017, their Figs. 4, 5 and 7). These findings, in conjunction with our own, suggest that Arctic sea ice loss in these PAMIP experiments may lead to changes in North American surface temperature variability.
**CESM2: U700 composite differences**

(a) Super Strengthen
\[ \Delta \text{freq} = +11.1\% \]

(b) Strengthen/Extend
\[ \Delta \text{freq} = +2.2\% \]

(c) Poleward Tilt
\[ \Delta \text{freq} = +3.5\% \]

(d) Weaken/Retract
\[ \Delta \text{freq} = -4.1\% \]

(e) Equatorward Shift
\[ \Delta \text{freq} = -10.8\% \]

(f) Poleward Shift
\[ \Delta \text{freq} = -10.0\% \]

Fig. 4 Difference between futArcSIC and piArcSIC full field U700 composite maps, based on the daily anomalous U700 variability patterns (shading). Contours represent the piArcSIC daily U700 composite maps, and bold face on the \( \Delta \text{freq} \) represents significance at 90% confidence. All results are for CESM2.

---

**CESM2: \( T_s \) anomaly composites**

(a) Super Strengthen

(b) Strengthen/Extend

(c) Poleward Tilt

(d) Weaken/Retract

(e) Equatorward Shift

(f) Poleward Shift

Fig. 5 Composite maps of daily anomalous North American \( T_s \) from the piArcSIC experiment, based on the daily anomalous U700 variability patterns (shading). Contours represent the piArcSIC January–February mean surface temperatures. All results are for CESM2.
3.2 Downstream surface temperature variability

If the same relationships between the North Pacific jet stream and North American surface temperatures found by Griffin and Martin (2017) hold true in the PAMIP atmosphere-only experiments, we would expect that under futArcSIC there will be an increase in the occurrence of the North American warm west/cold east temperature dipole and a decrease in occurrence of anomalously cold temperature events over west and northern North America. In order to establish whether this is the case, we use the results of the North Pacific anomalous U700 k-means cluster analysis for CESM2, CanESM5 and HadGEM3. Once again, the CESM2 results are shown as representative of the three models (see SM5–SM8 in supplemental figures for all three models’ T variability results). Using the days assigned to each pattern for the piArcSIC experiment, we calculate the composite maps of the downstream, North American surface temperature daily anomalies associated with each North Pacific U700 variability pattern (Fig. 5). The shading represents the composite daily anomalies of $T_s$, and the contours represent the piArcSIC January-February climatology. As in Figs. 2, 3 and 4 the three strengthened jet patterns are in the top row (Fig. 5a–c), and the three weakened jet patterns are in the bottom row (Fig. 5d–f).

Figure 5 shows that the relationships shown in Griffin and Martin (2017) are also found in the CESM2 PAMIP experiments, and the same is true of the other models (see Fig. SM5–SM6 in supplemental materials). The Super Strengthen jet events are associated with the warm west/cold east North American temperature dipole (Fig. 5a), the Weaken/Retract jet events are associated with cold temperatures over Canada and the northern United States and warm temperatures in the south and southeast (Fig. 5d), and the Equatorward Shift pattern is associated with cold temperature anomalies over most of the continent and anomalously warm temperatures over northern Alaska and northwest Canada (Fig. 5e). The Poleward Tilt pattern is associated with anomalously warm temperatures over Alaska and western Canada, and cool temperatures to the south (Fig. 5f). Thus, we expect that in futArcSIC we would see an increased frequency of temperature patterns with warm temperatures to the west and northwest and cold temperatures to the east and south (Fig. 5a, c). We also expect to see a decreased frequency of anomalously cold temperatures over North America (Fig. 5d, e). To test this we next apply our cluster analysis approach to the daily surface temperature anomalies over North America.

Similar to the k-means cluster analysis done for the North Pacific U700 anomalies, we remove the daily ensemble mean $T_s$ from both piArcSIC and futArcSIC daily $T_s$ fields and combine the resulting anomalies across the two forcing simulations. For consistency, we again use six centroids to describe the six main patterns of variability within the North American surface temperatures, allowing for a comparison to the composite patterns shown in Fig. 5. Again, we show results from CESM2 as they are representative of all three models (see Fig. SM7–SM8 in supplemental materials). It is important to note that there are drivers of wintertime North American surface temperatures other than the North Pacific jet stream, and thus there is no expectation that the six $T_s$ anomalies patterns will perfectly match the composite patterns shown in Fig. 5, though we do find strong similarities. Thus the patterns shown in Fig. 6 are roughly oriented to most closely match the patterns in Fig. 5 when possible. The shading represents the daily anomaly patterns associated with each centroid, and the contours are the piArcSIC January–February climatology. Percent frequency changes considered significant at the 90% confidence level are bolded in the titles. The centroid resulting from the k-means cluster analysis is assigned a pattern name, which will be used in the remainder of this paper.

The six main patterns of daily anomalous $T_s$ variability over North America are as follows:

(a) Warm W/Cold E warm temperatures in the west and cool temperatures in the east. This is the amplified January–February climatology pattern, i.e. the North American temperature dipole.

(b) Warm Air Outbreak strong warm temperature anomalies over Canada and northern USA.

(c) Cold NW/Warm SE anomalously cold temperatures over Alaska and western Canada, with anomalously warm temperatures over most of the United States.

(d) Cold Air Outbreak anomalously cold temperatures over most of North America, with warmer temperatures over northern Alaska.

(e) Warm NW/Cold SE anomalously warm temperatures over Alaska and northwest Canada, and anomalously cold temperatures over the rest of the continent.

(f) Warm East anomalously warm temperatures over the eastern half of North America.

While many of the patterns appear consistent with those in Fig. 5, we note that none match well with the Poleward Tilt pattern in Fig. 5c. However, there are some strong similarities in the other patterns, as can be seen when comparing panels a, b, d–f in both Figs. 5 and 6. There is also consistency in the percent changes in frequency for the matching patterns. The Warm W/Cold E and Warm Air Outbreak patterns (Fig. 6a, b) are similar in structure to the temperature composites of the Super Strengthen and Strengthen/Extend patterns (Fig. 5a, b), and also increase in frequency, as we had expected. The Cold Air Outbreak, Warm NW/Cold SE and Warm East patterns (Fig. 6d–f) are also structurally similar to the
composites found for Weaken/Retract, Equatorward Shift and Poleward Shift patterns (Fig. 5d–f), and all three also decrease in frequency of occurrence.

Figure 7 shows the percent change in frequencies for each of the six patterns for all three models. As in Fig. 3, darker colours and the single asterisk represent significance at 80% confidence, and the double asterisk represents significance at the 90% confidence level. Only two patterns show consistency in changes of frequency across all three models. The Warm W/Cold E pattern (first cluster in Fig. 7a) increases significantly across all three models, which is consistent with the increased frequency of the Super Strengthen pattern (Fig. 3a). Additionally, the Cold Air Outbreak pattern (first cluster in Fig. 7b) decreases in frequency in all three models at the 90% confidence level, again consistent with the decreased frequency of the Weaken/Retract pattern in Fig. 3b for the same models. Model disagreement in the percent frequency changes for the remaining four patterns (Warm Air Outbreak, Cold NW/Warm SE, Warm NW/Cold SE, and Warm East, Fig. 7) could be due to multiple factors. Notably, the surface temperatures are driven by a multitude of factors, not just by the North Pacific jet stream, which is the framework we have used here for the cluster analysis.

4 Discussion and conclusions

This work uses recently available results from four models running the PAMIP atmosphere-only, time-slice experiments. We compare two sets of experiments, future versus pre-industrial, both with present day sea surface temperature forcing but differing Arctic sea ice concentrations. We examine the changes in sub-seasonal variability for both lower-level North Pacific zonal winds and North American surface temperatures between the two experiments, and identify changes caused by Arctic sea ice loss. While the changes in January-February mean North Pacific U700 are small across all of the models (on the order of 2 ms⁻¹, Fig. 1), they are still considered significant at a 99% confidence level. The large number of ensemble members for each experiment is what allows us to distinguish a forced response from the internal noise of the system. Further, the strong model agreement in the mean U700 change adds to our confidence that the changes we are seeing, specifically a strengthened and extended North Pacific eddy-driven jet stream with negative wind anomalies along the poleward flank, are truly forced by the Arctic sea ice loss. This is of particular interest, as the previous assumption was that the midlatitude jet streams would weaken and/
or shift equatorward in response to Arctic warming and sea ice loss (e.g. Screen et al. 2018b). This assumption is also the basis of the “tug-of-war” concept, where the warming Arctic and warming tropics have opposite and competing influences on the jet streams, with the upper-tropospheric tropical warming shifting the jets poleward (e.g. Woollings and Blackburn 2012; Barnes and Screen 2015; McGraw and Barnes 2016; Peings 2018). While our results suggest that this tug-of-war concept may not be accurate, particularly in the North Pacific, this set of sea ice loss experiments do not include changes to sea surface temperatures, i.e. the role of the background state, and ocean feedbacks, both of which have been shown to play an integral role in the remote responses to Arctic sea ice loss and Arctic warming (e.g. Deser et al. 2015, 2016; Smith et al. 2017, 2019). Thus, further work using the PAMIP experiments with changes to SST’s, as well as the future coupled experiments, are needed to address the tug-of-war assumption currently in use.

Similar to the small mean changes, the changes in frequency of various sub-seasonal U700 variability patterns over the North Pacific are also generally small. Those
significant at an 80% confidence level for individual models range from differences of only 0.8–3.88% of all days. However, by comparing these frequency changes across the four models, all of which show very similar patterns of internal variability (see Fig SM1-SM8 in the supplemental materials), we gain more confidence that some of the frequency changes, albeit small, are forced by Arctic sea ice loss:

- **Super Strengthen** all four models show an increased frequency, and three of the models exhibit significance with at least 80% confidence.
- **Poleward Tilt** all four models show an increased frequency, and two of the models exhibit significance of 90% confidence.
- **Weaken/Retract** three of four models show a decreased frequency, two of which are significant at 90% confidence.
- **Equatorward Shift** all four models show a decreased frequency, and all are significant at 90% confidence.

While we consider the changes in frequency of the North Pacific U700 anomaly patterns to be directly forced by Arctic sea ice loss in these PAMIP experiments, it is important to consider the limitations of applying these findings to the observations. There are fewer than 3000 January-February days in the observational record (compared to the 5900 or even 9000 days in these experiments), in addition to the fact that sea ice loss in these experiments is abrupt, rather than evolving over time. Low frequency variability, such as the El Niño Southern Oscillation, is also not included in these experiments, along with atmosphere-ocean coupling, both of which are key components in examining North Pacific jet stream variability, as well as North American surface temperatures. Further, the observed sea ice loss has inter-annual temporal variability, whereas these experiments are run with a forcing that only varies as a function of season. With that said, these experiments provide examples of what we may expect as consequences of Arctic sea ice loss acting in isolation, and these conclusions are strengthened by the consistency between models. In addition, the sea ice loss in these modelling experiments is based on a multi-model mean, and is a more conservative estimate than what observations have shown, as discussed in Smith et al. (2019).

Arctic sea ice loss also leads to changes in North American surface temperature sub-seasonal variability in the PAMIP simulations. Again, using significance thresholds for individual models and comparing those to the multi-model agreement increases our confidence that some of the changes we see are indeed forced by Arctic sea ice loss. Specifically, in the PAMIP futArcSIC experiment we see an increase in the number of days with anomalously warm temperatures to the west and cold temperatures to the east, i.e. the North American temperature dipole, and a decreased number of days with anomalous cold air outbreaks over central and western North America. Both of these changes are consistent with the upstream changes in the North Pacific, as previously shown in Griffin and Martin (2017), supporting the theory that, since North American weather regimes depend, in part, on atmospheric activity over the North Pacific (e.g. Jaffe et al. 2011; Griffin and Martin 2017; Swain et al. 2017; Chien et al. 2019), Arctic sea ice loss can indirectly affect North American weather via changes in the North Pacific circulation patterns (e.g. Lee et al. 2015).

While the forced response of the atmosphere to Arctic sea ice loss in the atmosphere-only time-slice PAMIP experiments is very small, particularly compared to the internal variability of the system, the large number of ensemble members allowed us to identify some robust responses. However, the level of noise within the system in these experiments makes identifying and exploring the underlying physical processes that may account for the responses seen extremely difficult. Thus, we are unable to comment on the mechanism(s) which may cause the North Pacific jet to strengthen in response to Arctic sea ice loss. A future subset of PAMIP experiments are planned, in particular, the coupled ocean-atmosphere extended experiments, which may provide some insight (see Table 1 in Smith et al. 2019). Of particular interest are the possible changes in North Pacific wavebreaking associated with the sea ice loss, which previous work has linked to a strengthened jet (Ronalds and Barnes 2019).

**Acknowledgements** The authors thank Lantao Sun at Colorado State University for the use of the CESM2 data. BR and EAB are supported by the Climate and Large-scale Dynamics Program of the National Science Foundation under Grant AGS-1545675, and EAB is also supported under Grant AGS-1749261. RE is supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra and also supported by APPLICATE (European Union Horizon 2020 research and innovation program, Grant Number 727862). YP is supported by the National Science Foundation Grant AGS-1624038 and the Department of Energy Grant DE-SC0019407.

**References**

Barnes EA, Screen JA (2015) The impact of Arctic warming on the midlatitude jet-stream: can it? Has it? Will it? Wiley Interdiscip Rev Clim Change 6:277–286. https://doi.org/10.1002/wcc.337

Blackport R, Screen JA (2019) Influence of arctic sea ice loss in autumn compared to that in winter on the atmospheric circulation. Geophys Res Lett 46(4):2213–2221

Chien YT, Wang SYS, Chikamoto Y, Voelker SL, Meyer JDD, Yoon JH (2019) North american winter dipole: observed and simulated changes in circulations. Atmosphere 10(12):793

Ciavarella A, Christidis N, Andrews M, Groenendijk M, Rostron J, Elkington M, Burke C, Lott FC, Stott PA (2018) Upgrade of the HadGEM3-A based attribution system to high resolution and a new validation framework for probabilistic event attribution. Weather Clim Extremes 20:9–32
Sanchez C, Smith C, Stratton R, Tennant W, Tomassini L, Van Weverberg K, Vosper S, Willett M, Browse J, Bushell A, Carslaw K, Dalvi M, Essery R, Gedney N, Hardiman S, Johnson B, Johnson C, Jones A, Jones C, Mann G, Milton S, Rumbold H, Sellar A, Ujiie M, Whitall M, Williams K, Zerroukat M (2019) The met office unified model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. Geosci Model Dev 12(5):1909–1963

Wang SY, Hipps L, Gillies RR, Yoon JH (2015a) Probable causes of the abnormal ridge accompanying the 2013–2014 california drought: ENSO precursor and anthropogenic warming footprint. Geophys Res Lett 41:3220–3226

Wang SYS, Huang WR, Yoon JH (2015b) The north american winter ‘dipole’ and extremes activity: a CMIP5 assessment. Atmos Sci Lett 16(3):338–345

Wang SYS, Yoon JH, Becker E, Gillies R (2017) California from drought to deluge. Nat Clim Change 7(7):465–468

Wilks DS (2016) “The stippling shows statistically significant grid points”: how research results are routinely overstated and overinterpreted, and what to do about it. Bull Am Meteorol Soc. https://doi.org/10.1175/BAMS-D-15-00267.1

Woollings T, Blackburn M (2012) The north Atlantic jet stream under climate change and its relation to the NAO and EA patterns. J Clim 25:886–902. https://doi.org/10.1175/JCLI-D-11-00087.1

Wu Y, Simpson IR, Seager R (2019) Intermodel spread in the northern hemisphere stratospheric polar vortex response to climate change in the CMIP5 models. Geophys Res Lett 46(22):13290–13298

Zappa G, Pithan F, Shepherd TG (2018) Multimodel evidence for an atmospheric circulation response to arctic sea ice loss in the CMIP5 future projections. Geophys Res Lett 45(2):1011–1019

Zhang P, Wu Y, Simpson IR, Smith KL, Zhang X, De B (2018) A stratospheric pathway linking a colder siberia to Barents-Kara sea ice loss. Sci Adv 4(7):eaaat6025

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.