Responses to Reviewers

First of all, we wish to thank both reviewers for their careful reading of the paper and for the insightful and constructive suggestions. In the following pages, we provide point-by-point responses to every comment from the reviewers. Our responses are in italics; the reviewers' comments are in regular font. Line numbers (LXX-LYY) refer to revised manuscript.

Anonymous Referee #1
Received and published: 4 October 2018

Overview and broad comments
This study by Walsh et al uses a long historical record of sea ice coverage data to estimate the amount of variance explained in two (mainly September) quantities, the Beaufort Sea Index and pan-Arctic sea ice extent, by the sea ice extent in different Arctic sub-regions. They consider both concurrent correlations and lagged correlations of these quantities with sea ice conditions for previous months, and separate these correlations into a total and interannual estimation. They argue that a piece-wise linear trend is most appropriate for detrending the different time series for estimating the latter.

The noteworthy pieces of information to come out of this study are (1) a piece-wise linear trend is the best option for detrending the time series considered in the study, (2) the break-point year emergent in that analysis which can be interpreted as an acceleration of the ongoing negative trend in sea ice cover is in the mid- to late- 1990's, (3) that interannual variability in the BSI can be explained by June sea ice coverage in the Beaufort Sea and by July coverage in the Chukchi Sea, (4) consistent with other studies the September pan-Arctic ice extent has significant autocorrelation back to July (about two months), and (5) that the Laptev and East Siberian Seas explain the most concurrent correlation in September pan-Arctic SIE.

As it stands though, in my opinion the paper requires substantial revisions and additional analyses before being re-submitted.

The key goal of the study seems to be to provide baseline metrics against which sea ice prediction studies can be evaluated. However, that baseline has already been established for one of the two predictands considered in this study, pan-Arctic SIE, in several other studies based on autocorrelation (i.e. persistence). No reference to these other studies is ever made. The one thing that separates the result about the pan-Arctic SIE predictand in this study from others is that a very long historical record was used here, but that point should be emphasized when motivating the study (it's currently not mentioned). The rest of the lagged correlation analysis for pan-Arctic SIE, which was between it and the SIE for the various subregions, didn't yield higher correlations than lagged pan-Arctic sea ice extent itself. So, for evaluating prediction skill, is it not the autocorrelation of pan-Arctic SIE that is the important baseline to beat? What therefore is the motivation to consider the lagged correlation analysis with the different sub-regions?

We agree that the baseline for persistence-based predictions have been established in previous studies (e.g., Blanchard-Wrigglesworth et al., 2011; Day et al., 2014; Bushuk et al., 2017, 2018), and the revision places added emphasis on the use of a longer record length (back to 1953 rather than 1979), cf. new paragraphs: L122-134; L161-167; L 316-346, as well as new Fig. 5). The use of autocorrelations and cross-correlations was essentially a vehicle for illustrating the issues associated with detrending in a predictive framework. The main intent of the paper is to show how detrending is a key step in the depiction of persistence-based statistical predictions. We illustrate the effect of detrending for both pan-Arctic ice exten{t} and regional metrics in order to show that predictive applications on both scales must address detrrending in a rigorous way, and that there are various alternatives for detrending. While these alternative detrending strategies are known, the relative effectiveness of the various alternatives has not been addressed in previous studies, as we state in the new paragraph on p. 6, L149-167. Goldstein et al., 2016, The Cryos. Disc.; 2018, Sci. Rep.) come closest by comparing representations based on linear trends and discontinuities in the mean. An addition novel outcome of the present study is the synthesis of break-point information across the various Arctic subregions and (in a reviewer-based revision) across seasons.

With respect to the BSI predictand, that analysis seems okay and is fine to report on, but I'm not sure all of the information presented (particularly in the tables; see specific comments below) is needed to reach the conclusions made.

As indicated in the detailed responses below, we have relegated two of the tables (previously Tables 3 and 4) to Supplementary Material.

I was also a bit surprised (and disappointed) to see that only these two predictands were considered if the goal is to provide baseline skill numbers, especially considering the first predictand has already received considerable attention in other studies. Like in the supplementary material of Bushuk et al 2018 (reference below), it would be good to extend the analysis to treat all of the different regions as separate predictands and compute the autocorrelations for each of those regions. This would help put the results of the Bushuk et al study into context, as they used a shorter and different observational dataset than is used here. As of now, and considering some of my suggestions to remove certain parts of the results section (see specific comments below), this additional analysis would help strengthen that section and would provide useful baseline numbers for future studies.

This suggestion is incorporated into the revision, and it is especially helpful in emphasizing that a novel feature is the use of longer time series than have been used on other recent studies that used only the post-1978 satellite data. In the revision, we present comparisons of the regional autocorrelations for two periods, 1953 onward and 1979 onward, with explicit reference to the Bushuk et al. (2018) regional autocorrelations (new paragraph L316-346 and new Fig. 5). Our use of the two time periods leads to the conclusion that the springtime "predictability barrier" in regional forecasts based on persistence of ice extent anomalies is not reduced by the inclusion of several decades of pre-satellite data. We highlight this result in the Abstract (L20-22) and in the Conclusion (Section 6), L475-477.

I was missing from the introduction a sufficient argument for the need for this study in the context of current literature on sea ice prediction. The additional references to, and discussion of some key studies listed below are needed to place the goals of this study and its findings into perspective. Specifically here are studies on persistence/autocorrelation of observed and modeled pan-Arctic and regional sea ice extent:

Blanchard-Wrigglesworth, Edward, et al. "Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations." Journal of Climate 24.1 (2011): 231-250.
– Includes estimates of autocorrelation for pan-Arctic SIE and SIA in both observations and models. Also includes contributions of lagged Beaufort Sea sea ice thickness to september SIA and SIE variability.
– This paper was referenced in the study, but it's odd that only the SST re-emergence result was mentioned.

Sigmond, M., et al. "Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system." Geophysical Research Letters 40.3 (2013): 529-534.
– Their Fig. 3 shows persistence forecasts of SIE from two observational datasets.

Bushuk, Mitchell, et al. "Regional Arctic sea–ice prediction: potential versus operational seasonal forecast skill." Climate Dynamics (2018): 1-23.
– In their supplementary material, they show autocorrelations for individual regions from observations and compare against models.

These references are indeed relevant. We have carefully gone through each one and identified points of comparison for inclusion (with references) in the revision. These references are not cited at various points in the text, beginning with a new Section 2 (Previous Work), L148-167. As noted above, we also include quantitative comparisons of the post-1953 and post-1979 autocorrelations in order to broaden the temporal context relative to the previous studies. We distinguish between the persistence information derived from models and from observational data, as studies such as Blanchard-Wrigglesworth et al. (2011), Sigmund et al. (2013) and Bushuk et al. (2018) include both. We also add information (including explicit references) to other relevant papers that have addressed trends and autocorrelation issues in a predictive context: Day et al. (2014, J. Climate), Bushuk et al. (2017, J. GRL), Goldstein et al. (2016, The Cryos. Disc.).

Throughout the paper, the word "predictability" is used somewhat misleadingly. It should be made clear early on that what is being referred to as predictability throughout the rest of the paper is a very specific estimate of predictability – that is, the predictability of a certain sea ice coverage quantity based on the lagged cross-correlation or auto-correlation (often referred to as memory or persistence) between that quantity and other sea ice coverage quantities. Otherwise, it sounds like the authors are referring to predictability in general, which encompasses all sources of predictability and estimates of theoretical predictability limits.

Agreed. We have included a new paragraph (L122-134) clarifying that we are addressing only persistence-based statistical predictability using the observational record, and we have expanded the introductory discussion (L72-106) so that it states clearly that there are other sources of predictability, especially those captured by physical-dynamical models of the coupled atmosphere-ocean-ice system. We have inserted "persistence-based" in front of forecasts and predictability at various points in the text, beginning with L10 in the Abstract. We do note that the original manuscript was "up front" in referring to "forecast skill achieved by other methods such as more sophisticated statistical formulations, numerical models, and heuristic approaches" (Abstract, first sentence).

Guemas, Virginie, et al. "A review on Arctic seaice predictability and prediction on seasonal to decadal timescales". Quarterly Journal of the Royal Meteorological Society 142.695 (2016): 546-561.
– Provides a review on sources of predictability, including memory/persistence of sea ice coverage and thickness. See references therein.

Bushuk, Mitchell, et al. "Regional Arctic sea–ice prediction: potential versus operational seasonal forecast skill." Climate Dynamics (2018): 1-23.
– Focuses on regional predictability.

The revised Introduction cites the review paper by Guemas et al. (2016) on L106, and it highlights Bushuk et al.'s (2017) finding that the ocean surface and subsurface state contributes to predictability in the North Atlantic subarctic (L103-105).

The de-trending analysis could be expanded on. The 1979-onward period is what is commonly used in prediction studies, and there is an ongoing debate in the literature about what de-trending method is most appropriate for pan-Arctic SIE, particularly when the most recent few years are included. Extending the de-trending analysis to focus also on this shorter period would be helpful for future prediction studies. While most authors choose to fit a linear trend over that period, Dirkson et al 2017 suggested a quadratic fit; others suggest de-trending with a high-pass filter but this has the unfortunate effect of removing the first or last sample from the data over an already short period. It would be worthwhile, and would add to the paper, if the authors could make a case for the piece-wise linear trend over this shorter period if indeed it is much better than a linear or quadratic fit. Also, as prediction studies are beginning to focus more on regional sea ice prediction than on pan-Arctic SIE/SIA, it would be helpful to know what choice is most appropriate for the different regions and whether this choice depends on the month or season. For instance, Bushuk et al 2017 and Dirkson et al 2017 argued that linear de-trending was sufficient for regional SIE and local SIC over a similar period considered here.

Dirkson, Arlan, William J. Merryfield, and Adam Monahan. "Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions." Journal of Climate 30.3 (2017): 1001-1017.

Bushuk, Mitchell, et al. "Skillful regional prediction of Arctic sea ice on seasonal timescales." Geophysical Research Letters 44.10 (2017): 4953-4964.

Following this suggestion, we have expanded on our evaluation of the benefits of a piecewise linear fit relative to a single linear trend line. The new text discussing the expanded results is on L316-346. Specifically, we have identified all cases (across regions and the winter, spring and summer seasons) in which the piecewise fit reduces the residual variance by more than 5% relative to a single linear trend. Only these cases are included in our (revised) summary figure containing – for each season separately – the temporal histograms of the break-point years. The revised figure, containing plots for the three seasons preceding September, replaces the original Figure 5, which combined the results for the three seasons into a single bar plot.

Specific Comments

• It should be stated in the abstract early on (before L11-13) the prediction method that will be used in this paper.

The text beginning on L9 of the Abstract has been modified to: "In this study, we use observational data to evaluate the contribution of the trend to the skill of persistence-based statistical forecasts of monthly and seasonal ice extent on the pan-Arctic and regional scales."

• L19: "statistical predictability" should be replaced with "statistical skill..." based on said approach. It's not statistical predictability in the sense of a theoretical limit, which isn't directly possible to estimate.

Agreed. We now use "statistical skill" here (L16) and in other instances where "predicatability" had been incorrectly used. The previous item (revision of lines 9-11 in the Abstract) is an example of the rewording to "skill" rather than "predicatability".

• L17-18; Which month(s) are being referred to here?

"September" is now specifically stated (L18, L22).

• L31-32; Statement requires a reference.

The AMAP (2017) reference in the preceding sentence will be inserted here (L32). We have also added a reference on L30 to a new paper on Arctic indicators: Box et al. (2018, "Key indicators of Arctic climate change, 1971-2017", Env. Res. Lett., in press).

• L37-39; Should cite:
Maslanik, James, et al. "Distribution and trends in Arctic sea ice age through spring 2011." Geophysical Research Letters 38.13 (2011).

Maslanik et al. (2011) is now cited on L42, and also the more recent NOAA (2018) Arctic Report Card (L44).

• L47-50; This positive trend in the Bering Sea in winter is disappearing if more recent years are taken into account... Ah, this is stated in the conclusions, but is probably more appropriate to place here instead.

New statement about the Bering Sea has been added (L48-49): "However, the positive trend of Bering Sea ice largely vanishes when the most recent winters (especially 2017-18) are included." We also mention this again in the Conclusion (L455-457)

• The departures are also affected by antecedent sea ice conditions themselves. Please see the references given in the general comments above related to persistence.

We assume this comment refers to lines 74-76 of the original manuscript. We have added ",,,in addition to antecedent sea ice conditions themselves" in the revision (L77-78).

• L79-82; This is not correct. "Ice-ocean model" implies that both the ice and ocean evolve freely as determined by the model (only the atmospheric forcing is required). Also, the statement ignores the fact that fully coupled models, which determine both the atmospheric and ocean/ice conditions prognostically, are now used more often for sea ice prediction. These models are also limited by the chaotic nature of the climate system, but this is typically accounted for by running ensembles. There is important literature on predictability using said models that isn't referenced or discussed here. Please see general comment above highlighting predictability literature.

Text has been modified as follows (L80-91): "Even ice-ocean models initialized to current sea ice and ocean conditions require atmospheric forcing in order to predict future ocean states. Moreover, fully coupled models, which determine both the atmospheric and ocean/ice conditions prognostically, are now used increasing often for seasonal sea ice predictions. Ensembles of coupled simulations are generally run because of the chaotic nature of the climate system. These models can be run for much longer time periods than the observational sea ice record, so they can provide statistics of sea ice persistence (autocorrelations) subject to the "perfect model" assumption. Examples of studies employing the "perfect model" approach are Holland et al. (2011) Blanchard-Wrigglesworth et al. (2011), Day et al. (2014), Bushuk et al. (2017) and Bushuk et al. (2018). In these model simulations, autocorrelation of sea ice anomalies tends to be greater in the model results than in observational data (e.g., Blanchard-Wrigglesworth et al., 2011, their Fig. 2; Day et al., 2014, their Fig. 1)."

• L83-84; It's difficult to see how this paper is an "extension" of Drobot 2003; while this study also focuses on the Beaufort and Chukchi Seas, and extends an analysis of statistical prediction skill in that region based on more recent observational data, the analysis done in Drobot 2003 (statistical prediction using multiple linear regression with many more predictors) is not repeated here.

We have deleted this sentence, but have retained a subsequent summary of Drobot's (2003) relevant study (L102-117).

• L88-92; The caution raised in the Drobot 2003 study I think definitely deserves

attention given Arctic sea ice and Arctic climate have changed, as stated. Although this study distinguishes between prediction of sea ice coverage with and without the trend, as written it sounds like the authors will carry out an analysis similar to Drobot 2003, and determine the impact of the changing conditions on the statistical relationships found in that study, which is not what is done here.

Text has been revised (L110-112) to "While the present study will not include the type of multiple-predictor evaluation carried out by Drobot (2003), it will provide a more regionally comprehensive and updated assessment of sea ice anomaly persistence in a predictive context".

• L95-96; This is incorrect. While the Drobot 2003 study didn't consider the effects of detrending as stated, Blanchard-Wrigglesworth et al 2011 did. Specifically they detrended the model projections of SIE by subtracting the ensemble mean at each point in time (removing the forced signal). Additionally, they detrended the SIE observations by subtracting the long-term linear trend.

Sentence originally in lines 95-96 has been deleted.

• L97-100; I find this overview too vague and is broader in scope than what is actually done in the paper. I think it would be more accurate and helpful to the reader if the authors provided, in order of appearance in the paper, what will specifically be done.

Lines 97-100 has been replaced by the following paragraph (L1223-134) containing a more specific summary of the manuscript's subsequent sections: "In the present paper, we use the autocorrelation statistic to quantify the skill of persistence as a control forecast of pan-Arctic and regional sea ice extent. In addition to utilizing the more conventional metric of ice extent in regional and pan-Arctic domains, we include a regional sea ice index developed in the 1970s to capture interannual variations of marine access in the Beaufort Sea.. A primary focus of the evaluation is the method of detrending the data, as various alternative methods have not been fully explored in the literature. We show that the piecewise linear method generally results in the smallest residual variance about the trend line, and we then perform an across-region synthesis of information on the break-points in the trend lines in different seasons. Our period of analysis extends back to 1953, which results in a considerably larger sample of years than the more commonly used satellite period (1979 onward). Finally, we examine lagged cross-correlations to determine whether pan-Arctic ice extent or Beaufort Sea summer ice conditions are foreshadowed in a statistical sense by antecedent ice conditions in particular subregions of the Arctic."

• L102; "seasonal climatologies, persistence, and trend" of what? Should clarify that you are referring to these quantities for sea ice coverage itself.

Revised (L136-137) to "arising from climatological sea ice coverage, sea ice persistence, and sea ice trends".

• L104; "ice-ocean models" should be replaced with "dynamical models". Ice-

ocean models are a specific subset of these, but fully-coupled models are used more often.

Revised (L138) to "dynamical models" to include fully coupled models as well as ice-ocean models.

• L105; The Sea Ice Outlook was managed by the Sea Ice Prediction Network, and starting in 2018 became managed by the Sea Ice Prediction Network–Phase 2. The Sea Ice Outlook is not a previous name for Sea Ice Prediction Network. See https://www.arcus.org/sipn/sea-ice-outlook.

Revised (L139-141) to "The Sea Ice Outlook, coordinated by the Sea Ice Prediction Network now in its Phase 2 (https://www.arcus.org/sipn/sea-ice-outlook, accessed 27 Dec 2018), provides an annual compilation…"

• L118; Suggest changing "maps" to "gridded records". Presumably what is being referred to here is data in digitized form and not the presentation of the data literally as maps.

 "maps" changed to "digitized records"(L175)

• L124-127; Say the product name here. The product is described in the next sentence without actually saying explicitly that it is used.

Revised (L182-186) to "…we compute ice extent using the gridded Arctic-wide sea ice concentration product known as "Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward (Walsh et al., 2015), referred to in the National Snow and Ice Data Center (NSIDC) catalog as G10010. This dataset is based on observations from approximately 15 historical sources…".

• L154-155; While this is true, previous studies have shown that there are shorter persistence time scales for pan-Arctic SIE than for pan-Arctic SIA due to high frequency dynamical influences that change SIE, but not SIA (Blanchard-Wrigglesworth et al 2011; their Fig. 14). This should be mentioned here.

New text inserted in L214-218: "It should be noted, however, that persistence time-scales of pan-Arctic sea ice area have been shown in previous studies (e.g., Blanchard-Wrigglesworth et al., 2011) to be longer than those of pan-Arctic sea ice extent because high-frequency forcing can change ice extent more than it changes ice area (i.e., by converging or diverging ice floes in the absence of ridging or melt)."

• L157ff; Was there any interpolation done to get the G10010 dataset onto the same grid used to define the MASIE sub-regions?

No. The MASIE region mask defines the regions on its grid. For our comparisons, we used the region code of the closest MASIE grid cell for each grid cell in the G10010 grid. No data needed to be interpolated to re-project the MASIE region mask onto the G10010 grid.

• L182; Somewhere early on in the methods section it should be stated what years are considered in the study.

Text has been modified (L246-247) to say "As shown in Figure 3, Arctic ice extents have generally been decreasing over the post-1953 period of this study."

• L191-192; Please clarify. I think what is meant is that the presence of a trend in a time series inflates forecast skill when using the anomaly correlation coefficient to assess that skill. It should be noted though that for distance metrics like the root mean square error, skill can actually be inflated by detrending the data if the two time series have different magnitudes of trends.

This statement referred to the use of simple persistence (autocorrelation) of the time series to be predicted, in which case there is no second time series. However, we see the reviewer's point for cases in which a second variable (e.g., another region's ice extent) is the predictor. The reworded L254-255: "However, a trend can inflate persistence-based forecast skill when a variable is used to predict itself (assuming the historical trend continues into the future)".

• L198-202; please refer to general comment on de-trending, but this paragraph would be a good place to reference choices made in other sea ice prediction studies.

While the revision includes new references and descriptions of their relevance to the present study (cf. response to this reviewer's earlier paragraph beginning with "The key goal of this study…"), we have also replaced the original Lines 202-204 with the following (L263-271): "The previous studies cited in Section 2 (e.g., Blanchard-Wrigglesworth et al., 2011; Sigmund et al., 2013; Day et al., 2014; Bushuk et al., 2017, 2017) have generally relied on least-squares linear fits for detrending. Goldstein et al. (2016, 2018), by contrast, showed that discontinuous changes in the mean better captured time series (such as open water area) characterized by abrupt changes. In the spirit of the Goldstein et al. studies, we explore various options for detrending a time series such as those in Figures 2 and 3, for which the changes are more pronounced in recent decades than in earlier decades. In such cases, a single multi-decadal trend line cannot be expected to optimally represent the historical evolution."

• L209ff; As of now, I find the description of the fitting methodology a bit hard to follow. It could be more straight-forward both in terms of its description and implementation. If one thinks of the function as a piece-wise linear trend defined by

$$y = a_1 x + b_1, \quad x < x_b \qquad (1)$$

and

$$y = a_2 x + b_2, \quad x > x_b \qquad (2)$$

where xb is the breakpoint year, continuity of y at xb can be ensured by letting b2= (a1−a2)xb+b1. Substituting b2 into the equation yields 4 parameters to be estimated (a1, b1, a2,xb), which can be done using the 'curve_fit' function. This avoids having to also use both the 'curve_fit' and 'lin_regress' function, as the parameters found by 'curve_fit' can be used to describe the two lines completely. Regardless of which method is used, the equation(s) used in 'curve_fit' should be shown in the paper.

The procedure outlined by the reviewer essentially captures the procedure we used. The "curve_fit" function is defined in lines 504-794 of the file https://github.com/scipy/scipy/blob/master/scipy/optimize/minpack.py, which we have added to the description of our methodology after L285. This function performs a least-squares fit to the function by modifying the function's parameters. A starting "guess" of the function parameters is provided by the user. The linear algebra methods of the scipy numerical library is then used. The slopes and break points that emerge will be the same as those obtained by the use of (1) and (2). .

• L227-229; This was just stated on L207-208.

There is a slightly different message in the two statements, with the first (previous Lines 207-208) referring to single time series while the second (previous Lines 227-229) refers the across-region comparison of the break points. The revision clarifies this distinction by changing the text (L304-306) to: "Because the break-points are computed separately for each region, the use of the two-piece linear fit allows comparisons of the timing of the break-points across the various subregions".

• L239ff; This is an interesting result, but is it necessary to include the break-point year for all calendar months from January-September in Figure 5? The piece-wise equation seems like a sensible approach to determine a breakpoint year in the summer months (except maybe in the central Arctic). However, for winter and spring months is it really the case that a two-piece linear trend yields a better mean squared error than a simple linear trend? If not, then the breakpoint year for those months won't carry any significant meaning other than that it's an emergent parameter you get from fitting the two-piece linear trend. Is that perhaps what is being seen in Fig. 5 for years before the 1990's and after the early 2000's? Breaking down the information in Fig. 5 for each month and region would make the results easier to interpret and more informative.

The reviewer raises an important point concerning the seasonality of the break-points. In response, we have expanded this section (L347-365) by constructing histograms of break-points for the three antecedent seasons: Jan-Mar, Apr-Jun, and Jul-Sep. The revised Figure 6 (formerly Figure 5) now contains four panels, including one for each of the three antecedent seasons. (We limited the break-points included in the plots to those for which the two-piece fit reduced the variance about the trend lines by at least 5% relative to a single linear trend line). The result is that the clustering of break-points in the 1990s is more apparent in summer and winter than in spring, although the number of plots satisfying the variance-reduction criterion is greater in the summer than in the winter.

• L250; Is the reason for showing Figure 6 not also to see what regions contribute
the most explained variance to pan-Arctic SIE?

Yes. The original line 250 is revised (L366-367) to "In order to illustrate the effect of detrending and to show which regions contribute the most explained variance to pan-Arctic sea ice extent, Figure 7 shows…".

• L256-260; Why say the significance level for this specific autocorrelation value?
Is it the maximum autocorrelation value for all of the samples considered?

We are not sure which "specific autocorrelation" value the reviewer is referring to, as lines 255-256 provided the ranges of our correlations. The confusion may arise from the significance thresholds on lines 256-258, for which we also provide a range that encompasses the corresponding range of autocorrelations (from 0 to 0.40). We have inserted a statement (L375-376) that "None of the regional or pan-Arctic ice extent correlations exceeded 0.40."

• L260-262; This is the significance value when the sample has an autocorrelation of zero. Do
the detrended time series have no autocorrelation? Also, this statement is true for most regions,
but not all according to Fig. 6.

Figure 7 (formerly Figure 6) shows cross-correlations between regional and pan-Arctic ice extent, not autocorrelations. The detrended time series have very small autocorrelations (0.15 or less), which results in effective sample sizes greater than 50 (following Bretherton, 1999), in which case the 95% significance threshold increases only from 0.26 to 0.28 – with no changes to our conclusions about statistical significance.

• L263-265; This is true when the trend is included, but not when the time series
are detrended according to Fig 6. Please clarify which is being referred to here.
Can the authors speculate at all why it is the East Siberian Sea and Laptev Sea
explain the most synchronous interannual variance in pan-Arctic SIE? Is it the
region that has the most interannual variability in September?

Yes, this statement needed to be revised for clarification, as noted by the reviewer. The new text (L381-388) reads as follows: "According to Figure 7, the regions contributing most strongly to September pan-Arctic sea ice variations (including trends) are the Beaufort, Chukchi and East Siberian Seas. After the data are detrended, the regions contributing most to September pan-Arctic sea ice variations are the East Siberian and Laptev Seas". The somewhat surprisingly large contribution of the Laptev Sea is consistent with the "dynamical preconditioning" hypothesis of Williams et al. (2016), which we cite in the revision. The variances of the detrended September extents of East Siberian and Laptev Seas are indeed among the largest of all the regions, although the Chukchi Sea's interannual variance is essentially as large".

• L266-275; Is Figure 7 really necessary? It seems reasonable to consider the
contributions of variability in different regions to variability in pan-Arctic SIE, but
is there any physical basis for thinking that regions far away from the Beaufort

Sea would show any explained variance in the BSI, apart from the trend? It's worthwhile to know how far from the Beaufort Sea variations matter, but those regions are shown in Fig. 9 already for lag 0.

While we are open to removing material (see following comment), we would prefer to retain this figure (now Figure 8) if possible because it shows that regions of significant explained variance include the Canadian Archipelago to the east, as well as the Chukchi Sea to the west. There is a difference in the "scale of influence" in Figures 7 and 8 that is worth noting, and we have added a statement to that effect (L397-400).

• Are tables 1-4 necessary?

No, the tables are not necessary except for supporting the finding that a larger fraction of September pan-Arctic variance is explained by antecedent pan-Arctic extent than by antecedent regional extent (see comment below on Figure 9, formerly Figure 8). Assuming the journal allows Supplementary Material, we have relegated Tables 3-4 to Supplementary Material (Tables S1 and S2); otherwise we will omit Tables 3 and 4, but would prefer to retain Tables 1 and 2 to support the above finding.

For Tables 1 and 3, when the full time series are compared (non-detrended), the fact that there will be correlation between the different regions and the 2 predictands when both predictor and predictand contain trends is not really a surprise, is it? For Tables 2 and 4, the only relevant information to prediction of the BSI is limited to a few (expectantly nearby) regions, and for the most part to a short number of lag months. For those areas and lags where there is any explained variance over 10% , that information is already plotted in Figs. 8 and 9. Why not just say when describing Figs 8 and 9 that no explained variance values greater than 10% were found in other regions?

It is not surprising that the trends inflate the correlations, although the extent to which the correlations decrease with detrending might not have been anticipated without actually evaluating the cross-correlations. There is indeed a limited spatial scale of coherence in the BSI results in Tables 2 and 4 (now Table S2). We have followed the reviewer's suggestion and shorten the text by deleting Lines 314-320 and replacing it with "The BSI variance explained by all other regions is less than 10%" (L447-448)

If an argument to keep the tables is made, the shading scheme should be explained.

The two levels of shading simply denote explained variances that exceed 10% , 20%, 30%,… This is now explicitly stated in the captions of the tables.

• L284ff and Figure 8; I'm surprised to see the values of 0.41 and 0.05 for the Beaufort Sea in Fig. 8 (and identical values for the Canadian Archipelago in the tables 1, 2) for lag months January-April. How can this be if sea ice extent is not variable in the Beaufort Sea and the Canadian Archipelago in the winter?

These values were actually misleading, and they were consequences of very small digitization artifacts. In one of the earlier sources of the pre-satellite era, the winter sea ice concentrations were digitized with a slightly different land mask. This biased digitization impacted both the regional (Beaufort) and Canadian Archipelago) as well as the pan-Arctic ice extents. We have revised the figures (Figs. 9 and 10) to show correlations based on corrected values. The revised correlations are zero where they should be.

• Figure 8; I think it makes sense to also show the autocorrelation for September pan-Arctic SIE (recognizing it is in Tables 1 and 2; see comment above). It would put the contributions of these other regions into context for prediction purposes. According to Table 2, Pan-Arctic SIE contributes at different lags contributes more to September SIE than any individual region. Should that result itself be mentioned as the most significant (albeit arguably expected) result in terms of prediction? This result should be compared against the studies mentioned in the general comments, but it seems to agree with them. Also, the lagged Laptev Sea SIE contributes more to September Pan-Arctic SIE than the Beaufort Sea, so why not show it too?

As suggested, we have added pan-Arctic and Laptev Sea panels to Figure 9. The revised text (L426-433) also highlights the result that the pan-Arctic extent of July and August indeed correlates more highly than any regional extent with September pan-Arctic ice extent in both the non-detrended and the detrended data (Tables 1 and 2). This is indeed the rationale for retaining Tables 1 and 2. The finding that the lagged pan-Arctic correlations exceed the lagged regional vs. pan-Arctic correlations is consistent with the perfect-model results in Bushuk et al.'s (2017) Figure 2, which we now cite, although we note that this comparison is not apples-vs.-apples: Bushuk et al. show the skill pf predictions of regional extent (not pan-Arctic extent) in their regional panels. The same is true for Day et al.'s (2014) Fig. 11 and for Bushuk et al.'s (2018) Figs. 6/9/10/11.

A physical explanation for Laptev Sea can be found in Williams et al 2016:

Williams, James, et al. "Dynamic preconditioning of the minimum September sea-ice extent." Journal of Climate 29.16 (2016): 5879-5891.

The revised text cites this paper and its "dynamical preconditioning" hypothesis (L385-388).

• L288-291; "...reaching zero by...". Near zero, not zero exactly. It's also odd that these are not zero exactly when ice covers these regions completely through April

Revision of Line 289: "...approaching zero by 3-4 months". The non-zero values are another consequence of the land-mask difference noted above (cf. response to comment on L284ff and Figure 8), and the correlations in the revision are based on corrected (temporally consistent) values of ice extent.

• L319-320; Doesn't the BSI contain information from earlier months than September? That is, it's not a strictly September prediction metric.

Yes, the BSI includes pre-September information such as the length of the navigation season in the Beaufort. In the revision, we add the following (L446-449): "This percentage of explained variance is even less than one might have anticipated, given that the BSI includes information on the length of the navigation season, which can begin well before September, i.e., as early as July in some years".

• L336-338; Specify that this is a general finding; it's certainly not true for all "ice extents" considered.

Revision of Line 336 (now L466-467): "Based on the raw (not detrended) time series, the antecedent ice extents in a substantial fraction of the Arctic regional seas provide significant predictive skill…".

• L338-340; "(and even years, given the multidecadal scales of the trends)"... While probably true, this isn't a direct conclusion/finding of this study.

Will delete the parenthetical phrase, so the revised statement (L468-469) is limited to "…the regional extents of prior seasons".

• 342-345; The study being referenced here is based on six years of hindcasts which were created by averaging over many different techniques/models, and the results are not reflective of the predictive skill shown by several other prediction studies. To put the results of this paper into context, it should be stated that numerous other studies have shown higher skill than anomaly persistence forecasts (which is essentially the method used here). See references in the introduction of Bushuk et al 2018 (reference in general comments).

We will grant that other sea ice prediction efforts have outperformed persistence and now state this explicitly (L478-480), citing Bushuk et al. (2018) and studies such as Tivy et al. (2011), Shroeder et al. (2014), Yuan et al. (2016) and Petty et al. (2017). However, in this context, perfect-model studies do not seem to fit the point being made; persistence-derived predictability is greater in perfect models than in corresponding operational forecasts, as even some of the perfect-model studies show. In this respect, the SIPN is the acid test of the current state of sea ice prediction (at least for September pan-Arctic ice extent). A compilation of SIPN results for the past eight years (the "state of the art"), completed after the 2018 sea ice minimum, shows that, on balance, the SIPN consensus forecasts outperform detrended anomaly persistence by only a small amount. While that persistence metric is based on year-to-year September variations, the SIPN forecasts for September are made in June, July and August – less than a season ahead, and on the favorable side of the springtime "prediction barrier". So, while the revised text will affirm the reviewer's point, we have added new text (L480-499) placing the improvement over persistence into context.

• L355; The influence of trend predictive skill is well recognized in the sea ice pre-
diction community, and skill results based on detrending are commonly presented
in sea ice prediction studies, so I'm not sure I see how challenge (1) follows from
this particular study.

The reviewer makes a valid point here. We have deleted the entire paragraph to which the reviewer refers.

Technical corrections

• Figure 3: I think it would be easier to see and compare magnitudes of the anomalies and trends if Fig 3 were split into two sub-panels: one for the Beaufort and one for pan-Arctic SIE. The sub-panel for pan-Arctic SIE could have double y-axis (one for March, one for September) so that there isn't such a large vertical space between lines.

Agreed. We have remade Figure 3 into two panels as suggested.

• L194-197; Sentence is worded awkwardly, specifically the "Because..., so ...".

Revision into two sentences (L257-260): "One of our main interests in this study is whether or not interannual variations of preceding regional ice extents correlate with later BSI values. In order to exclude the effect of the overall trends in the correlation of these time series, we detrend the data and explore various methods for doing so".

• Figure 4: It looks like there is a kink in the piece-wise linear functions near the breakpoint. This shouldn't be there.

Yes, we see that in the lower panel and have corrected the two-piece linear trend line. Thank you for catching the glitch.

• L255; "accrual".. what is accruing here? Why not just say "corresponding to correlations..."

That was a typo. Should have been "actual", but "corresponding" is even better. Line 372 has been corrected.

• L257 and 258; typo? "a 60-year samples" should be "a 60-year sample" (L374)

Revised by deleting "s" at end of "samples".

• L296; Change to ..."when sea ice extent" for the Beaufort and East Siberian Seas are predictors. The seas themselves aren't predictors.

You are right. We have revised to say that sea ice extents in those seas are the predictors (L420)

• L319-320; incomplete sentence

"that" (second word) was a typo. Sentence has been revised (L447-450).

• L333-335; "(increases of trends)" not needed as it is explained what is meant by break-points in the second part of the sentence.

"(increases of trends)" deleted in revision (L463).

Anonymous Referee #2

Received and published: 1 November 2018

˘"Seasonal sea ice prediction based on regional indices" is an intriguing new look
into the statistical predictability of sea ice conditions as defined by the Barnett Severity
Index (BSI). The BSI is one of the longer duration sea ice metrics on record and I com-
mend the authors for reaching back prior to the satellite era to paint a more thorough
picture of sea ice conditions. We need more work in this area.
As the authors note, this paper somewhat extends previous work by Drobot but it is not a precise
analogue. The present paper has three main objectives: (1) to quantify predictability inherent in
antecedent spatial distributions of sea ice, (2) to distinguish predictability of pan-Arctic sea ice
from that of regional predictability, and (3) to distinguish quantitatively the trend-derived
predictability and predictability of departures from trend.

What is written in the paper is very well done. I have few critical issues with this aspect
of the paper. Assessing the predictability of de-trended data is a key baseline contri-
bution for further developing statistical sea ice forecasts. However, I am left wanting
with just this analysis. In the Drobot paper, sea ice data was supplemented with atmospheric
teleconnections and other data. This paper would be greatly strengthened by adding additional
predictors so we can begin to better understand the predictability of de-trended sea ice data.

*We acknowledge the reviewer's point, which is that the addition of other (atmospheric)
predictors, the paper would provide a more comprehensive assessment of statistical
predictability. However, previous work by Lindsey et al. (2008, JGR) and Drobot et al. (2006,
GRL), has shown that atmospheric modes of variability such as the NAO, AO, and PDO "were
found to have little value as predictors of the September Arctic SIE compared with ocean and ice
predictors" (Guemas et al., 2014, QJRMS, p. 555). The recent paper by Goldstein et al. (2016,
The Cryosphere Disc.), using Self-Organizing Maps (SOMs), also found that there was little
evidence of a useful signal of the atmospheric circulation in seasonal prediction of open water
season length. Summer wind patterns play a role in interannual variations of late-summer ice
extent, but the summer wind patterns in years with similar September pan-Arctic can be quite
different (Serreze et al., 2016, JGR) and wind pattern anomalies are largely unpredictable.
Ocean anomalies do have some predictive value, especially for the North Atlantic winter ice
extent (Bushuk et al., 2017, GRL), but the record of subsurface ocean variables is much shorter
than the 60+ year record length of sea ice variations examined in our study. Given that (1) the
aggregate of this evidence points to diminishing returns if the predictor suite is expanded beyond
sea ice and (2) such an expansion of scope would, in our view, detract from the paper's
intentional focus and main "punch lines", we believe that the paper's main messages would be
diluted by the expansion to include other types of predictors.*

Finally, we note that the paper's present focus on persistence generated a host of issues requiring further discussion and clarification, as is evident in the extensive comments of Reviewer 1. While a need for clarification is not a justification for a more limited scope, some of Reviewer 1's comments are fertile ground for further discussion (see Reviewer 1's comment on lines 342-345 and response), so we believe the paper's present content can generate reader interest.

For the above reason, I recommend this paper needs major revisions. To clarify, I find little fault with what is here – it's just that I don't think it's enough. By adding additional predictors, this will become a more complete package and one that will have high visibility moving forward.

Seasonal sea ice prediction based on regional indices

John E. Walsh[1], J. Scott Stewart[2], Florence Fetterer[2]

[1]Alaska Center for Climate Assessment and Policy, University of Alaska, Fairbanks, AK 99709 USA
[2]National Snow and Ice Data Center, University of Colorado, Boulder, CO 80303 USA

Correspondence to: John E. Walsh (jewalsh@alaska.edu)

Abstract. Basic statistical metrics such as autocorrelations and across-region lag correlations of sea ice variations provide benchmarks for the assessments of forecast skill achieved by other methods such as more sophisticated statistical formulations, numerical models, and heuristic approaches. In this study we use observational data to evaluate the contribution of the trend to the skill of persistence-based statistical forecasts of monthly and seasonal ice extent on the pan-Arctic and regional scales. We focus on the Beaufort Sea where the Barnett Severity Index provides a metric of historical variations in ice conditions over the summer shipping season. The variance about the trend line differs little among various methods of detrending (piecewise linear, quadratic, cubic, exponential). Application of the piecewise linear trend calculation indicates an acceleration of the winter and summer trends during the 1990s. Persistence-based statistical forecasts of the Barnett Severity Index as well as September pan-Arctic ice extent show significant statistical skill out to several seasons when the data include the trend. However, this apparent skill largely vanishes when the data are detrended. In only a few regions does September ice extent correlate significantly with antecedent ice anomalies in the same region more than two months earlier. The springtime "predictability barrier" in regional forecasts based on persistence of ice extent anomalies is not reduced by the inclusion of several decades of pre-satellite data. No region shows significant correlation with the detrended September pan-Arctic ice extent at lead times greater than a month or two; the concurrent correlations are strongest with the East Siberian Sea. The Beaufort Sea's ice extent as far back as July explains about 20% of the variance of the Barnett Severity Index, which is primarily a September metric. The Chukchi Sea is the only other region showing a significant association with the Barnett Severity Index, although only at a lead time of a month or two.

1 Introduction

One of the most widely monitored variables in the climate system is Arctic sea ice. By any measure, Arctic sea ice has decreased over the past few decades (Box et al., 2018). September sea ice extent during the past 5-10 years has been approximately 50% of the mean for the 1979-2000 period (AMAP, 2017). The recent decline is unprecedented in the satellite record, in the period of direct observations dating back to 1850 (Walsh et al., 2016), and in paleo reconstructions spanning more than 1400 years (Kinnard *et al*., 2011). The recent reduction of sea ice has been less in winter and spring than in summer and autumn, resulting in a sea ice cover that is largely seasonal (AMAP, 2017). The increasingly seasonal ice cover contrasts with the Arctic Ocean's predominantly multiyear ice pack of the pre-2000 decades. When compared to the reductions of the spatial extent of sea ice, the percentage reductions of ice volume and thickness are even larger. Ice thickness decreased by more than 50% from 1958-1976 to 2003-2008 (Kwok and Rothrock, 2009), and the percentage of the March ice cover made up of thicker multiyear ice (ice that has survived a summer melt season) decreased from 75% in the mid-1980s to 45% in 2011 (Maslanik *et al*., 2011). Laxon et al. (2013) indicate a decrease of 64% in autumn sea ice volume from 2003-08 to 2012. The portion of the Arctic sea ice cover comprised of older thicker ice has decreased from 45% in 1985 to 21% in 2017 (NOAA, 2018).

While the loss of sea ice is generally presented in terms of pan-Arctic metrics, regional trends can be quite different from the pan-Arctic trends. The Bering Sea, for example, showed a positive trend of coverage (fewer open water days) from 1979 through 2012 (Parkinson, 2014), However, the positive trend of Bering Sea ice largely vanishes when the most recent winters (especially 2017-18) are included. By contrast, the Chukchi and Beaufort Seas to the north of the Bering Sea have shown some of the largest decreases of summer ice coverage in the entire Arctic (Onarheim et al., 2018). Another area of strong decrease of ice coverage has been the Barents/Kara Sea region.

The Beaufort Sea serves as an illustrative example of the impacts of trends and variability of sea ice. The number of open water days immediately offshore of the Beaufort coast has been 60-120 in recent years. Parkinson's (2014) Figure 2 shows that the number of open water days increased by 20-30 days per decade over the period 1979-2013. However, as recently as the 1970s, there were summers with little or no open water in this region, as described by Crowley Maritime, one of the major barge operators in the Alaska region:

2

59   "With pipeline construction well underway in 1975, the Crowley summer sealift
60   flotilla to the North Slope faced the worst Arctic ice conditions of the century. In fleet
61   size, it was the largest sealift in the project's history with 47 vessels amassed to carry
62   154,420 tons of cargo, including 179 modules reaching as tall as nine stories and
63   weighing up to 1,300 tons each. Vessels stood by for nearly two months waiting for the
64   ice to retreat. Finally in late September the ice floe moved back and Crowley's tugs and
65   barges lined up for the slow and arduous haul to Prudhoe Bay. When the ice closed again,
66   it took as many as four tugs to push the barges, one at a time, through the ice".
67   *— From Crowley Maritime, 50 Years of Service in Alaska (2002)*

68   As will be shown, the contrast between present-day ice conditions and the Crowley
69   experience of the 1970s is largely a manifestation of the trend of Beaufort Sea ice cover.
70   However, sea ice also exhibits large year-to-year variability, which has been superimposed on
71   the recent trend towards less sea ice in the Arctic. This variability challenges users of coastal
72   waters in various sectors and lies at the heart of the sea ice prediction problem. While the
73   climatological seasonal cycle and even observed trends provide an initial expectation for the sea
74   ice conditions that will be present in a particular region at a particular time of year, the
75   departures from the climatological mean, whether or not the mean is adjusted for a trend, is
76   affected by the atmospheric forcing (winds, air temperatures, radiative fluxes) and oceanic
77   forcing (currents, water temperatures) of the particular year in addition to antecedent ice
78   conditions themselves. These departures have a large component of internal variability and hence
79   are difficult to predict over monthly and seasonal timescales (Serreze et al., 2016), raising
80   questions about the extent to which sea ice variations may be predictable. Even ice-ocean
81   models initialized to current sea ice and ocean conditions require atmospheric forcing in order to
82   predict future ocean states. Moreover, fully coupled models, which determine both the
83   atmospheric and ocean/ice conditions prognostically, are now used increasing often for seasonal
84   sea ice predictions. Ensembles of coupled simulations are generally run because of the chaotic
85   nature of the climate system. These models can be run for much longer time periods than the
86   observational sea ice record, so they can provide statistics of sea ice persistence
87   (autocorrelations) subject to the "perfect model" assumption. Examples of studies employing the
88   "perfect model" approach are Holland et al. (2011), Blanchard-Wrigglesworth et al. (2011), Day

3

et al. (2014), Bushuk et al. (2017) and Bushuk et al. (2018). In these model simulations, autocorrelation of sea ice anomalies tends to be greater in the model results than in observational data (e.g., Blanchard-Wrigglesworth et al., 2011, their Fig. 2; Day et al., 2014, their Fig. 1).

The skill of persistence-based statistical forecasts of sea ice variations beyond the mean seasonal cycle and ongoing trends is the main focus of this paper. While various prior studies (cf. Section 2) have utilized broader approaches to evaluating sea ice predictability and the skill of forecasts, the present study is limited specifically to statistical predictions of regional (and pan-Arctic) September sea ice based on auto-correlation (anomaly persistence, often referred to as "memory") and lagged cross-correlations between with other sea ice coverage quantities. Other approaches to sea ice predictability include the use of models, which can be initialized to obtain deterministic forecasts verifiable with observations or which can be run for long periods in a coupled mode to assess predictability of sea ice within the "model's world" (irrespective of observations). We also do not use atmospheric or oceanic predictors in our evaluation of persistence-based predictability. Atmospheric predictors in the form of known teleconnection patterns have been used by Drobot (2003) and Lindsay et al. (2008), while Bushuk et al. (2017) have shown that ocean temperature initialization contributes to skill of seasonal forecasts of sea ice in the North Atlantic subarctic seas. A review of the various approaches to sea ice prediction and sources of predictability has been provided by Guemas et al. (2016).

The present paper extends the temporal window of Drobot's (2003) study of the predictability of Beaufort-Chukchi sea ice. Drobot used data from 1979-2000 to assess predictability of a measure of Beaufort Sea summer ice severity (Section 3 below) based on antecedent sea ice conditions as well as several atmospheric indices. While the present study will not include the type of multiple-predictor evaluation carried out by Drobot, it will provide a more comprehensive and updated assessment of sea ice anomaly persistence in a predictive context. Drobot (2003) found that, in predictions based on indicators from the previous seasons, the limited sample of years used in developing the statistical models raises questions about broader applicability. In this regard, Drobot (2003, p. 1161) states "…if the Arctic climate changes, the methods described here will need to be altered". In fact, the Arctic climate and, in particular, its sea ice regime, have changed with the unprecedented retreat of sea ice in the post-2000 period. The impact of the trend on statistical predictability is a focus of the present paper. Another relevant study is that of Blanchard-Wrigglesworth et al. (2011), who found evidence that

120 persistence of ocean temperature anomalies across seasons has a detectable impact on sea ice

121 variations, implying some predictability over seasonal timescales.

122 In the present paper, we use the autocorrelation statistic to quantify the skill of persistence as

123 a control forecast of pan-Arctic and regional sea ice extent. In addition to utilizing the more

124 conventional metric of ice extent in regional and pan-Arctic domains, we include a regional sea

125 ice index developed in the 1970s to capture interannual variations of marine access in the

126 Beaufort Sea. A primary focus of the evaluation is the method of detrending the data, as various

127 alternative methods have not been fully explored in the literature. We show that the piecewise

128 linear method generally results in the smallest residual variance about the trend line, and we then

129 perform an across-region synthesis of information on the break-points of the two-piece linear

130 trend lines in different seasons. Our period of analysis extends back to 1953, which results in a

131 considerably larger sample of years than the more commonly used satellite period (1979

132 onward). Finally, we examine lagged cross-correlations to determine whether pan-Arctic ice

133 extent or Beaufort Sea summer ice conditions are foreshadowed in a statistical sense by

134 antecedent ice conditions in particular subregions of the Arctic.

135 More generally, the results presented here can serve to provide a baseline for distinguishing

136 contributions to seasonal sea ice forecast skill arising from climatological sea ice coverage, sea

137 ice persistence, and sea ice trend. This baseline can, in turn, serve as benchmarks for measuring

138 improvements achieved by more sophisticated prediction approaches such as dynamical models,

139 analog systems, neural networks and other more comprehensive statistical methods. The Sea Ice

140 Outlook, coordinated by the Sea Ice Prediction Network now in its Phase 2

141 (https://www.arcus.org/sipn/sea-ice-outlook, accessed 27 Dec 2018)), provides an annual

142 compilation of seasonal sea ice forecasts, which are grouped into three categories:

143 physical/dynamical models, statistical methods, and heuristic approaches. While the

144 methodology used in this paper falls into the statistical category, the distinctions between (a)

145 pan-Arctic and regional skill and (b) trend-derived and interannual forecast skill are relevant to

146 all three approaches to sea ice prediction.

147

148 **2. Previous work**

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

149   The baseline for persistence-based predictions have been established in previous studies
150   (e.g., Blanchard-Wrigglesworth et al., 2011; Day et al., 2014; Bushuk et al., 2017, 2018). These
151   studies have generally focussed on the post-1979 period of satellite data, while the present study
152   a longer record length (back to 1953 rather than 1979). The main intent of the paper is to show
153   how detrending is a key step in the depiction of persistence-based statistical predictions. We
154   illustrate the effect of detrending for both pan-Arctic ice extent and regional metrics in order to
155   show that predictive applications on both scales must address detrrending in a rigorous way, and
156   that there are various alternatives for detrending. While these alternative detrending strategies
157   are known, the relative effectiveness of the various alternatives has not been addressed in
158   previous studies. Goldstein et al. (2016; 2018) come closest by comparing representations based
159   on linear trends and discontinuities in the mean. An additional novel outcome of the present
160   study is the synthesis of break-point information.
161   The extension back to 1953 is especially noteworthy because of the recent reduction of
162   Arctic sea ice coverage has occurred almost entirely in the post-1978 period of satellite coverage.
163   On both pan-Arctic and regional scales, ice extent was relatively stable during the 1950s, 1960s
164   and 1970s, although interannual variations were still a prominent feature of the time series
165   (Walsh et al., 2016). While Drobot (2003) and Lindsay et al. (2008) made use of sea ice data
166   extending back to the 1950s, there has been no systematic comparison of sea ice anomaly
167   persistence during the satellite era with anomaly persistence over longer time periods.

3. Metrics of sea ice coverage

169   Historical variations of sea ice are documented using various metrics, including sea ice
170   extent, ice-covered area, and thickness. Sea ice extent is the total area within the ice edge, which
171   is typically taken to be the 15% contour of sea ice concentration. Ice extent is readily obtainable
172   from satellite measurements, as is the actual ice-covered area if the open water within the ice
173   edge is accurately depicted. Surface-based observations from ships or coastal locations typically
174   capture only the ice edge and are therefore useful primarily in the mapping of ice extent. While
175   digitized records of ice extent exist back to the 1800s, there are no such historical products for
176   ice thickness. *In situ* measurements of ice thickness are sparse in space and time, as are
177   submarine sonar measurements, which are not only sparse but often remain unavailable.
178   Satellite-derived estimates of ice thickness are subject to considerable uncertainty and have only

6

179 recently come into use (e.g., CryoSat), while dynamic-thermodynamic model-based

180 reconstructions of historical sea ice thickness variations have only recently been attempted

181 (Schweiger et al., 2018).

182   To explore the statistical skill that may be inherent in the spatial distribution of sea ice, we

183 compute ice extent using the gridded Arctic-wide sea ice concentration product known as

184 "Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward (Walsh et al., 2015), referred

185 to in the National Snow and Ice Data Center (NSIDC) catalog as G10010. This dataset is based

186 on observations from approximately 15 historical sources between 1850 and 1978: the earliest

187 are whaling records, and the most complete, in terms of coverage, are the Arctic-wide analyses

188 that the U.S. National Ice Center (NIC) began in the early 1970s. Beginning in 1979, sea ice

189 concentrations from passive microwave data are used exclusively in G10010. Ice concentration

190 fields on the 15th of each month are taken from the *NOAA/NSIDC Climate Data Record of*

191 *Passive Microwave Sea Ice Concentration, Version 2 (Meier et al., 2013)*.

192   Prior to the 1950s, most observations were from near or just within the ice edge. If only the

193 ice edge position was known, a gradient of ice concentration within the edge was imposed in

194 order to integrate the observations into G10010. The gradient was based on a climatology

195 constructed from the passive microwave data. Spatial and temporal gaps in observations were

196 filled using an analog technique that is described in the data product documentation. Each

197 month's sea ice concentration field in G10010 is an estimate of conditions at one time in the

198 month, nominally the 15th day of the month (or as close to the 15th as data were available). The

199 fields are at quarter-degree resolution. From these fields one can derive monthly sea ice extent

200 values. Sea ice extent is computed as the area, in sq km, covered by all cells that contain ice in

201 any concentration greater than 15%. Sea ice extent is always greater than or equal to the actual

202 ice-covered area, which excludes the area of open water within the main ice pack.

203   Various studies (e.g. Partington et al., 2003; Agnew and Howell 2002) have shown that

204 passive microwave-derived sea ice data tend to underestimate ice concentration when compared

205 with operational analyses. The *Climate Data Record of Passive Microwave Sea Ice*

206 *Concentration* is a blend of output from two algorithms that results in higher ice concentrations

207 overall for a better match with the operational analyses that predate the satellite record. Even

208 so, one might expect to see a discontinuity in the G10010 time series of ice extent when the

209 passive microwave record starts in 1979, but this is not evident (see Fig. 10 in Walsh et al.,

Formatted: Font: Not Italic

210 2016). While G10010 gives a record of ice extent that has realistic variability back to 1850, it is

211 difficult to assign an uncertainty to the concentration fields and ice extent values derived from

212 them. Ice extent will be more accurate than actual ice-covered area because there are many more

213 observations of the ice edge than of the concentrations within interior pack. For this reason, we

214 base our analysis on ice extent. It should be noted, however, that persistence time-scales of pan-

215 Arctic sea ice area have been shown in previous studies (e.g., Blanchard-Wrigglesworth et al.,

216 2011) to be longer than those of pan-Arctic sea ice extent because high-frequency forcing can

217 change ice extent more than it changes ice area (i.e., by converging or diverging ice floes in the

218 absence of ridging or melt).

219   G10010 was used to compute the time series of monthly sea ice extent for the pan-Arctic

220 domain and various Arctic subregions in which sea ice is at least a seasonal feature. The

221 regionalization adopted here follows that of the MASIE (Multisensor Analyzed Sea Ice Extent)

222 product available from the National Snow and Ice Data Center

223 (http://nsidc.org/data/masie/browse_regions, accessed 27 Dec 2018). MASIE (NIC and NSIDC,

224 2010) is produced in cooperation with the NIC, and its regions are defined on the basis of NIC

225 operational analyses areas. We use the following MASIE regions: (1) Beaufort Sea, (2) Chukchi

226 Sea, (3) East Siberian Sea, (4) Laptev Sea, (5) Kara Sea, (6) Barents Sea, (7) East Greenland Sea,

227 (8) Baffin Bay/Davis Strait, (9) Canadian Archipelago, (10) Hudson Bay, (11) central Arctic

228 Ocean and (12) Bering Sea. There are several other MASIE regions (Baltic Sea, Yellow Sea,

229 Cook Inlet) that are not used here because they are not geographically connected with the main

230 Arctic sea ice cover. Figure 1 shows the regions.

231   We also make use of the long ice extent record provided by G10010 to investigate the extent

232 to which the Barnett Severity Index, or BSI, may be statistically predictable from antecedent ice

233 extent. The BSI is directly relevant to offshore navigation applications in the Beaufort Sea. It is

234 a metric of the severity of ice conditions, such as conditions encountered by barges resupplying

235 the North Slope. The BSI is determined once per year, at the end of the summer shipping season,

236 by analysts at the NIC. It is a unit-less linear combination of five parameters: 1) the distance in

237 nautical miles from Point Barrow northward to the ice edge on 15 September, 2) the distance

238 from Point Barrow northward to the 4/8th ice concentration line on 15 September, 3) the number

239 of days the entire sea route from the Bering Strait to Prudhoe Bay is ice-free in a calendar year,

240 4) the number of days the entire sea route to Prudhoe Bay is less than or equal to 4/8th ice

241 concentration in a calendar year, and 5) the temporal length of the navigable season, defined as

242 the time period from the initial date the entire sea route is less than 4/8th ice concentration to 1

243 October (Barnett, 1980). Figure 2 is a time series of the BSI reconstructed from gridded sea ice

244 concentration data (see Appendix). Higher values indicate less severe ice conditions.

4 Methods

246 As shown in Figure 3, Arctic sea ice extents have generally been decreasing over the post-

247 1953 period of this study. The Beaufort Sea is a prime example of a region in which summer

248 and autumn sea ice coverage has been decreasing, although winter (March) sea ice extent in the

249 Beaufort Sea shows no trend or variability because the ice edge extends to the coastline in March

250 of every year, essentially eliminating year-to-year variations. Consistent with the September

251 decrease of Beaufort ice extent, the BSI has been increasing over the past few decades (Figure

252 2). Two time series containing trends over time can show a correlation simply because the trends

253 are present in the time series. A trend can be used as a predictive tool by assuming its

254 continuation into the future. However, a trend can inflate persistence-based forecast skill when a

255 variable is used to predict itself (assuming the historical trend continues into the future). Indeed,

256 depictions of time-variations of a quantity such as sea ice extent are often shown as departures

257 from a trend line in order to highlight the interannual variations. One of our main interests in this

258 study is whether or not interannual variations of preceding regional ice extents correlate with

259 later BSI values. In order to exclude the effect of the overall trends in the correlation of these

260 time series, we detrend the data and explore various methods for doing so.

261 The choice of a function with which to de-trend the time series should be determined by

262 features of the series itself. The detrended time series should exclude the general tendency to

263 change over time, but preserve a measure of the year-to-year variability of the series. The

264 previous studies cited in Section 2 (e.g., Blanchard-Wrigglesworth et al., 2011; Sigmund et al.,

265 2013; Day et al., 2014; Bushuk et al., 2017, 2017) have generally relied on least-squares linear

266 fits for detrending. Goldstein et al. (2016, 2018), by contrast, showed that discontinuous changes

267 in the mean better captured time series (such as open water area) characterized by abrupt

268 changes. In the spirit of the Goldstein et al. studies, we explore various options for detrending a

269 time series such as those in Figures 2 and 3, for which the changes are more pronounced in

270 recent decades than in earlier decades. In such cases, a single multi-decadal trend line cannot be

271 expected to optimally represent the historical evolution.

272 We explored several functional forms which fit the time series, including linear, quadratic,

273 cubic, and exponential functions.   We found that a simple two-piece linear function – wherein

274 the data are modeled by two line segments that intersect at a 'break-point' year – had the lowest

275 average RMS difference between the time series and the fitted function, although fits using other

276 functions had only slightly larger RMS differences. This choice of the detrending fit has the

277 additional feature of giving a sense of when the ice extent began to change more rapidly.

278 The two-piece linear fits were obtained by using standard statistical algorithms. A

279 function defined by two intersecting half-lines can be specified by the coordinates of one point

280 on each half-line and the intersection point. With the x-axis as time, and the y-axis as the value

281 of the sea ice extent, the x-values of the non-intersecting points can be chosen to be 1953 and

282 2013, the first and last years of the BSI dataset. This leaves four values for the function to fit:

283 the series value in 1953, the series value in 2013, and the year and value at the intersection point,

284 also referred to here as the break-point. We note that the break-point is not specified by the user

285 but is determined by the algorithm so that the fit to the time series is optimized. The "curve_fit"

286 function is defined in lines 504-794 of the file

287 https://github.com/scipy/scipy/blob/master/scipy/optimize/minpack.py This function performs a

288 least-squares fit to the function by modifying the function's parameters. A starting "guess" of the

289 function parameters is provided by the user. The linear algebra methods of the *scipy* numerical

290 library is then used.

291 The two-piece linear fit was generated by allowing the SciPy 'curve_fit' routine (Jones et

292 al., 2001) to iterate to a solution. The "curve_fit" function is defined in lines 504-794 of the

293 minpack.py routine (https://github.com/scipy/scipy/blob/master/scipy/optimize/minpack.py,

294 accessed 27 Dec 18). This function performs a least-squares fit to the function by modifying the

295 function's parameters. A starting "guess" of the function parameters is provided by the user. The

296 linear algebra methods of the scipy numerical library are then used to detrend the BSI values as

297 well as the time series of the regional and pan-Arctic ice extents. In Figure 4, we show the

298 piecewise linear fit together with quadratic, cubic and exponential fits to the time series of the

299 BSI and the September Beaufort Sea ice extent. In the case of the two-piece linear fit, the break-

300 point (chosen to minimize the departures from trend) is in the early 1990s for both sea ice

301 metrics. It is visually apparent from Figure 4 that all four fits are comparable in terms of the

302 overall magnitudes of the departures from the trend lines. The root-mean-square departures from

303 the various trend lines indeed differed by less than 10%. Given the small differences between the

304 fits, we choose the two-piece linear for the remainder of this study. Because the break-points are

305 computed separately for each region, the use of the two-piece linear fit allows comparisons of the

306 timing of the break-points across the various subregiona.

307 After using the 'linregress' method from the SciPy (Jones et al., 2001) software library to fit

308 a line to regional monthly extent values and the BSI, we computed correlations between the

309 departures of the two time series from their respective two-piece trend lines. For comparison, we

310 also computed correlations between the "raw" (with trends) time series. The square of the

311 Pearson correlation coefficient (R^2) was computed using the 'stats' method from the SciPy

312 package and was used to determine whether and how strongly the two time-series are correlated

313 with each other.

314

315 ## 5 Results

316 As noted in Section 2, previous studies (e.g., Bushuk et al., 2018) have evaluated the

317 persistence of regional ice extent over the post-1978 period of satellite observations. Here we

318 extend this evaluation to encompass a longer period dating back to 1953 in order to assess the

319 stability of the persistence statistics. Specifically, for each region in Figure 1, we have correlated

320 the September ice extent with the ice extent of antecedent months for the 1953-2013 and 1979-

321 2013 periods. Figure 5 compares these persistence values (autocorrelations at multimonth lags),

322 for the antecedent months of March, May and July in a subset of regions. Because the regions

323 chosen were those that have interannually varying ice cover in September, regions such as the

324 Bering Sea, Hudson Bay, the Sea of Okhotsk and the Baltic Sea were excluded. The correlations

325 were computed before and after a detrending of the data, although only the results for the non-

326 detrended data are shown in Figure 5. For most of the regions, the inclusion of the earlier

327 decades does not have a notable impact on the persistence from July to September. However, the

328 March-to-September and May-to-September correlations change substantially in a few regions.

329 The Baffin Bay March-to-September correlations increase from 0.00 to 0.34 when the earlier

330 decades are eliminated, largely as a result of the post-1979 trend: the post-1979 correlation is

331 statistically significant ($p < 0.05$), while the corresponding correlation based on detrended data is

11

not significant. The pan-Arctic correlations for all three antecedent months also increase when the earlier decades are eliminated. In the Greenland Sea, the correlations from March and May decrease substantially and lose statistical significance when the earlier decades are eliminated. In this case the March-to-September and May-to-September correlations are again reduced to insignificance by detrending. Although the results for the detrended data are not shown graphically, the detrending generally reduces the significance of the correlations between September and the earlier months, both for the longer post-1953 periods and the shorter post-1979 periods: The March-to-September correlations based on the detrended data for the longer/shorter periods are: -0.05/0.20 for Baffin Bay, 0.20/0.13 for the Barents Sea, 0.00/0.00 for the Beaufort Sea (no March variance), 0..00/0/00 for the Canadian Archipelago (no March variance), -0.15/0.00 for the Chukchi Sea, 0.07/0.21 for the East Siberian Sea, 0.25/-0.03 for the Greenland Sea, 0.03/0.03 for the Kara Sea, and 0.07/0.18 for the Laptev Sea. The corresponding 5% significant levels are 0.26/0.33. Evidently, the springtime "predictability barrier" (Lindsay et al., 2008; Day et al., 2014; Bushuk et al., 2018) in regional forecasts based on persistence of ice extent anomalies is not reduced by the inclusion of several decades of pre-satellite data.

Because changes of trend have not been addressed systematically in previous evaluations of Arctic sea ice trends, we synthesized the break-point information across all regions and calendar months (January-September) included in our study. The synthesis was limited to only those regions and calendar months in which the two-piece linear fit reduced the root-mean-square residual by at least 5% relative to the one-piece linear best fit. Figure 6 groups the break-points into five year periods ending in 1955, 1960,…, 2015. In order to capture the seasonality of the break-points, we present separate plots for (a) the entire January-September period, (b) January-March (winter), (c) April-June (spring), and July-September (summer). As shown in panel ((a), nearly all the break-points occur in the second half of the study period, with a maximum in 1991-1995. The 1991-1995 period has the most break points of any 5-year period, and the 1990s have nearly as many break points as all the other decades combined. The winter and summer seasons are the primary contributors to the maximum in the 1990s, as the spring break points are evenly distributed through the latter half of the study period. However, spring has the fewest (12) break-points overall, while the summer has the most (26). The break-points for our focal metrics, the BSI and September pan-Arctic ice extent, are 1991 and 1996, respectively, consistent with the distribution in Figure 6. These two metrics are included in the results summarized in Figure

6. One may conclude that the 1990s, and to a lesser early 2000s, represent the shift to a more rapid rate of sea ice loss. If one is to argue for a "regime shift" in Arctic sea ice loss (Lenton, 2012), this period would be the leading candidate.

In order to illustrate the effect of the detrending and to show which regions contribute the most explained variance to pan-Arctic sea ice extent, Figure 7 shows the squares of the correlations (R^2) between September pan-Arctic ice extent and the concurrent ice extent in each of the subregions. The R^2 metric is used rather than R because R^2 corresponds to the explained variance. The figure shows values of R^2 before detrending (upper numbers, regular font) and after detrending (lower numbers, bold font). With the trend included, the R^2 values are relatively high in most regions (except for the Bering Sea), ranging from 0.32 to 0.71; the corresponding correlations (R) range from 0.57 to 0.84. These correlations all exceed the 95% significance thresholds, which range from 0.26 ($R^2 = 0.07$) for a 60-year sample with no autocorrelation to 0.38 ($R^2 = 0.14$) for a 60-year sample with an autocorrelation of 0.4. None of the regional or pan-Arctic ice extent autocorrelations exceeded 0.40. Because these correlations are dominated by the trend, the larger values appear in the regions with trends that are most similar to the pan-Arctic trend. When the data are detrended, the correlations are much smaller (R^2 values in bold font in Figure 7) although still larger than the 95% significance thresholds for a 60-year sample (R = 0.26, $R^2 = 0.07$). These smaller values indicate the relative contributions of regional variations to the interannual variations of pan-Arctic ice extent. According to Figure 7, the regions contributing most strongly to September pan-Arctic sea ice variations (including trends) are the Beaufort, Chukchi and East Siberian Seas. After the data are detrended, the regions contributing most to September pan-Arctic sea ice variations are the East Siberian and Laptev Seas. The somewhat surprisingly large contribution of the Laptev Sea is consistent with the "dynamical preconditioning" hypothesis of Williams et al. (2016). The variances of the detrended September extents of East Siberian and Laptev Seas are indeed among the largest of all the regions, although the Chukchi Sea's interannual variance is essentially as large.

Figure 8 shows the squares of the correlations between the annual BSI and regional September ice extent before the detrending of both variables (top numbers) and after detrending (bottom numbers). While the actual correlations between the BSI and regional extent are generally negative, the R^2 values plotted in Figure 8 are positive. Large values of R^2 appear in most regions when the trend is included (upper numbers) because the BSI has a strong positive

13

394 trend over time while September ice extent in most regions has a negative trend. The R^2 values

395 are much weaker in regions away from the Beaufort Sea when the trends are removed (lower

396 numbers in Fig. 8). The detrended R^2 values show the spatial representativeness of the BSI as a

397 measure of interannual variations. Figure 8 shows that the regions of significant explained

398 variance include the Canadian Archipelago to the east as well as the Chukchi Sea to the west.

399 However, the "scale of influence", if measured by the area of significant correlation, is smaller

400 for the BSI in Fig. 8 than for pan-Arctic ice extent in Fig. 7.

401   Because the potential for seasonal predictions is a key motivation for this study, we

402 examine cross-correlations in which the predictands (pan-Arctic ice extent and the BSI) lag

403 potential predictors (regional ice extents) by intervals ranging from zero (no lag) to several

404 seasons. Cross-correlations between non-detrended and detrended September pan-Arctic and

405 regional ice extents are summarized in Tables 1 and 2 respectively. Cross-correlations between

406 non-detrended and detrended BSI and regional ice extent are given in Tables S1 and S2

407 respectively. In all cases, the numerical values are the R^2 values. In order to illustrate the

408 contribution of the trend to the apparent forecast skill, we present these correlations graphically

409 for the regions which show the strongest associations with the September predictands. Figure 9

410 shows the R^2 values for cases in which September pan-Arctic ice extent lags by 0, 1, 2,…,8

411 months the ice extent in four subregions: the Beaufort, Chukchi, East Siberian and Barents Seas.

412 The red bars correspond to correlations computed from the data with the trends included. Not

413 surprisingly, the R^2 values are largest at zero lag. The rates at which the correlations decrease

414 with increasing lag vary regionally, reaching zero by 3-4 months for the Beaufort, Chukchi, and

415 East Siberian Seas. The zero-month lag values are quite large for the Beaufort, Chukchi, and East

416 Siberian regions, where they exceed $R^2 = 0.7$ (R = 0.84).

417   However, after detrending (using the two-piece linear best fits), most of the apparent forecast

418 skill is lost. As shown by the blue bars in Figure 9, nearly all the predictability from the Barents

419 and Chukchi Seas vanishes with the detrending, while only small fractions of explained variance

420 remain at non-zero lags when sea ice extents for the Beaufort and East Siberian Seas are the

421 predictors. For example, when the regional extent leads by two months (July), the fractions of

422 explained variance are approximately 0.16 and 0.10 (R ~ 0.40 and 0.32) for the East Siberian and

423 Beaufort Seas, respectively. The implication is that the persistence of interannual variations about

424 the trend line makes only small contributions to interannual variations of pan-Arctic sea ice extent,

14

and that these small contributions result mainly from the Pacific sector of the Arctic. As indicated

by Figure 9, the pan-Arctic extent of July and August correlates more highly than any regional

extent with September pan-Arctic ice extent in both the non-detrended and the detrended data (see

also Tables 1 and 2). The finding that the lagged pan-Arctic correlations exceed the lagged regional

vs. pan-Arctic correlations is consistent with the perfect-model results in Bushuk et al.'s (2017)

Figure 2, although this comparison is not apples-vs.-apples: Bushuk et al. show the skill of

predictions of regional extent (not pan-Arctic extent) in their regional panels. The same is true for

Day et al.'s (2014) Fig. 11 and for Bushuk et al.'s (2018) Figs. 6, 9, 10 and 11.

The lagged R^2 values relevant to predictions of the Barnett Severity Index are shown in

Figure 10. Because the BSI is based primarily on ice conditions in the Beaufort Sea in August

and September, it is not surprising that the correlation is largest for the Beaufort's ice extent in

September, when the R^2 value is approximately 0.8 for data that are not detrended. The August

and September values for the Chukchi are essentially as large as the corresponding Beaufort

values, indicating a spatial coherence of the variations (with trends included) in the two regions.

The antecedent extents in the East Siberian and Barents regions also explain significant fractions

of the variance when the trends are included.

The blue bars in Figure 10 are the lagged R^2 values based on the detrended data. Because the

trend's contribution to the forecast skill has been removed, these correlations provide the most

meaningful assessment of the seasonal forecast skill if the BSI based on antecedent ice

conditions. The largest correlations are for the Beaufort Sea, where the explained variances

decrease from about 0.55 (R ~ 0.74) in September to about 0.10 (R ~ 0.32) in June. The

correlations for the Chukchi are only slightly smaller, but the BSI variance explained by all other

regions is less than 10%. The percentage of explained variance is less than one might have

anticipated, given that the BSI includes information on the length of the navigation season,

which can begin well before September, i.e., as early as July in some years.

6 Conclusion

The substantial decrease of Arctic sea ice over the past several decades is well documented

(Cavalieri and Parkinson, 2012; Parkinson, 2014; Onarheim et al., 2018). Of all the regions

considered here, only the Bering Sea does not show a negative trend (Onarheim et al., 2018, their

Table 1), although the extreme minima of Bering Sea ice during the past two winters (2016-17

15

and 2017-18) are starting to bring the Bering's trend into alignment with the other regions of the Arctic.

The prominence of the trends in the time series of regional as well as pan-Arctic ice extent makes it important to distinguish the contribution of the trend from other sources of forecast skill. In this study we explored the use of several methods of detrending in order to evaluate the use of ice anomaly persistence (autocorrelation) and regional cross-correlations as predictors of ice variations. The two-piece linear trend evaluations generally have break-points in the 1990s, indicating that the rate of ice loss has been greater in the past two decades than in the earlier portion of the satellite era that began in 1979.

Based on the raw (not detrended) time series, the antecedent ice extents in a substantial fraction of the Arctic regional seas provide significant predictive skill for September pan-Arctic ice extent as well as for the Barnett Severity Index, which is more specific to the Beaufort Sea. Significant portions of variance of both September metrics are explained by the regional ice extents of prior seasons. However, this predictive "skill" is attributable primarily to the trends in the data. Removal of the trend leaves little forecast skill beyond a month or two when the forecast method is limited to the relatively simple statistical correlations utilized here. The low skill for the detrended September pan-Arctic ice extent is consistent with the findings of Stroeve et al. (2014) based on the Sea Ice Outlook as part of the Study of Environmental Arctic Change (SEARCH). Moreover, our inclusion of data back to the early 1950s shows that springtime "predictability barrier" in regional forecasts based on persistence of ice extent anomalies is not reduced by the inclusion of several decades of pre-satellite data.

It must be noted that other sea ice prediction approaches have outperformed persistence (e.g., Tivy et al. 2007; Shröder et al., 2014; Yuan et al., 2016; Petty et al., 2017; and Bushuk et al., 2018). These studies have either used other predictors or made use of the perfect model approach. With regard to the latter, persistence-derived predictability is greater in perfect models than in corresponding operational forecasts, as even some of the perfect-model studies show (Blanchard-Wrigglesworth et al., 2011; Bushuk et al., 2018). With regard to the former, the SIPN is the acid test of the current state of sea ice prediction (at least for September pan-Arctic ice extent) because many contributions utilized predictors other than persistence. A compilation of SIPN results from 2008-2018 shows that, on balance, the SIPN consensus forecasts outperform detrended anomaly persistence by only a small amount. (In this case, persistence was evaluated

16

from the yearly September mean ice extents in the National Snow and Ice Data Center's G02135_v3.0: ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/seaice_analysis/). The mean absolute error of the median SIPN forecasts issued in July of 2008-2018 is 0.32 million km^2, while the error of a forecast of persistence of the previous September's deviation from the trend line is 0.37 million km^2. Simple persistence of the previous year's actual value has an error of 0.40 million km^2. The corresponding root-mean-square errors are 0.57, 0.68 and 0.67 million km^2. While those persistence metrics are based on year-to-year September variations, the SIPN forecasts for September are made in June, July and August -- less than a season prior to September, and on the favorable side of the springtime "prediction barrier". At least in this particular application, which represents the state of the art in seasonal sea ice forecasting, sea ice anomaly persistence is a challenging control forecast and may even be regarded as a respectable competitor.

While there is statistical significance in the trend-derived skill at lead times of several seasons and also in the remaining (detrended) skill at lead times of a month or two, statistical significance does not equate to usefulness. Potential users of sea ice forecasts include local communities engaging in offshore subsistence and travel activities, marine transport companies, offshore resource extraction, and the tourism industry. The relatively small fractions of variance predictable several months in advance using detrended data (Figures 6-9) will likely leave uncertainties that are too great for many users. However the trend-derived skill, which can represent 50% or more of the variance, may enable decisions if the interannual variations superimposed on the trend represent acceptable risks for users of sea ice forecasts.

Acknowledgments

References

Agnew, T. A., and Howell, S.: Comparison of digitized Canadian ice charts and passive microwave sea-ice concentrations. Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International 1: 231- 233. doi: 10.1109/IGARSS.2002.1024996, 2002.

AMAP: Snow, Water, Ice and Permafrost in the Arctic: 2017 Update. Arctic Monitoring and Assessment Programme, Oslo, Norway, xiv + 269 pp., 2017.

Barnett, D. G.: A long-range ice forecasting method for the north coast of Alaska. Sea Ice Processes and Models, R. Pritchard, Ed., University of Washington Press, 402–409, 1980.

Blanchard-Wrigglesworth, E., Armour, K. C., and Bitz, C. M.: Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. J. Climate, 24, 231-250, 2011.

Box, J. E., and 19 Coauthors: Key indicators of Arctic climate change, 1971-2017. Environmental Research Letters, in press.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., and Yang, X.: Skillful regional predictions of Arctic sea ice on seasonal timescales. Geophysical Research Letters, 44, 4953-4964, doi:10.1002/2017/GL073155, 2017.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., and Gudgel, R.: Regional Arctic sea-ice prediction: potential versus operational seasonal forecast skill. Climate Dynamics, https://doi.org/10.1007/s00382-018-4288-y, 2018.

Cavalieri, D. J., and Parkinson, C. L.: Arctic sea ice variability and trends, 1979-2010. Cryosphere, 6, 881-889, 2012.

Crowley Maritime Corporation, 2002. From http://www.crowley.com/content/download/11926/80932/version/1/file/Alaska-50-Years.pdf accessed 10 May 2018.

Day, J. J., Tietsche, S., and Hawkins, E.: Pan-Arctic and regional sea ice predictability: Initialization month dependence. J. Climate, 27, 4371-4390, 2014.

Drobot, S.: Long-range statistical forecasting of ice severity in the Baaufort-Chukchi Sea. Weather and Forecasting, 18, 1161-1176, 2003.

Drobot, S. D., J. A. Maslanik, and Fowler, C. F.: A long-range forecast of Arctic summer sea0ice minimum extent. Geophysical Research Letters, 33, L10501, doi:10.1029/2006/GL026216, 2006.

Goldstein, M. A., Lynch, A. H., Arbetter, T. E., and Fetterer, F.: Abrupt transitions in Arctic open water area. The Cryosphere Discussion, doi:10.5194/tc-2016-108, 2016.

Goldstein, M. A., Lynch, A. H., Zsom, A., Arbetter, T., Chang, A., and Fetterer, F.: The step-like evolution of Arctic open water. Nature Scientific Reports, 8:16902, doi:10.1038/s41598-018-35064-5, 2018.

Guemas, V., and 12 Coauthors: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Quarterly J. Royal Meteor. Soc., 142, 546-561, 2016.

Holland, M. M., Bailey, D. A., and Vavrus, S.: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. Climate Dynamics, 36, 1239-1253, 2011.

Jones E., Oliphant, E. Peterson E. *et al.:* SciPy: Open Source Scientific Tools for Python, 2001-, http://www.scipy.org/ [Online; accessed 2018-03-29].

Kinnard, C., Zdanowicz, C. M., Fisher, D. A., Isaksson, E., De Vernal, A., and Thompson, L. G.: Reconstructed changes in Arctic sea ice over the past 1,450 years. Nature, 479, 509-512, 2011.

Kwok, R., and Rothrock, D. A.: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958-2008. Geophys. Res. Lett., 36, L15501, 2009.

Laxon, S.W., Giles, K. A.. Rideout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J.,. Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and

565 Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume. Geophys. Res.

566 Lett., *40*, doi: 10.1002/grl.50193, 2013.

567 Lenton, T. M.: Arctic climate tipping points. Ambio, 41, 10-22, 2012.

568 Lindsay, R. W., Zhang, J., Schweiger, A. J., and Steele, M. A.: Seasonal predictions of ice extent

569 in the Arctic Ocean. Journal of Geophysical Research (Oceans), 113, C02023,

570 https://doi.org/10.1029/2007/C004259, 2008.

571 Maslanik, J., Stroeve, J., Fowler, C., and Emery, W: Distribution and trends in Arctic sea ice age

572 through spring 2011. Geophys. Res. Lett., 38**,** L13502, doi:10.10029/2011GL047735, 2011.

573 Meier, W., Fetterer, F., Savoie, M., Mallory, S., Duerr, R., and Stroeve, J: NOAA/NSIDC

574 Climate Data Record of Passive Microwave Sea Ice Concentration, Version 2. Boulder,

575 Colorado USA. NSIDC: National Snow and Ice Data Center.

576 doi:http://dx.doi.org/10.7265/N59P2ZTG.   (G02202), 2013.

577 National Ice Center and National Snow and Ice Data Center. Compiled by Fetterer, F., Savoie,

578 M., Helfrich, S., and Clemente-Colón,: *Multisensor Analyzed Sea Ice Extent - Northern*

579 *Hemisphere (MASIE-NH), Version 1*. Boulder, Colorado USA. NSIDC: National Snow and Ice

580 Data Center. doi: https://doi.org/10.7265/N5GT5K3K, (G02186), 2010 (updated daily)

581 NOAA: Arctic Report Card 2017. National Oceanic and Atmospheric Administration,

582 ftp://ftp.oar.noaa.gov/arctic/documents/ArcticReportCard_handout2017.pdf (accessed 27 Dec

583 2018)

584 Onarheim, I. H., Eldevik, T., Smedsrud, L. H., and Stroeve, J. C.: Seasonal and regional

585 manifestations of Arctic sea ice. J. Climate, 31, 4917-4932, doi:10.1175/JCLI-D-17-0427.1,

586 2018.

587 Parkinson, C. L.: Spatially mapped reductions in the length of the Arctic sea ice season.

588 Geophys. Res. Lett., 41, 4316-4322, 2014.

Formatted: Left

Formatted: Left

Partington, K., Flynn, T., Lamb, D., Bertoia, C., and Dedrick, K.: Late twentieth century Northern Hemisphere sea-ice record from U.S. National Ice Center ice charts. J. Geophys. Res. 108(C11): 3343. doi:10.1029/2002JC001623, 2003.

Petty, A. A., Schröder, D., Stroeve, J., Markus, T., Miller, J., Kurtz, N., Feltham, D., and Flocco, D: Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. Earth's Future, 5, 254-263, 2017.

Schröder, D., Feltham, D. L., Flocco, D., and Tsamados, M.: September Arctic sea-ice minimum predicted by spring melt-pond fraction. Nature Clim. Change, 4(5): 353–357, 2014.

Schweiger, A. J., Wood, K. R., and Zhang, J.: Arctic sea ice volume variability over 1901-2010: A model-based reconstruction. J. Geophys. Res.-Oceans, in review, 2018.

Serreze, M. C., and Stroeve, J.: Arctic sea ice trends, variability and implications for seasonal sea ice forecasting. Phil. Trans. Royal Society A, 373, 20140159, 2015.

Serreze, M. C., Stroeve, J., Barrett, A. P., and Boisvert, L. N.: Summer atmospheric circulation anomalies over the Arctic Ocean and their influences on September sea ice extent: A cautionary tale. J. Geophys. Res.-Atmospheres, 121(19): 11,463-11,485, doi:10.1002/2016JD025161, 2016.

Sigmund, M., Fyfe, J. C., Flato, G. M., Kharin, V. V., and Merryfield, W. J.: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. Geophysical Research Letters, 40, 529-534, doi:10.1002/grl.50129, 2013.

Stroeve, J., Hamilton, L. C., Bitz, C. M., and Blanchard-Wrigglesworth, E.: Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook, 2008-2013. Geophys. Res. Lett., 41, 2411-2418, 2014.

Tivy, A., Alt, B., Howell, S. E. L., Wilson, K., and Yackel, J.: Long-range prediction of the shipping season in Hudson Bay: A statistical approach. Weather and Forecasting, 22, 1063-1075, 2007.

Formatted: Font: 12 pt

614   Walsh, J. E., Fetterer, F., Stewart, J. S., and Chapman, W. L.: A database for depicting sea ice
615   variations back to 1850. Geograph. Rev., 107, 89-107, 2016.

616   Walsh, J. E., Chapman, W. L., and Fetterer, F.: Gridded Monthly Sea Ice Extent and
617   Concentration, 1850 Onward, Version 1. Boulder, Colorado USA. NSIDC: National Snow and
618   Ice Data Center. doi: http://dx.doi.org/10.7265/N5833PZ5, (G10010), 2015 (updated 2016).

619   Williams, J., Tremblay, B., Newton, R., and Allard, R.: Dynamic preconditioning of the
620   minimum September sea ice extent. J. Climate, 29, 5879-5891, 2016.

621   Yuan, X., Chen, D., Li, C., and Wang, W.: Arctic sea ice seasonal prediction by a linear markov
622   model. J. Climate, 29, 8151-8173, 2016.

623 **Appendix.** Reconstruction of the Barnett Severity Index, 1953-2013

624

625 As described in Section 2, the Barnett Severity Index (BSI) is a combination of five metrics of ice

626 coverage in the Beaufort Sea. Drobot et al. (2003) used the BSI through 2000 in their evaluation

627 of predictability based on multilinear regression against various measures of sea ice cover. In

628 order to update the BSI for use in this study, we base a reconstruction on the digital grids of sea

629 ice concentration in the Historical Sea Ice Atlas (HSIA) for Alaska (http://seaiceatlas.snap.uaf.edu/

630 accessed 27 Dec 2018). As with the regional ice extent calculations using G10010 (Section 3), we

631 use the HSIA because it extends the record 26 years back in time before the start of the satellite

632 passive microwave record. While the sources of the ice concentration data in the HSIA are the

633 same as in G10010, a notable advantage of the HSIA is its weekly temporal resolution (vs. the

634 monthly resolution of G10010). The HSIA also has a spatial resolution of ¼° latitude by ¼° degree

635 longitude. Because of the weekly time resolution, the distance metrics (3)-(5) of the BSI are

636 truncated to the nearest week. Similarly, the distance metrics (1) and (2) are truncated to the

637 nearest 27.8 km (15 n mi). One of the within-month dates of the HSIA grids is the 15th of each

638 month, so no temporal interpolation is necessary for metrics (1) and (2). The reconstructed values
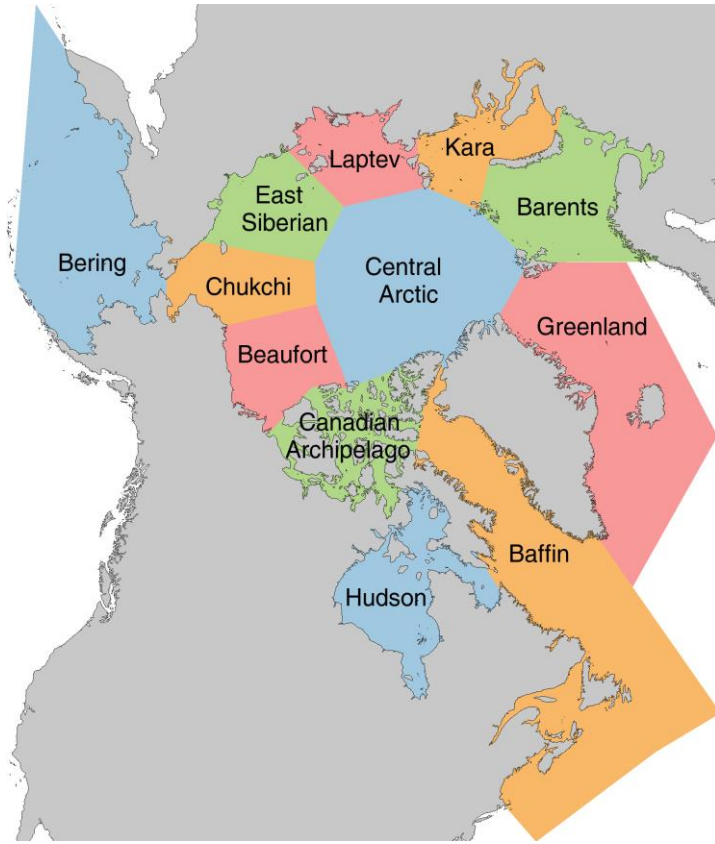
639 of the BSI are listed in Table A1.

640

641

Figure 1. The MASIE subregions used in the study (NIC and NSIDC, 2010).

645
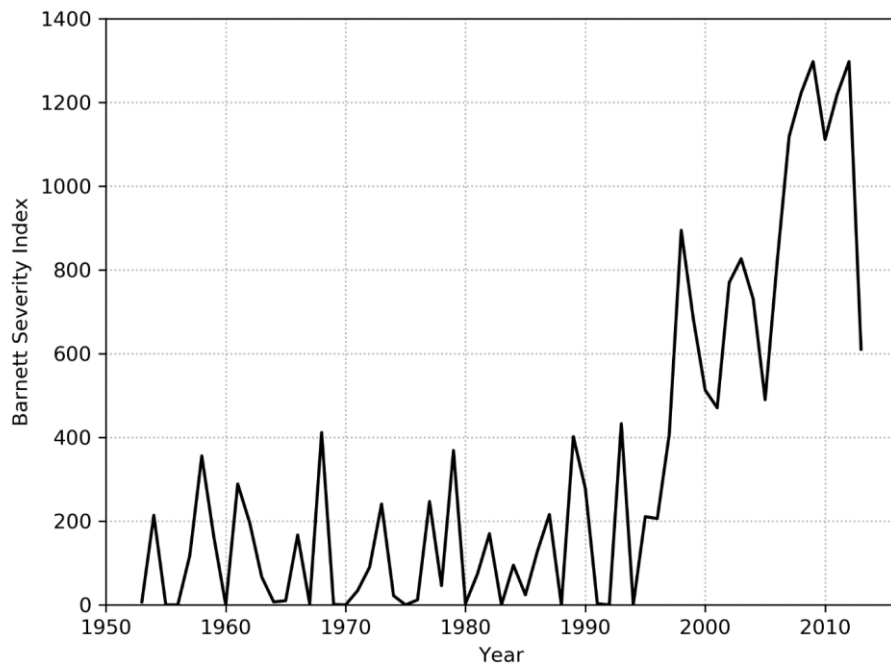
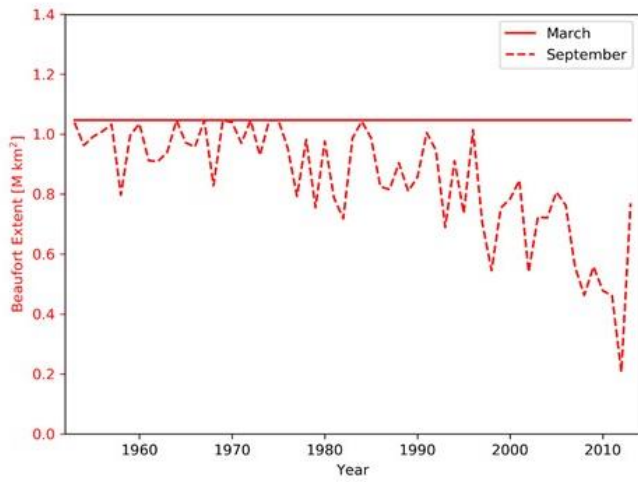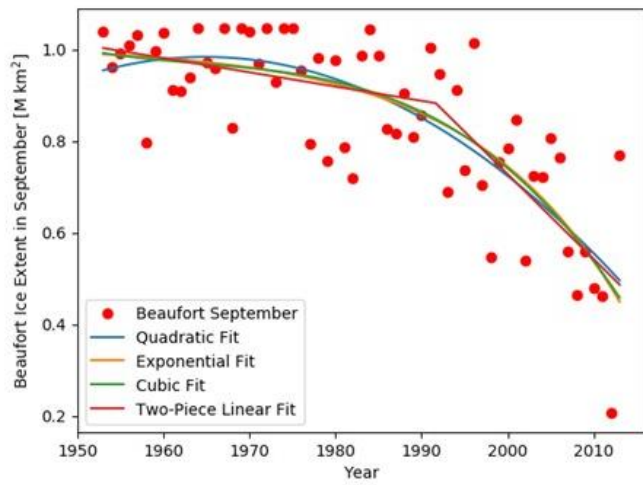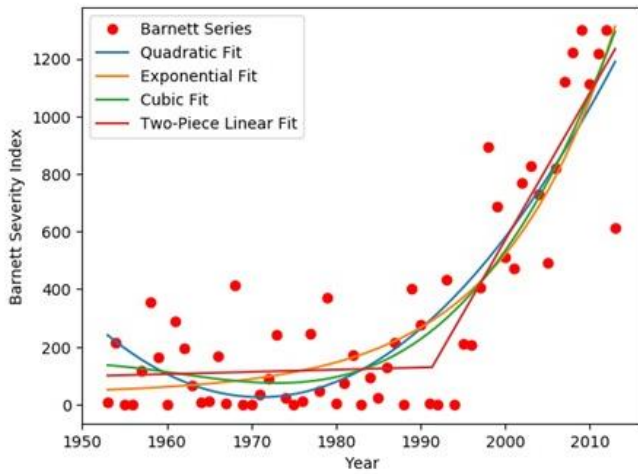**Figure 2.** Time series of the Barnett Severity Index (BSI), 1953-2013.
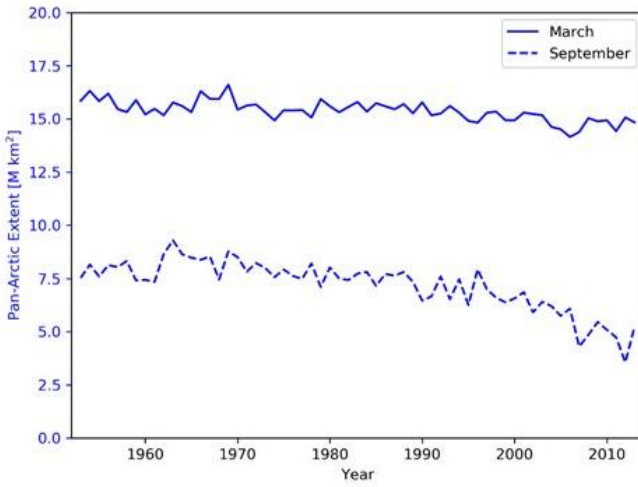
(a) Pan-Arctic

(b) Beaufort

652

Figure 3. (a) Total Arctic sea ice extent and (b) the extent of ice in the Beaufort Sea during
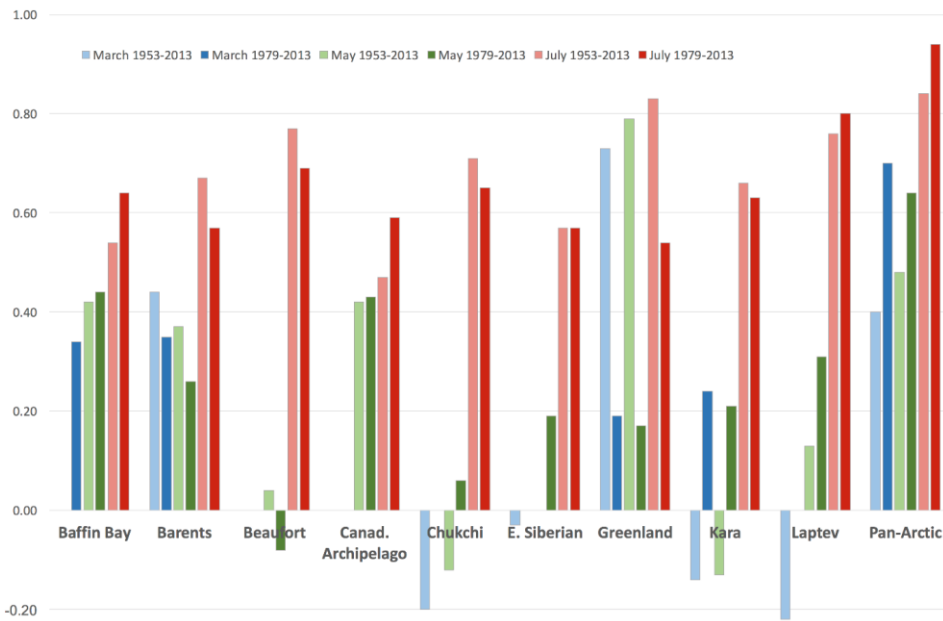March(solid lines) and September (dashed lines) .

655

656

657

Figure 4. Examples of different fit methods (see legend) applied to the BSI (upper panel) and
the September Beaufort ice extent time series (lower panel).

658

659

660

Figure 5. Correlations of September ice extents in individual seas with ice extent in the same region in March (green bars), May (blue bars) and July (red bars). Correlations are also shown for Pan-Arctic extent (far right). The correlations are based on non-detrended data. In each case, light-colored bars are for 1953-2013 and dark-colored bars are for 1979-2013. The absence of a bar indicates a correlation of zero.
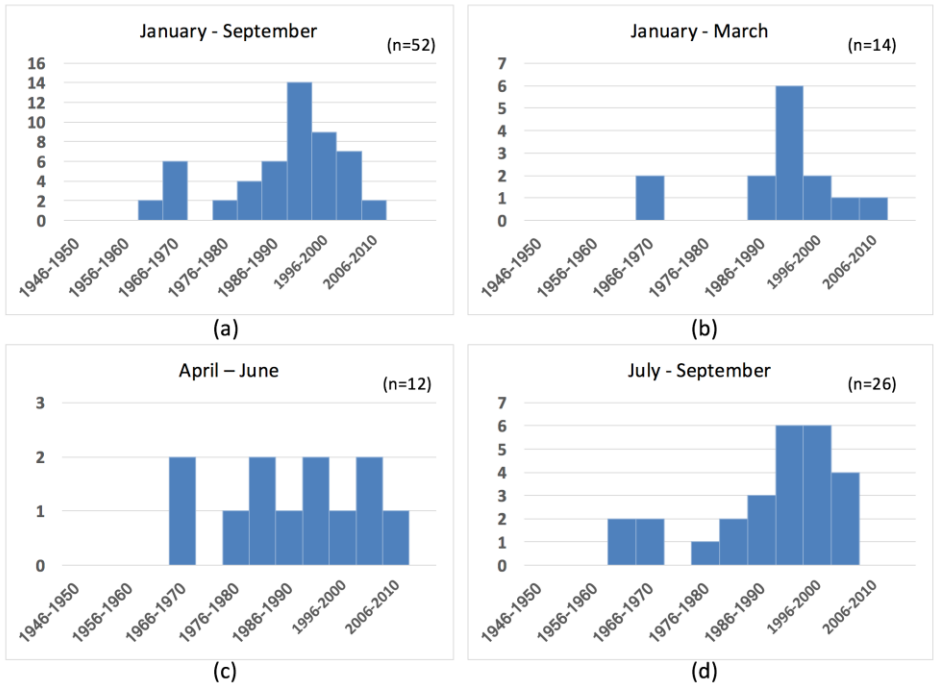
Figure 6. The distribution of break-point years across all regions for (a) January-September and its three subperiods: (b) January-March, (c) April-June, (d) July-September). Only cases for which detrending using two lines, rather than one, reduced the rms error by 5% or more are included. Note that y-axes have different scales.

676

677
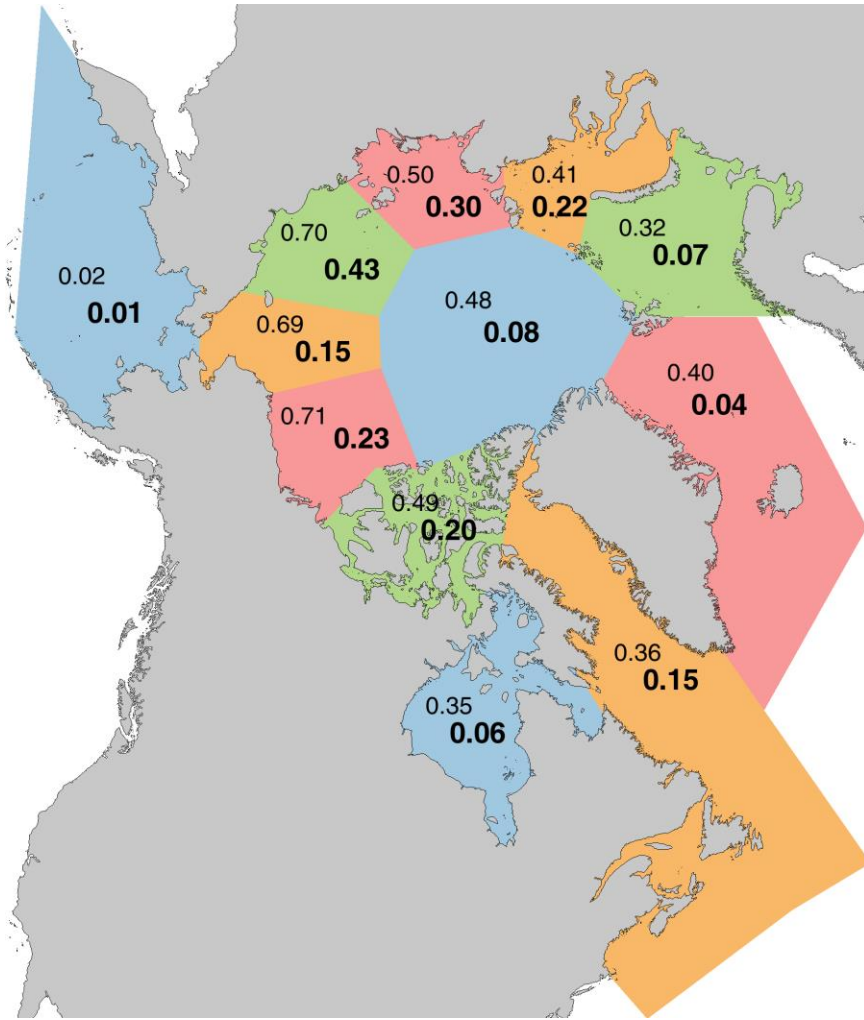
Figure 7. Squares of correlations (R^2) between September pan-Arctic ice extent and September regional ice extent based on ice extents including trends (upper numbers in normal font) and detrended (lower numbers, bold font).

678

679

680

681
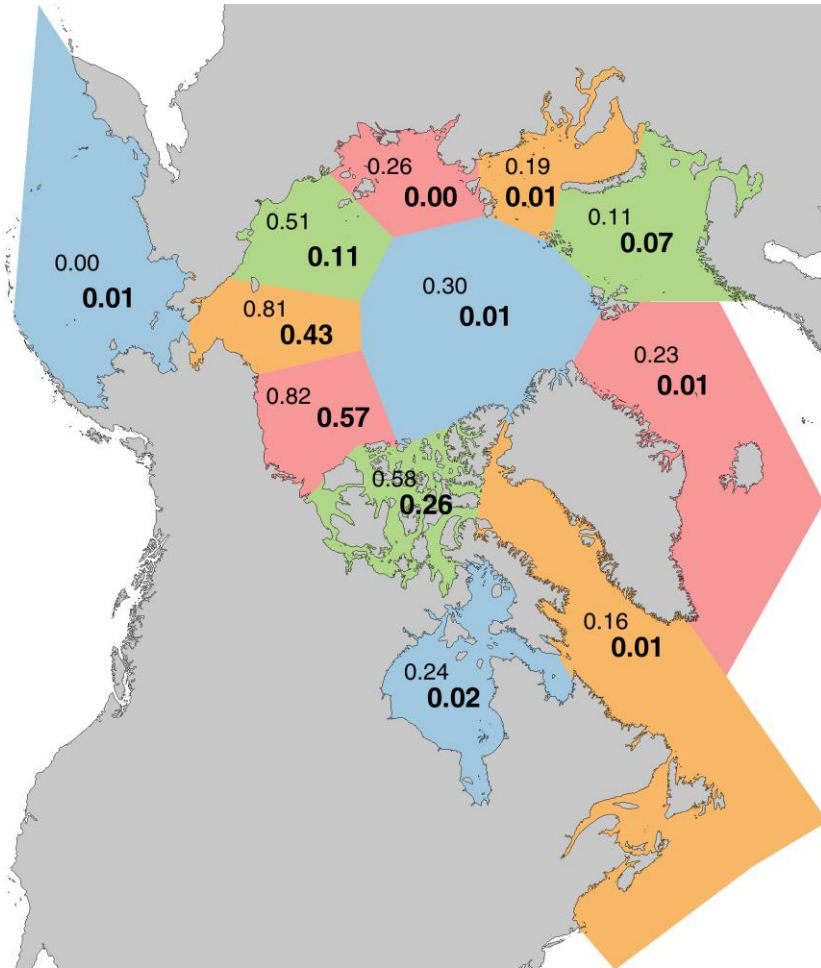
Figure 8. As in Figure 7, but for squares of correlations between the annual BSI and September
regional ice extents based on raw (not detrended) time series (upper numbers) and
detrended time series (lower numbers, bold font).
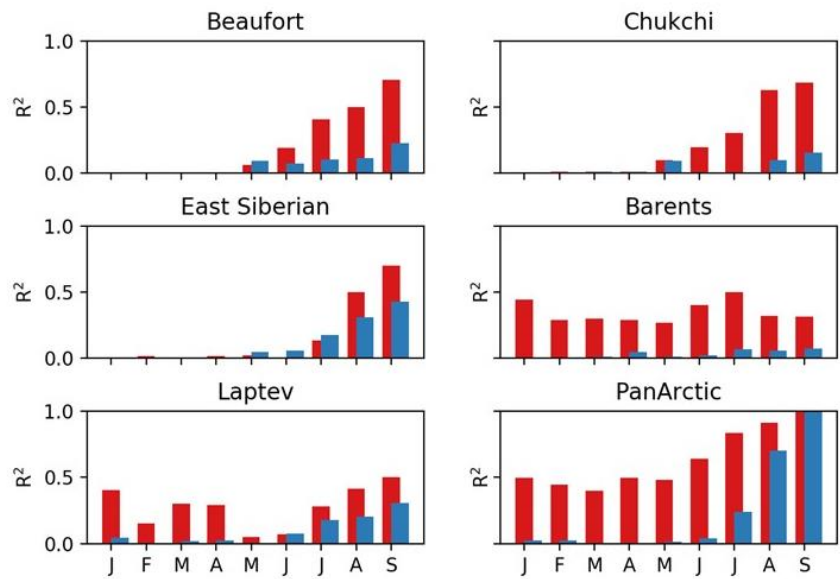
Figure 9. Examples of variances of September pan-Arctic ice extent and explained by correlations with antecedent regional ice extent in individual calendar months from September back to January (pan-Arctic extent lagging by 0, 1, 2, …, 8 months). Correlations are plotted as fractions of explained variance (squares of correlations). Red bars are correlations with trends included, blue bars are correlations after removal of trends.

695
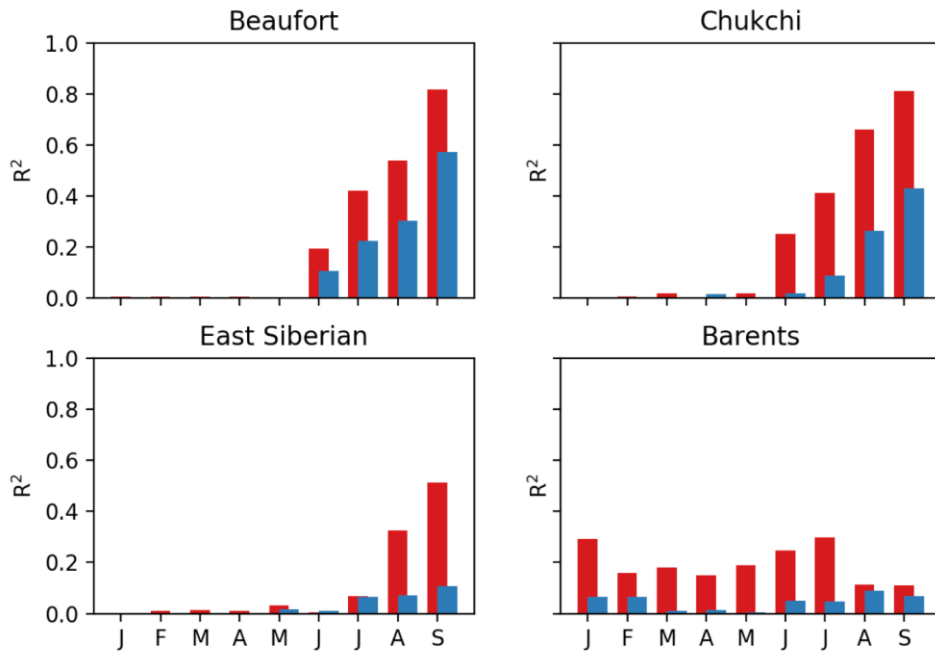
696
697

698 **Figure 10.** Examples of variances explained by correlations between the Barnett Severity Index
699 and regional ice extent in individual calendar months from September back to January
700 (BSI lagging by 0, 1, 2, …, 8 months). Correlations are plotted as fractions of
701 explained variance (squares of correlations). Red bars are correlations with trends
702 included, blue bars are correlations after removal of trends.

703

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Baffin-St. Lawrence	0.08	0.06	0.02	0.16	0.32	0.49	0.61	0.52	0.36
Barents	0.45	0.29	0.30	0.29	0.27	0.41	0.50	0.32	0.32
Beaufort	0.41	0.41	0.41	0.41	0.06	0.19	0.41	0.50	0.71
Bering	0.00	0.00	0.01	0.00	0.01	0.12	0.08	0.01	0.02
Canadian Archipelago	0.41	0.41	0.41	0.41	0.00	0.13	0.09	0.52	0.49
Central Arctic	0.21	0.11	0.18	0.01	0.02	0.02	0.15	0.07	0.48
Chukchi	0.00	0.01	0.01	0.01	0.10	0.20	0.31	0.63	0.69
East Siberian	0.00	0.02	0.01	0.02	0.02	0.00	0.14	0.50	0.70
Greenland	0.47	0.53	0.50	0.48	0.43	0.41	0.45	0.29	0.40
Hudson	0.05	0.41	0.41	0.26	0.03	0.32	0.66	0.37	0.35
Kara	0.00	0.03	0.11	0.04	0.10	0.09	0.44	0.42	0.41
Laptev	0.40	0.15	0.30	0.29	0.05	0.07	0.28	0.42	0.50
Pan-Arctic	0.50	0.44	0.40	0.50	0.48	0.64	0.84	0.91	1.00

Table 1. Correlations between monthly regional ice extent and pan-Arctic ice extent expressed as explained variance (R^2). Cases where at least 10% of the variance in pan-Arctic ice extent is explained by regional ice extent in a given antecedent month are highlighted with bolded region names. Levels of shading of boxes denote values exceeding 0.10, 0.20, 0.30,…

Region	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Baffin-St. Lawrence	0.09	0.04	0.08	0.06	0.01	0.00	0.03	0.16	0.15
Barents	0.00	0.01	0.01	0.05	0.01	0.02	0.07	0.06	0.07
Beaufort	0.05	0.05	0.05	0.05	0.10	0.08	0.11	0.11	0.23
Bering	0.01	0.01	0.08	0.03	0.02	0.00	0.01	0.01	0.01
Canadian Archipelago	0.05	0.05	0.05	0.05	0.01	0.02	0.02	0.16	0.20
Central Arctic	0.02	0.02	0.11	0.03	0.02	0.04	0.07	0.00	0.08
Chukchi	0.00	0.00	0.01	0.01	0.10	0.00	0.00	0.10	0.15
East Siberian	0.00	0.00	0.00	0.00	0.05	0.06	0.18	0.31	0.43
Greenland	0.06	0.04	0.09	0.07	0.03	0.06	0.04	0.00	0.04
Hudson	0.00	0.05	0.05	0.01	0.05	0.01	0.11	0.07	0.06
Kara	0.01	0.03	0.03	0.04	0.00	0.18	0.12	0.13	0.22
Laptev	0.05	0.00	0.02	0.02	0.01	0.08	0.18	0.21	0.30
Pan-Arctic	0.03	0.02	0.00	0.00	0.01	0.04	0.24	0.70	1.00

Table 2. Correlations between detrended monthly regional ice extent and detrended September pan-Arctic ice extent expressed as explained variance (R^2). Cases where at least 10% of the variance in September pan-Arctic ice extent is predictable by regional ice extent in a given antecedent month are highlighted with bolded region names. Shading of boxes is as in Table 1.

Formatted: Centered

1953	7	1984	95
1954	213	1985	24
1955	0	1986	178
1956	0	1987	216
1957	117	1988	0
1958	356	1989	402
1959	163	1990	278
1960	0	1991	3
1961	289	1992	0
1962	195	1993	434
1963	66	1994	1
1964	7	1995	211
1965	10	1996	206
1966	167	1997	407
1967	3	1998	895
1968	412	1999	685
1969	1	2000	513
1970	0	2001	471
1971	34	2002	770
1972	90	2003	827
1973	240	2004	731
1974	22	2005	490
1975	0	2006	819
1976	13	2007	1119
1977	247	2008	12239
1978	46	2009	12989
1979	368	2010	1112
1980	3	2011	1219
1981	74	2012	1298
1982	170	2013	611
1983	0		

Table A1. Yearly values of the Barnett Severity Index (BSI). Source: Rebecca Rolph, Geophysical Institute, University of Alaska, Fairbanks.

5

Formatted: Justified