WCD-2021-18

Identification, characteristics, and dynamics of Arctic extreme seasons

Response to the Reviewers' comments by Katharina Hartmuth, Maxi Boettcher, Heini Wernli, and Lukas Papritz

We thank all three reviewers for their insightful and helpful comments. We address each comment point by point below. The reviewers' comments are given in blue and our responses in black. The most important aspects of our replies and revisions are:

1) As suggested by reviewer 1, we used the North et al. approach to show that the first two PCs are statistically distinguishable from the others.
2) We now better explain the several subjective choices that were necessary for our analyses.
3) We clarify our choice to use a multivariate approach to explore different types of extreme seasons.

Reviewer 1

General comments

The authors evaluate the atmospheric conditions during anomalously extreme seasons in the Arctic. This is performed using a regional principal component (PC) analysis (PCA) from ERA5 data of the first two PCs of all seasons from 1979-2018. Furthermore, the PCA uses six key surface variables and divided spatially into 9 Arctic sub-regions. The sub-regions are subjectively chosen, but based on climatological sea ice conditions in either the Nordic Seas, Kara-Barents Seas, and the rest of the Arctic. Results identify 2-3 extreme seasons for each season (DJF, MAM, JJA, SON) in each sub-region. The PCA applied here provides a quantification of which variables contribute most to the extreme conditions of the respective season, and how consistent those conditions are during those particular seasons. The authors then choose two extreme seasons in the Kara-Barents sea during winter (DJF) to further investigate the synoptic weather conditions that were occurring and how they might have lead up to the resulting seasonal extremes. The chosen seasons are picked based on their orthogonal, yet extreme, projections onto the PCs.

This research nicely demonstrates how PCs can be used to identify seasonal anomalies and extremes in certain regions of the Arctic. It furthermore demonstrates how to use that information to provide an expectation of how an extreme season was characterized with regard to one of the six variables and how consistent those conditions were. It is certainly a nice way to be able to identify extreme seasons that might be worth analyzing in further detail at shorter time and space scales if desired. Overall, I think these results can make a contribution and be published once some remaining issues are addressed. In particular:

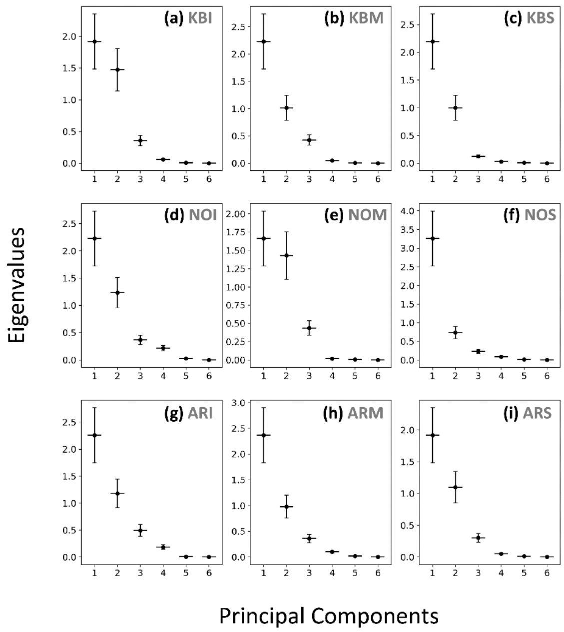
1) Picking the first two principal components is subjective and does not necessarily isolate most of the variance. It first needs to be established that the first two principal components are the only significant ones. I do not doubt that this is the case given that on line 282, it is stated that they usually explain 80-90% of the variance. However, it should be shown that they are indeed statistically distinguishable from the others. North et al. (1982) provide a well-established method of statistically distinguishing the first few eigenvalues from the others.

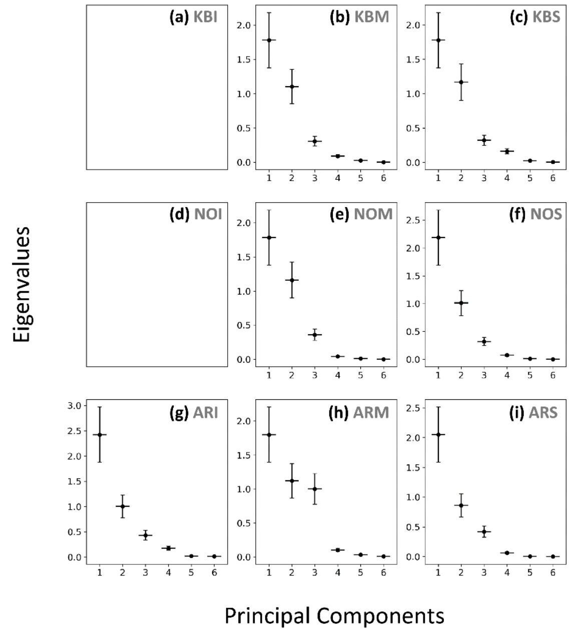
Thank you for this comment and for pointing us to the method introduced by North et al. We applied this method and the results reveal that the first two PCs in DJF and JJA are, with the exception of sub-region **ARM** in JJA, always statistically distinguishable from the others. We will add this information to the revised manuscript. Here we provide further details about our results from applying the North et al. (1982) method.

Figures R1 and R2 show the standard errors for each eigenvalue in our PCA as introduced by North et al. (1982) for each sub-region in DJF (left-hand side) and JJA (right-hand side), respectively. The estimate for the standard error is given by

$$\delta\lambda_{\alpha} \approx \lambda_{\alpha}(2/N)^{1/2},$$

where λ_{α} denotes the respective eigenvalue and N the sample size, which in our case corresponds to the 39 realizations of the four seasons of the study period. Along the y-axis, the eigenvalue for each Principal Component (PC) is given and the error bars represent the estimated standard error. For both seasons, the first two eigenvalues, λ_1 and λ_2, are either clearly distinguishable or their error bars show only a very small overlap, except for sub-regions **KBI** and **NOM** in winter. Further, λ_1 and λ_2 are always clearly distinguishable from the third eigenvalue λ_3. The only exception is the sub-region **ARM** in JJA for which the error bars of λ_2 and λ_3 have a significant overlap. Thus, we can show that the first two PCs, which we use for the definition of our extreme seasons and which explain between 80-90% of the variance in the respective sub-regions, are almost always statistically distinguishable from the remaining eigenvalues. We conclude that PC1 and PC2 isolate most of the variance and the corresponding eigenvalues are statistically distinct.





Figures R1 and R2: Standard errors for each PCA eigenvalue for all sub-regions in DJF (Fig. R1) and JJA (Fig. R2). The number of the Principal Component is given along the x-axis and the eigenvalue of each Principal Component along the y-axis. Error bars denote the estimated standard error following North et al. (1982).

2) Section 3 and generally throughout: The values of all correlations and their p-values that are described should be listed in a table.

We agree that it would be helpful to add a list containing all correlations and respective p-values for the described relations between the different parameters. We will thus add such a table to the supplementary material and refer to it in the paper.

3) Figures 5 and 6 are a very nice way to illustrate the seasonal anomalies and the variability that may have also been occurring within those seasons. Having never seen these diagrams before, it at first takes a little bit of time to understand. It would be very helpful if there were a schematic showing the "phase space" of the interpretation that illustrates what is said in words on lines 251-258 (i.e., regions on the graph where there would be anomalies that tend to be continuous, where there would be warm episodes alternating with weak cold episodes, where there would be several intense warm and cold episodes that nearly cancel, where they would be near the climatology, etc.).

Thank you very much for pointing this out. To better understand and interpret the figures, we will add lines of a constant ratio of the seasonal-mean anomaly and the seasonal-mean absolute anomaly ($\frac{\overline{\chi^*}}{\overline{|\chi^*|}}$) to the diagrams, such as you can see in the schematic figure below. We will further adapt lines 247-258 as follows:

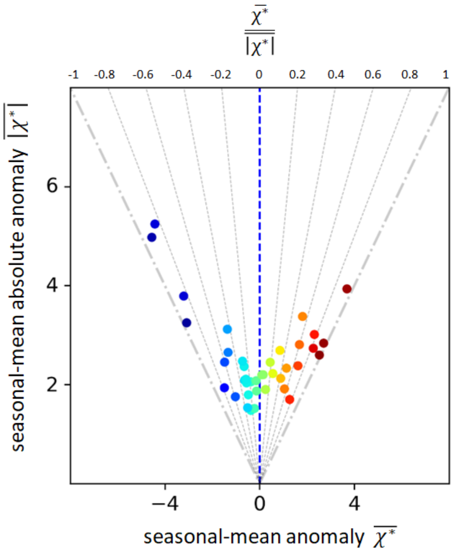


Figure R3: Schematic figure showing seasonal-mean anomalies ($\overline{\chi^*}$, along x-axis) vs. seasonal-mean absolute anomalies ($\overline{|\chi^*|}$, along y-axis). The dots in this schematic figure are colored by the ratio of both measures, $\frac{\overline{\chi^*}}{\overline{|\chi^*|}}$, which is additionally visualized by grey dashed lines. $\frac{\overline{\chi^*}}{\overline{|\chi^*|}} = \pm 1$ is shown by stippled grey lines and $\frac{\overline{\chi^*}}{\overline{|\chi^*|}} = 0$ is shown by a dashed blue line.

'To better understand the seasonal substructure of Arctic winters and summers, we compare the seasonal-mean anomalies ($\overline{\chi^*}$) with the seasonal-mean absolute anomalies ($\overline{|\chi^*|}$) for T_{2m}, P and E_S in selected sub-regions in DJF (Fig. 5) and JJA (Fig. 6). The ratio of seasonal-mean and seasonal-mean absolute anomalies, $\frac{\overline{\chi^*}}{\overline{|\chi^*|}}$, is indicative of the temporal persistence of an anomaly throughout a season. Thus, the location of a season in the diagrams provides information about the substructure of the season in terms of the considered parameter. In general, the further to the right, the more positive is the seasonal-mean anomaly of the shown parameter and the further to the left, the more negative. The closer the seasonal-mean anomaly is to the seasonal-mean absolute anomaly (dots close to the outer stippled grey line representing $\frac{\overline{\chi^*}}{\overline{|\chi^*|}} = \pm 1$), the more persistent the anomaly is throughout the season. Thus, we define seasons with $0.8 \leq \frac{\overline{\chi^*}}{\overline{|\chi^*|}} \leq 1$as seasons with a "continuous" anomaly. With a decreasing absolute value of$\frac{\overline{\chi^*}}{\overline{|\chi^*|}}$, the seasons are located further away from the outer stippled grey lines, meaning that positive or negative anomalies in the respective parameter occur more episodically throughout the season. The closer a season is positioned towards the blue dashed line where $\overline{\chi^*} = 0$ and thus $\frac{\overline{\chi^*}}{\overline{|\chi^*|}} = 0$, the more positive and negative daily anomalies cancel each other, leading to a weak overall seasonal anomaly. The value of $\overline{|\chi^*|}$ is further indicative of the magnitude of the daily anomalies throughout a season. A season located at the top of the plot shows stronger daily anomalies than a season with the same $\frac{\overline{\chi^*}}{\overline{|\chi^*|}}$ratio but a smaller $\overline{|\chi^*|}$.

For example, a season can be anomalously warm because the daily-mean T_{2m} values are larger than the climatology on almost all days of the season , resulting in $\frac{\overline{T_{2m}^*}}{\overline{|T_{2m}^*|}} \approx 1$. With a decreasing ratio of both anomaly metrics, e.g.$\frac{\overline{T_{2m}^*}}{\overline{|T_{2m}^*|}} = 0.5$, the season is still anomalously warm, but it results from several warm episodes alternating with weaker and/or shorter periods with negative T_{2m}* values. If $\frac{\overline{T_{2m}^*}}{\overline{|T_{2m}^*|}} \approx 0$, cold and warm episodes cancel each other leading to a weak overall seasonal anomaly. Comparing two seasons with the same $\frac{\overline{T_{2m}^*}}{\overline{|T_{2m}^*|}}$, the season which is positioned further up in the plot (showing larger values of $\overline{|T_{2m}^*|}$ and $\overline{T_{2m}^*}$) shows a larger variability in T_{2m} with more intense warm and/or cold episodes compared to the season which is located further down.'

4) The justification for choosing two winter cases is weak. Perhaps this is because the anomaly values are smaller in the summer. But is it not dM and the standardized anomalies that determine how extreme a season is with these methods? These are just as strong in the summer (Table 2). I can see that the results shown in Figure 5 are used to pick the cases, but again, this seems contrary to the main setup of this paper of using the PCs to determine the extremes. Also, why 2011/12 in the Kara Barents Sea when this categorizes as "anomalous" rather than "extreme" in these methods? Regardless, it is hard to justify the title "Identification, characteristics, and dynamics of Arctic extreme seasons" when only the dynamics of winter extreme seasons were discussed.

We agree that our choice of the case study seasons (three winter seasons, two in the Kara and Barents Seas, and one in the High Arctic) is subjective. They are also to a certain extent a compromise between "many more seasons would be interesting to study in greater detail" and "the paper should remain readable and have a reasonable length". We think that choosing three winter seasons makes their comparison easier and their differences more revealing. We would like to show that even in the same region (Kara and Barents Seas) two anomalous/extreme winter seasons can have a completely different substructure and can be associated with different weather systems, emphasizing the inter-annual variability. Showing this for a winter and a summer season would be less surprising and interesting as this would mix seasonal and inter-annual variations. It is also important to us to evaluate the case studies in some depth, which limits the number of cases fitting into the paper to about three. Of course, we also looked at other seasons but then decided that the selected cases nicely illustrate the diversity and complexity of the involved processes, which is one of the key aims of our study. In the revised version we will better explain that the choice of the case studies is subjective and motivated by our intention to reveal the diversity and complexity of the involved processes. As a caveat, we will also mention that a more in-depth investigation of extreme summer case studies remains to be done.

Regarding the title, it's true that we identify and characterize extreme seasons in summer and winter, but then only discuss the dynamics in winter. However, as we now explicitly mention this limitation and better explain the key aim of the case studies, we think that the title is justified.

Specific comments:

1) Line 135: Why are only marine cold air outbreaks (CAOs) considered? There are also significant CAOs over land, described in Biernat et al. (2021).

We are only considering ocean and ice grid points and thus only marine cold air outbreaks, which are identified on grid points with less than 50% sea ice.

2) Line 186: Choosing a dM threshold of 3 seems quite subjective. How is this threshold picked? If each principal component has a significant anomaly of two standard deviations, this could provide an expectation for what would be significant when considering the PCs in combination.

The thresholds for anomalous resp. extreme seasons are indeed a rather subjective choice. However we find that with these thresholds we obtain on average 0-1 extreme seasons per sub-region (which equals 0-2.5% of all seasons) and 4-5 anomalous seasons per sub-region (equalling 15-17% of all seasons). Assuming a normal distribution, these values correspond to the range of 2-3σ for our extreme seasons and 1-2σ for our anomalous seasons. Further, with this number of extreme seasons, the return period of such a season corresponds to approximately 40 years. Several studies, e.g., Röthlisberger et al. (2021) used this as an adequate measure for defining extreme seasons.

As a side note, we mention that preliminary analyses of 1000 years of (present-day) CESM large-ensemble data show that our chosen threshold of d_M=3 results on average in a return period of around 70-90 years. We are, thus, confident that classifying the seasons with d_M>3 as "extreme" is well justified.

3) Line 219: Be more specific about "almost always." What percentage of the time is it true? Same thing for line 225... what percentage of the cases translates to `usually'?

Thank you for pointing this out. We will adjust the manuscript in the indicated section to clarify the mentioned relationships between the different variables.

4) Line 262: How close to the |P*| = P* line does a season need to be in order to be called "continuous?" For example, the 2016/2017 winter season was pretty close, but not exactly on it. On the other hand, there are very few cases of |T2m*| = T2m* being exactly equal in the summer while it is described as "continuous" on line 260.

There are indeed only very few cases where the seasonal-mean and seasonal-mean absolute anomalies of a season are equal. Thus we changed our definition of a "continuous anomalous season" from $\overline{|\chi^*|} = \overline{\chi^*}$ to $\overline{|\chi^*|} \approx \overline{\chi^*}$, including seasons with a ratio of $0.8 \leq \frac{\overline{\chi^*}}{\overline{|\chi^*|}} \leq 1$(see previous comment). We will add these changes to the revised manuscript.

5) Line 307: Would also be useful to point out that there is very little 2-m temperature variability over the Arctic sea ice in the summer. This could imply that temperature variability may not play a major role in sea ice loss, which has very large interannual variability in the summer.

Figure 8 does indeed suggest that T_{2m} has only little variability in regions with $SIC_{clim}>0.9$ in summer compared to other sub-regions. However, T_{2m} is capped above sea ice as the air is cold and the excess energy goes into the melting of the ice if T_{2m} is above the freezing point, which essentially limits (near-surface) temperature variability. We further assume that the sea ice loss in summer is equally strong in the other sub-regions (especially the mixed sub-regions with very variable SIC), which show a larger variability in SIC. As there exists only one sub-region with $SIC_{clim}>0.9$ in summer, we think that additional analyses would be needed to make such a statement.

6) The justification of how an extreme season is chosen on Lines 310-314 should be moved up to Section 2.3.

We already explain this in Sect. 2.3. The text in lines 310ff is meant as a reminder. We now clarify this by writing “As explained in Sect. 2.3, …”.

7) Line 315: Which season does Figure 2 show? This could also be referenced here along with Table 2.

Figure 2 in the manuscript is only a schematic biplot which does not refer to a specific season nor region. We slightly changed the figure caption to clarify that this is only an idealised plot.

8) Figures 9, 10, 14: Would be helpful to label the x-axis with the month/date instead of the day of the season, esp. to be consistent with the text.

We mostly use “on day 12,15, 20…” throughout the text and only rarely real dates. Thus, we adapted the manuscript such that we don’t use specific dates anymore, as we think that this ensures better readability.

9) Line 367: How are blocking, cyclone, and CAO frequencies computed exactly? Need references and a short description.

A common feature of our weather system identification schemes is that they produce a two-dimensional binary field, often referred to as the “mask” of the weather systems, where grid points that belong to a system have a value of 1 and the others have a value of 0. Simple time averaging of these binary fields then automatically delivers the weather system frequency field. For example, if a cyclone mask covers a grid point at 25% of all times, then averaging 25% times a value of 1 and 75% a value of 0 leads to a frequency of 0.25. For the specifics of the identification scheme, we will add a few sentences for each weather system and give the relevant references to the papers that introduced these schemes.

10) Line 389: "Several episodic precipitation events..." But wouldn't Fig. 5h suggest consistent precipitation events?

Thank you for pointing this out. We will replace “episodic” by “recurrent” to clarify the constant occurrence of precipitation events throughout the season.

11) Line 431: Remove "it is obvious that"

We changed “it is obvious that” to “it can be seen that”.

12) Line 440: Please also label JJA 2016 in Figs. 6 and 8.

To make it clear that JJA 2016 is somehow connected to our case study DJF 2016/17, we will label it in Figs. 6 and 8 in the manuscript.

13) Lines 441-445: It is misleading to say that there were positive temperature anomalies over large parts of the Arctic in JJA 2016. This and the blocking was more centered over the Kara-Barents Sea region, while much of the central Arctic was not exceptionally warm and had frequent cyclones.

We are not sure if we understood your remark correctly, as we do not state that the positive surface temperature anomalies in JJA 2016 occurred over large parts of the Arctic, but only in the Kara and Barents Seas. We then state that in autumn 2016 (mainly during October and November), positive temperature anomalies occurred across the whole Arctic region as already shown by Tyrlis et al. (2019). We will clarify this further in the text by replacing “during autumn 2016” by “during October and November 2016”.

14) Section 5.3: If the blocking frequency was greatest over Scandinavia, why were the warmest temperature anomalies over the Kara-Barents (KB) region and not co-located with the blocking? Seems like there should have been northerly flow over much of the KB region from air flowing over sea ice. Is it surprising that the air mass was not modified by the time it reached KB?

Thank you for pointing this out. First of all we want to emphasize that DJF 2016/17 was not a particularly warm season, but experienced several episodic warm events. Blocking over Scandinavia influenced the surface temperatures in the Kara and Barents Seas, especially during the warm episode in February 2017. Trajectories show that a majority of the air causing this warm episode originated over Scandinavia and was undergoing subsidence (we will add a short evaluation of some air mass trajectories to the supplement; see answer to comment (15) of reviewer 2 and Fig. R6). However, the pattern of blocking and cyclone anomaly patterns as shown in Fig. 12 in the manuscript does also support northerly flow into the region as you correctly assume, causing for example the period with a strong CAO in mid-February 2017, when cold air is transported from the High Arctic towards the South, facilitated by a block over Scandinavia and a cyclone in the eastern part of the Kara and Barents Seas. Please have a look at the supplementary animation S2 where we show the synoptic evolution for each day throughout the season. We will also try to further shape section 5.3 to better highlight in which way the synoptic patterns influenced the surface temperatures in our case study sub-regions.

15) Lines 120, 541: Is this approach really novel given that (Graf et al. 2017) first introduced it in a similar application?

Using a PCA for finding dominant variability modes has been done in several studies such as, e.g., Graf et al. (2017). However it has never been used to define anomalous or extreme seasons based on the combination of several parameters. Thus, in terms of defining extreme seasons, this approach is novel. We will rephrase the mentioned lines such that this becomes clearer.

Technical corrections:

1) Table 1: 2 m temperature --> 2-m temperature

We followed the WCD submission guidelines (see “House standards” for hyphen usage: “It is our house standard not to hyphenate modifiers containing abbreviated units (e.g. “3-m stick” should be “3 m stick”)).

2) Table S1: Caption states standardized values are in brackets, but they are instead in parentheses.

Thank you for spotting this. We replaced “brackets” with “parentheses” in the caption of Tables 2 and S1.

3) Section 2 should be "Data and methods" given that there is more than one method used to complete the analysis.

Changed as suggested.

4) Figure 1 caption: State what the green and red boxes denote.

We added the following sentence to the caption of Figure 1: “Green and red boxes denote the areas of the Kara and Barents Seas and Nordic Seas, respectively.”

5) Line 135: There does not need to be a space between the number and the "%" symbol

Again, we followed the WCD submission guidelines (see Figure content guidelines: “Spaces must be included between number and unit (e.g. 1 %, 1 m).”).

6) Lines 140-141: What is the sign convention for the surface energy balance?

Thanks for hinting at this. We added the following sentence in line 141: “Positive signs denote net energy fluxes into the surface, whereas negative signs indicate net energy fluxes into the atmosphere.”

7) Line 183: There should be a period at the end of the equation.

A period has been added at the end of the equation.

8) Lines 352, 465: normal --> average

Changed as suggested.

9) Line 393: This --> These

Thanks for spotting this, we changed “this” to “these”.

10) Line 424: Remove "of"

“Despite of this” has been changed to “despite this”.

11) Line 453: Insert "of" after "Comparison"

“Comparison DJF 2011/12 and DJF 2016/17” → “Comparison of DJF 2011/12 and DJF 2016/17”.

12) Line 463: got --> became

Changed as suggested.

References:

Biernat, K. A., L. F. Bosart, and D. Keyser, 2021: A climatological analysis of the linkages between tropopause polar vortices, cold pools, and cold air outbreaks over the central and eastern United States. Mon. Wea. Rev., 149 (1), 189-206.

Graf, M. A., H.Wernli, and M. Sprenger, 2017: Objective classification of extratropical cyclogenesis. Quart. J. Roy. Meteor. Soc., 143 (703), 1047-1061.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. Mon. Wea. Rev., 110 (7), 699-706.

Reviewer 2

The paper presents an analysis of variability in three Arctic regions using 6 metrics. An input of those metrics into the dominant modes of variability and links between those metric are discussed. Overall, I am impressed by the amount and quality of work done in this study.

Here is what I like about the paper:

- Fig 5 and 6, which show that while strong anomalies may be observed in one or two metrics, other metrics may remain close to their climatological values;
- assessment of the input of the six metrics into the main modes of variability and relationships between them;
- case studies (particularly fig. 10,11, 14) and the discussion around them. An attempt to establish a connection between the weather and seasonal anomalies is valuable;

- a wide range of metrics used in the study - not only t2m/SIC/P, but also energy fluxes, cyclone frequency, CAO and a blocking index.

However, there is a couple of major concerns that need to be addressed before the paper can be accepted for publication:

1. I am not convinced that the approach, introduced in the paper, is a good way to select extreme seasons. Despite the use of a multivariate approach, it often comes to just one metric showing a strong seasonal anomaly, which was enough to identify the season as extreme or anomalous. Thus, without applying this approach, one may simply go through all 6 metrics and select the most extreme season(s) in each of them. I don't think I saw a proof that the seasons selected with the PCA analysis were more anomalous than those that showed a strong anomaly but were not picked up by the PCA approach. The latter may be even more anomalous than those, that were selected using the PCA.

On the other hand, there are seasons that were identified as anomalous though none of the variables showed a strong anomaly. Could it be proved that they are 'true' anomalous seasons and not artefacts of the method?

I am not asking for a change of the approach here, but I think more discussion around potential (dis)advantages of the proposed method is needed. In my opinion, this method identifies the dominant modes of variability and allows for assessment of the contribution of each of 6 metrics into those modes and a link between them. Section 5 explores a few seasons when one of the first two modes of variability was among the strongest.

We appreciate and fully understand this remark. It is a priori not clear how extreme seasons should be defined. An obvious choice, which we also use in other studies, is to simply choose the warmest or wettest seasons. This would prioritize one parameter (e.g. temperature or precipitation) and a justification would be given why this parameter is particularly relevant. Here we tried something else, something more "objective" in the sense that we did not want to pre-specify the most relevant parameter. Instead, we allow for the possibility that besides individual parameters, also their combination can be unusual. Thus, we were led by the hypothesis that our multivariate approach will lead to different types of extreme seasons (different in terms of their individual anomalies of T_{2m}, P and E_S), which, however, share a similar "anomalousness" as expressed by the parameter d_M. We don't think that this method produces artefacts; in order to reach a value of d_M>2 (or even d_M>3), at least one of the considered variables or a combination thereof must be clearly exceptional compared to the other seasons in the ERA5 time period. In the revised version we will make sure that this line of thought becomes obvious to the reader. At the same time, we cannot (and don't want to) prove that this approach is "better" than a more conventional one. If all that matters in a specific study is for instance the seasonal snow accumulation, then there is no need to work with our approach.

2. My other concern is the length of the manuscript. Considering the amount of work, it is hard to make it shorter, but I think the paper will benefit from it. Some plots (especially, Fig. 3) are too busy and are difficult to interpret. Section 3 and 4, while interesting, are hard to read, particularly when plots discussed in the text are a couple of pages away (which is inevitable). Please select the most robust and/or important relationships and focus on them. I understand that each plot provides a lot of information, but, unfortunately, human beings can only keep a few facts in mind at a time.

We will do our best to further streamline the text and make it as readable as possible. With regard to the length of the paper, we think that it is still fairly OK.

Other comments:

1) Abstract: 1.The abstract is a bit long, even if there is no word limit, a page-long abstract is not ideal.

Thank you, we will shorten the abstract by about 20%.

2) I think it is worth mentioning that 2016/17 winter was mostly anomalous in terms of precipitation and maybe in some other variables, otherwise, until you read the paper, it remains unclear why it was anomalous.

Thank you for this remark. In the revised version of the abstract we write "In contrast, winter 2016/17 started with a strongly reduced sea ice coverage and enhanced sea surface temperatures in the Kara and Barents Seas. This preconditioning, together with increased frequencies of cold air outbreaks and cyclones, was responsible for the large upward surface heat flux anomalies and strongly increased precipitation during this extreme season." This makes it clear that DJF 2016/17 was mainly anomalous in terms of precipitation and surface heat fluxes.

3) Sect. 2.3: For the PCA analysis, was each metric first averaged over the corresponding region? Meaning that the special structure of those anomalies was not accounted for.

Yes, we average over the region and therefore lose information about the spatial structure.

4) Fig. 3: As I already mentioned above, it is a very busy plot, which is hard to read. The only thing that is obvious to me is that in JJA the red/blue markers can be linked to positive/negative temperature anomalies. For DJF, what is obvious is a link between T2m and P anomalies and that the low right corner has predominantly negative Es anomalies. However, regional differences, discussed in the text, are very hard to see. If you decide to keep this plot, maybe splitting into different geographical locations or the sea ice concentrations helps.

Thank you very much for the suggestion. We adapted the figure (Fig. R4 below shows the revised Fig. 3 of the paper), and now show the correlations for each SIC_{clim} range in separate panels.

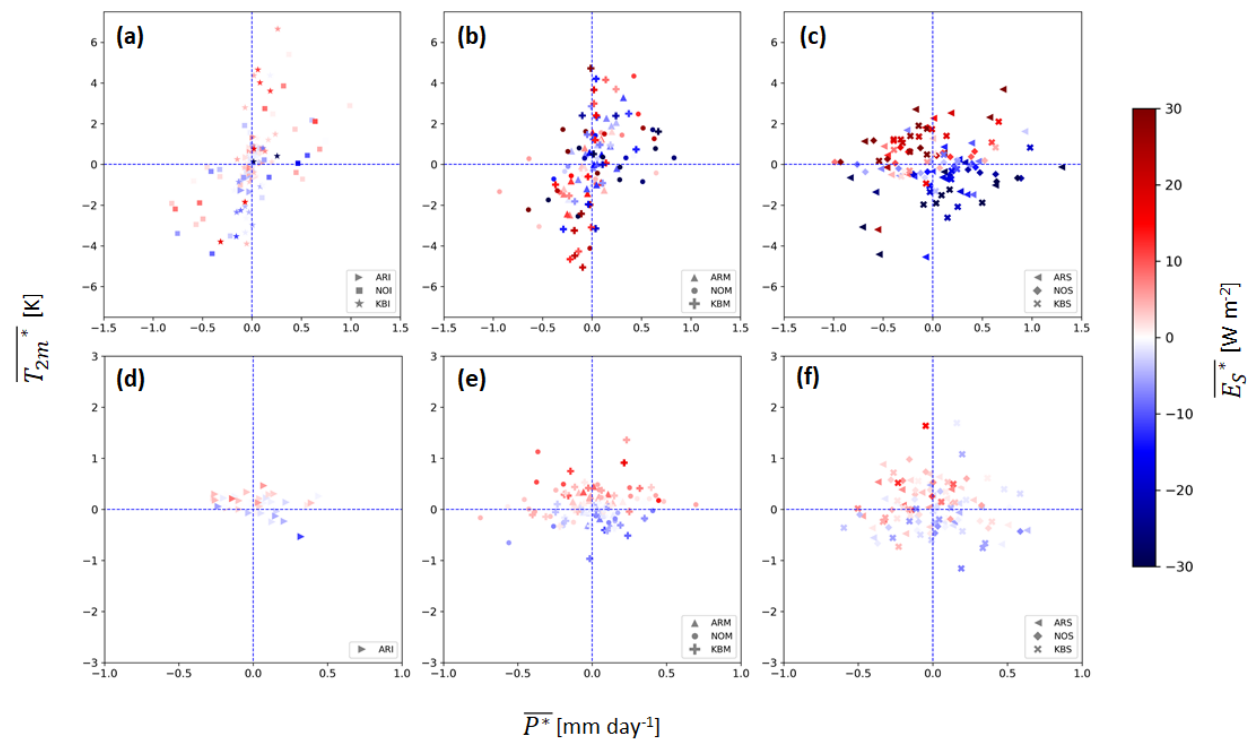


Figure R4: Seasonal-mean anomalies of P ($\overline{P}^*$; mm day^{-1}, along x-axis), T_{2m} ($\overline{T_{2m}}^*$; K, along y-axis) and E_S ($\overline{E_S}^*$; W m^{-2}, color) for 39 seasons in DJF (a,b,c) and JJA (d,e,f) for sub-regions with $SIC_{clim} > 0.9$ (a,d), $0.1 \le SIC_{clim} \le 0.9$ (b,e) and $SIC_{clim} < 0.1$ (c,f).

5) l.230: Despite good clustering in Fig. 4, this plot is again very busy. Maybe you can show the average location for each of the nine sub-regions on top of the existing plot.

Thanks for this suggestion. We added averages for each sub-region to Fig. 4 (see Fig. R5).

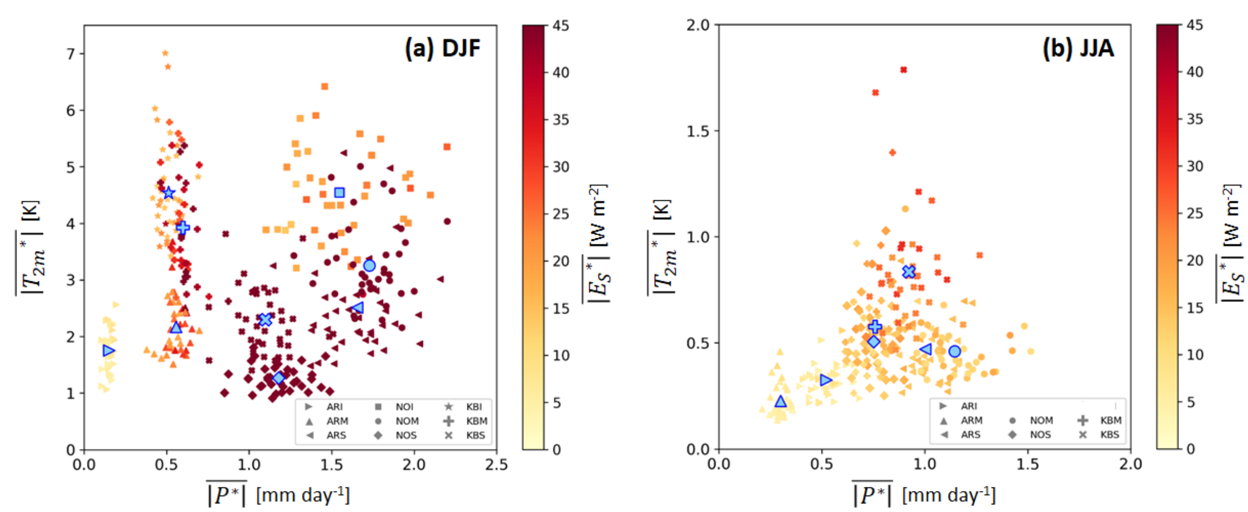


Figure R5: Seasonal-mean absolute anomalies for P ($\overline{|P^*|}$; mm day^{-1} ,along x-axis), T_{2m} ($\overline{|T_{2m}^*|}$; K, along y-axis) and E_S ($\overline{|E_S^*|}$; W m^{-2}, color) for 39 seasons and (a) 9 sub-regions in DJF and (b) 7 sub-regions in JJA. Blue symbols mark average values for each sub-region.

6) Fig.5, 'the seasonal-mean absolute anomalies': are these the seasonal-mean absolute *daily* anomalies, as in Fig. 4?

Yes. We define the seasonal-mean absolute anomalies as the seasonal mean of the absolute daily anomalies (lines 229-230), which is valid for all figures in Sect. 3.

7) Fig.5,6: why Nordic seas are not shown?

We do this in order to limit the length of the paper (see also your remark above), we selected some sub-regions for Fig. 5 (those that are most relevant for the case studies). The other sub-regions are now shown in the Supplement.

8) l.285-287: The statement on correlation between T2m and P comes from the fact that the corresponding blue lines are close to each other?

Yes. The more parallel two precursor vectors are, the stronger is the correlation of the precursors, provided that PC1 and PC2 explain a large part of the variance (Gabriel, 1971; 1972).

9) Regarding the comment on the weather systems creating extremes in the high Arctic, I would like to agree, though none of the AR seasons across all regions in fig. 5 look particularly extreme. How about other regions that have stronger extreme seasons often just in one parameter - can they be explained by anomalous weather patterns?

Yes, we agree, variability is distinctively smaller in the High Arctic in winter. But we think that one strength of our method is that it is still able to objectively quantify the "anomaly magnitude" of one season compared to another in a specific sub-region. A method using absolute thresholds would find most extreme seasons most likely at the poleward end of the storm tracks.

With respect to your 2nd question, we unfortunately don't understand what you mean by "other regions". For all seasons, which we investigated in detail, we found an important role of anomalous weather patterns.

10) l.303-307: Why the described connection between P and Rs over the sea, as well as between T2m and RL in KBM does not hold in the Kara-Barents Sea?

In **NOS** and **ARS** we can see an anti-correlation between P and R_S (corr(P,R_S)=-0.91 in **NOS** and corr(P,R_S)=-0.96 in **ARS**. Our argument for this anti-correlation is the presence of clouds during rainfall. As you correctly point out, there is no such anti-correlation in sub-region **KBS** (corr(P,R_S)=0.05). This would indicate that there is a large variation in cloud cover (and thus possible reduction in *Rs*) also during periods without precipitation. We did, however, not investigate this relationship in further detail.

We also want to point out that the mentioned correlations are only an approximation (see previous comment), which is more accurate the more of the total variance is explained by the first two Principal Components. As in the mentioned sub-regions this explained variance ranges between 85% and 88%, we assume that the correlations are good enough to use them for our interpretation.

11) Section 4: The relationship between 6 metrics during cold and warm seasons, gained from the PCA analysis, is interesting. Could correlations found in this section be confirmed by using the raw data?

We are not sure if we understand this question correctly. Yes, we can confirm some of the correlations found with the PCA. For example it is shown in Fig. 3 that in DJF T_{2m} and P are mostly positively correlated in **ice** and **mixed** sub-regions, whereas there is no such correlation in regions over the open ocean. This correlation can as well be seen in the PCA biplots for DJF in Fig. 7. However, as we use six different variables for the PCA analysis, and there seems to be no conventional method to illustrate the correlation of six variables, we only show T_{2m}, P and E_S in Fig. 3. It is further important to mention that we use detrended seasonal-mean anomalies for the PCA and thus remove the seasonality and a potential trend compared to the raw data. Therefore, it is not straightforward to compare the correlation of certain parameters for both data sets.

12) l. 322: “By design, extreme seasons have very large anomalies for at least one parameter… However, some anomalous seasons don’t show very strong anomalies in one particular parameter, which implies that for these seasons it is the combination of several parameters that makes them anomalous” I am not sure that the first sentence is true. Moderate anomalies in a few variables may also give an anomalous season and this is what happens in some cases.

Thank you for this remark. We will rephrase the mentioned section in order to clarify that indeed our **extreme** seasons have at least one large anomaly in one parameter (see previous comment about the approach) to reach a d_M value which is larger than 3. However, it is correct that this is not necessarily the case for **anomalous** seasons, where it is often the combination of several moderate anomalies resulting in d_M>2 (but smaller than 3).

13) l.367: I could not find a description of how cyclones, CAO and blocking events were defined.

For the specifics of the identification scheme, we will add a few sentences for each weather system and give the relevant references to the papers that introduced these schemes (see answer to specific comment (9) by reviewer 1).

14) l.372: Even during CAOs the temperature remained above the climatological mean, hence, I doubt that 38%-deficit in CAO can be responsible for the season being anomalous. During the first month (days1-27), there were no significant blocking events and CAOs, but T2m was well above average. To me it looks like there was a strong preconditioning. Furthermore, in the next

case, shown in Fig. 10, there is a high number of CAOs but they have relatively small effect on T2m, especially during the first half of the season, brings the temperature down by only, perhaps, 2-3 deg.

Thank you for this remark, it is certainly important to discuss this more thoroughly. As we state in section 5.3, where we discuss the synoptics throughout the winter 2011/12, one important feature is the pathway of the cyclones entering the Arctic from the North Atlantic, as they tend to slow down and get stationary in the region of the Nordic Seas, and their position relative to the Kara and Barents Seas. As a result, during several days of this winter, the warm sector of a cyclone is located in the Kara and Barents Seas whereas its cold sector is positioned in the Nordic Seas. This does not only explain partially the relative lack of CAOs, but also the overall increase in the surface temperature anomaly. If the cyclones were located further east, both the warm and the cold sectors would have been located in the region, likely resulting in no notable T_{2m} anomaly. Comparing the timeseries in Fig. 9 with the supplementary animation S1 shows that this synoptic situation especially occurs in December and in the second half of February, when the T_{2m} anomaly is very strong. For further studies it could thus be very useful to have a metric for the coverage of a region by a cyclones' warm sector as opposed to its cold sector (and thus the position of a cyclone with respect to that region). This would simplify the interpretation of a cyclones' influence on surface parameter anomalies in a distinct region. In the revised manuscript we will emphasize more that the impact of cyclones depends critically on their track relative to the region.

With regard to your comment on preconditioning in this season, we can say that this is most probably only a minor reason for the anomalous surface temperatures. Indeed, SON 2011 shows already slightly positive $\overline{T_{2m}}^{*}$ values and a slightly reduced SIC, but not to an extent that could explain the strong seasonal-mean T_{2m} anomaly during DJF 2011/12. The sea surface temperature reaches values of about +1-1.5 K above normal in September 2011, however returns to climatological values in October and shows no significant anomalies throughout November.

15) l.465: A seasonal blocking anomaly over Scandinavia is probably not enough to support the statement that 'Subsidence-induced warming [over Scandinavia] and long-range transport of warm air masses contributed to several warm episodes.'

This is indeed correct. To confirm this statement, we will add a short evaluation of some air parcel trajectories to the supplement, which show the importance of subsidence-induced warming and long-range transport during episodic warm events in DJF 2016/17.

Figure R6 shows air parcel trajectories for two warm episodes in DJF 2016/17 from 16-19 January 2017 (Fig. R6a) and from 11-14 February 2017 (Fig. R6b). In January, the influence of long-range transport of air parcels at lower levels, mainly from eastern Europe, can be observed. In February, subsiding air masses, favored by the presence of a blocking system over Scandinavia, additionally contribute to the warm event.

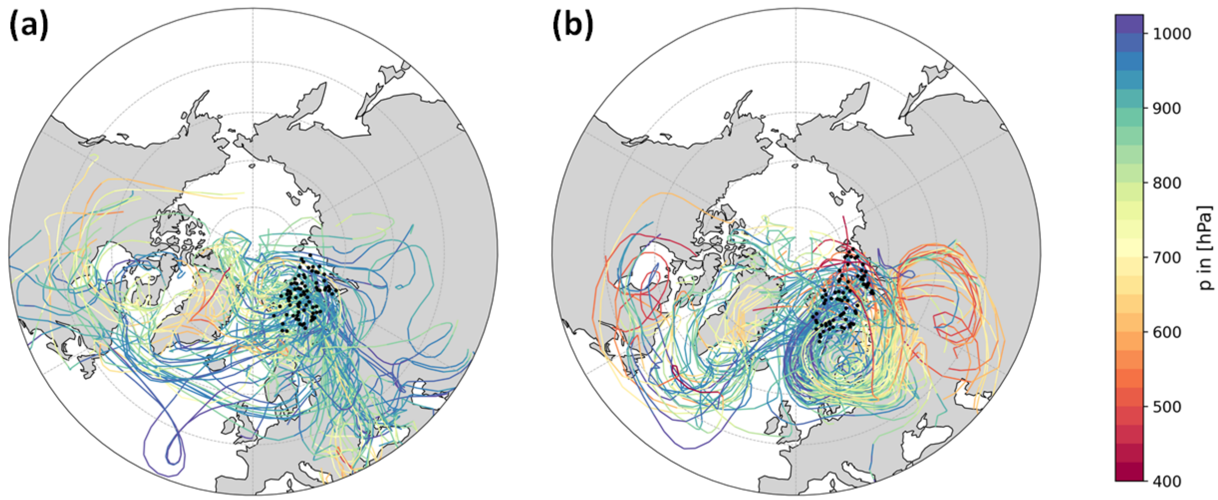


Figure R6: 10-day kinematic backward trajectories associated with positive daily mean T_{2m} anomalies in the region of the Kara and Barents Seas for the period (a) 16-19 January 2017 and (b) 11-14 February 2017 colored according to pressure. Trajectories are initialized every 6 hours at 25, 50, 75 and 100 hPa above ground for grid points with $T_{2m}^{*} \geq 2K$. Every 100th trajectory is shown with black dots denoting the starting point of each trajectory.

16) l.498: why a persistent high does not cause subsidence warming? and why there are no blocking events during Jan 2013 at the time of a persistent high? I can also see a number of cyclones in Feb, despite the text says that Feb was also calm. I agree that probably the main reason for decreasing t2m and low P is that the High Arctic remained isolated from the lower latitudes, however, none of the metrics in this study reflect an exchange between latitudes. I am not suggesting adding such metric at this stage, but it might be something to add in the future.

Thank you for these remarks. It would certainly be useful to have a measure which is indicative for latitudinal air mass exchange to better understand the processes leading to extreme seasons in the High Arctic.

Regarding your questions about the non-co-occurrence of the persistent high-pressure system as well as the lack of subsiding air, we analysed the geopotential height as well as the potential vorticity (PV) anomaly at upper levels throughout this winter. Figure R7 shows the geopotential height at 300 hPa (Z300) during the episode of the strong high-pressure system between 15 January and 25 January 2013 in the region of the Chukchi Sea and the High Arctic. Z300 does not show significantly enhanced values above the surface high, indicating that there is no strong upper-level forcing in the form of a persistent ridge which could have caused the formation of a block and the strong subsidence of air. The analysis of the vertically averaged potential vorticity anomaly (VAPVA) between 500 and 150 hPa does further support these results, as it reaches only small negative or even positive values in the same region (for the identification of a block following Sprenger et al. (2017), an area with VAPVA < -1.3pvu which persists for at least 5 days would be needed). Thus we assume that the strong high-pressure system at the surface is caused by very cold air below an inversion layer, decoupled from the synoptics in the upper troposphere. We can show that there exists a strong inversion layer very close to the surface in the center of the high pressure system by using a skew$T$-log$P$ diagram (see Fig. R8a), which supports our assumption that the air in this area experiences radiative cooling opposed to subsidence-induced adiabatic warming which one might expect in the presence of an upper-level block. Figure R8b shows skew$T$-log$P$ profiles at the edge of the high pressure system, which show a clear weakening of the inversion with increasing distance to the center.

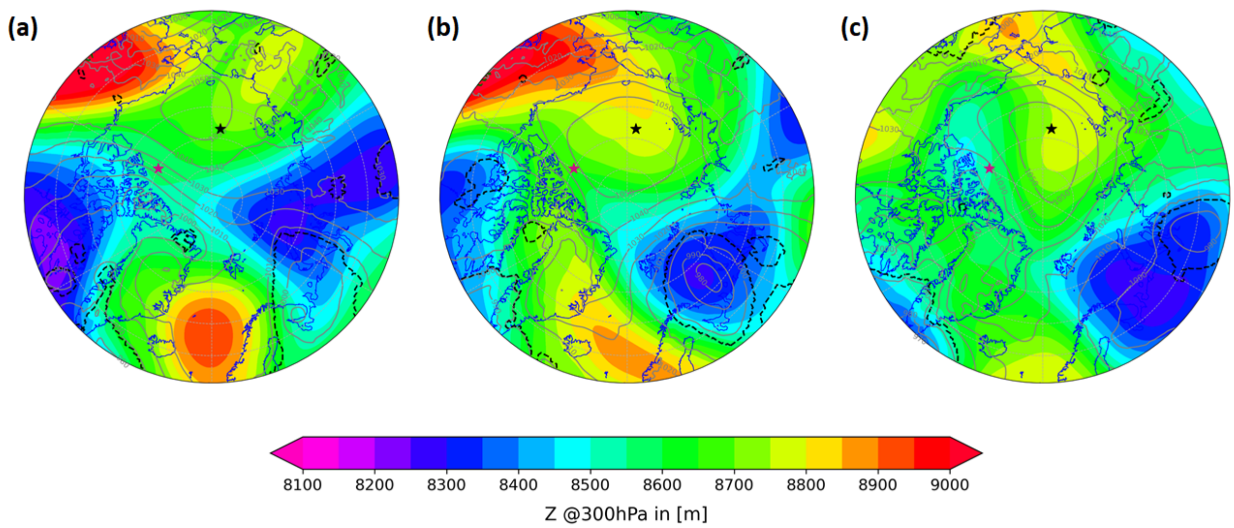


Figure R7: Synoptic situation on (a) 20 January 2013, (b) 22 January 2013, and (c) 24 January 2013. Daily mean geopotential height at 300 hPa (in hPa, color), instantaneous sea level pressure (grey contour, in intervals of 10 hPa), cyclone mask (dashed black contour) and blocking mask (dashed green contour) at 00 UTC. Black star at 173°E, 78.5°N and purple star at -120°E, 80°N show locations of skew$T$-log$P$ profiles in Fig. R8.

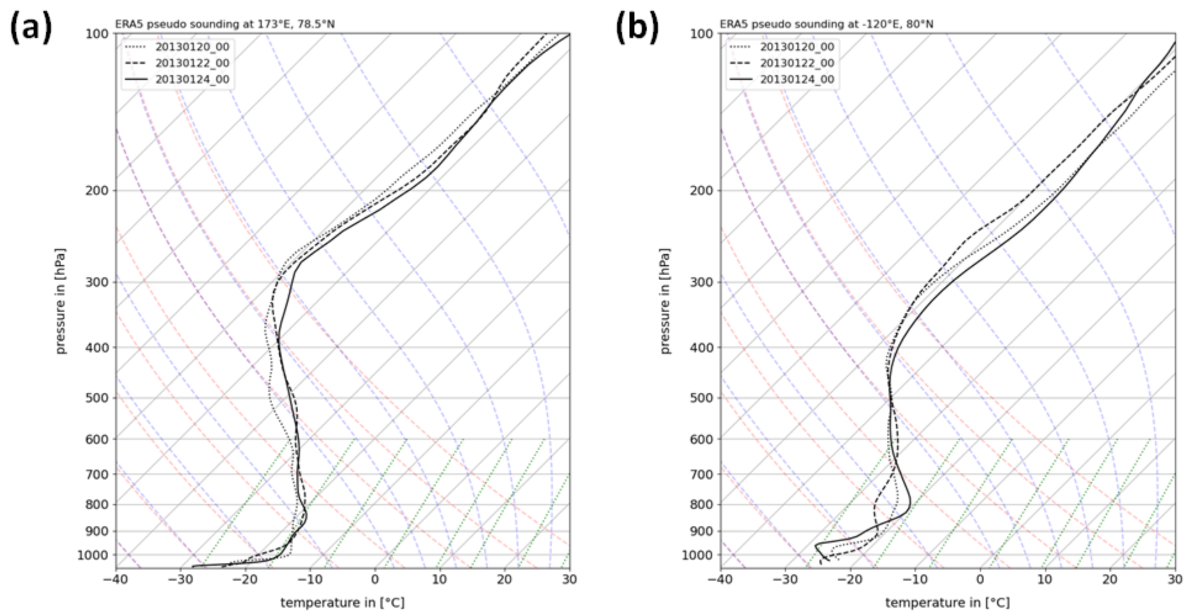


Figure R8: Skew*T*-log*P* diagram for dates shown in Fig. R6 at 00 UTC at (a) 173°E, 78.5°N (black star in Fig. R7) and (b) -120°E, 80°N (purple star in Fig. R7). Temperature is shown along the x-axis (in °C) and pressure along the y-axis (in hPa). Black lines show the ambient temperature profile for 20 January 2013 (dotted line), 22 January 2013 (dashed line), and 24 January 2013 (solid line). Grey lines show isobars (horizontal) and isotherms (skewed), respectively. Colored dashed lines denote dry (red) and moist (blue) adiabats, respectively. Green dotted lines denote constant saturation mixing ratios.

17) l.529-534: the paragraph first describes obvious seasonal differences (higher variability in winter due to stronger gradients) and then concludes ‘hence, it is reasonable to subdivide the Arctic into several regions considering these spatial differences to study anomalous Arctic winter seasons.’ But during summer the regions were also subdivided. I am not sure if this paragraph is needed at all.

Thank you very much for this remark. The mentioned paragraph is indeed a bit misleading and possibly not needed at all, which is why we will delete it. However, we still want to mention the difference in spatial variability between winter and summer and therefore will add 1-2 sentences in this regard to the previous paragraph.

18) l. 541: see my major comment on the PCA approach

See our response on p.11 of this reply document.

Minor comments:

1) l.61 ‘and of the feedback’: remove ‘of’

Changed “strongly affect the type of linkages between parameters and of the feedback processes” to “strongly affect the type of linkages between parameters as well as feedback processes”.

2) Table 1: Es should be added

Thank you for pointing this out, we added the variable “*Es*” to Table 1.

3) Table 2 is first mentioned in section 2.3 but is only shown in section 4. Replace ‘brackets’ with ‘parentheses’

Thank you, we changed “brackets” to “parentheses” in the Table caption.

Indeed we refer to Table 2 already in the method part to justify the detrending of our data set. However we prefer to show Table 2 only in the results part and not yet in the methods part as it basically shows the results of our analysis, based on the PCA biplots in Figs. 7 and 8.

4) l.160: it is not the entire ERA5 period, but the entire period covered by this paper

Thank you for pointing this out. We change the regarding sentence to “A distinction is made between areas where, on all days of the considered season in the time period covered by this study, mainly sea ice is present …”.

5) Please use either the Kara-Barents sea or the Kara and Barents seas

We now only use “Kara and Barents Seas”.

6) 406: I’d replace ‘single’ with ‘individual’

We replaced “single” by “individual”.

7) 432: on this date

Thank you, changed as requested.

8) Fig. 13 is mentioned earlier than fig. 12.

We will rephrase the mentioned part in line 437, to clarify that we only refer to section 5.3 here and not yet want to discuss Fig. 13.

9) Fig. 10,11, 14: I suggest showing months and days of months’ along Axis X, instead of days of season, as specific dates are often mentioned in the text (e.g, 9 Jan or 17 Feb). Can SLP be added to fig. 10,11? In fig. 14 the legend mentions CAOs, but they are not shown - could they be added?

Regarding days/dates: see our answer to specific comment (8) of Reviewer 1.

We added SLP to Figs. 10 and 11. Further we removed the CAO heatmap description from the caption. It does not make sense to show the marine air outbreak frequency for sub-region **ARI**, as this region is mainly ice-covered and as mentioned in the method section (lines 136-137) we define CAOs only for grid points with a sea ice concentration of less than 50%.

Reviewer 3

The authors have investigated seasonal extremes in the Arctic using PCA of six climate variables and analysis of some key dynamical elements – cyclones, blockings, and marine cold air outbreaks – to further investigate particular extreme seasons. This is an interesting and valuable framework for understanding the various causes of seasonal extremes, and it is very well presented. I recommend the manuscript for publication with some minor adjustments. My principal concerns relate to the justification of the many choices which needed to be made in this analysis, these are detailed below.

Specific comments:

1) L13: "respectively" – this doesn't quite follow when you say 2-3 extreme seasons for four seasons.

Thank you, we changed the wording to "...our approach identifies 2-3 extreme seasons for each of winter, spring, summer and autumn, with strongly differing characteristics…".

2) L15: I think a justification of why 2 winter seasons were chosen for the in-depth case studies is needed here.

See answer to general remark (4) by reviewer 1.

3) L117: It is very nice to have these questions in the Introduction to frame the paper, but as far as I could see the synoptic systems of interest are pre-defined in the study (cyclones, blockings, and marine CAOs), so perhaps this question should be reframed to reflect this.

This is indeed a good point. We will rephrase question 3: "In which way do synoptic-scale weather systems such as cyclones, blocks and marine cold air outbreaks determine the sub-structure of extreme seasons?"

4) L131: What was the method(s) of interpolation?

This interpolation is done by the ECMWF software when downloading the ERA5 fields from the MARS archive.

5) L155: What is the justification for choosing these regions?

As stated, a distinction between areas with differing sea-ice concentration is made, as surface heat fluxes and surface radiation are strongly dependent on the surface conditions. Further, we defined three different geographical regions, namely the Nordic Seas (**NO**), the Kara and Barents Seas (**KB**) and the remaining Arctic (**AR**). These are chosen based on the following main features: The **NO** region is the endpoint of the Atlantic storm track and important for deep water formation. The **KB** region has been strongly affected by changes in sea ice concentration and reacts very sensitively to atmospheric forcing. It is also a preferred region for atmospheric blocking and has its "own" storm track. Region **AR** is largely uncoupled from the mid-latitudes. Due to these different characteristics, it is useful to look at these regions separately when analysing the dynamical processes leading to Arctic extreme seasons.

6) L161: Are results sensitive to the choice of definition of ice, mixed, and sea? Why were these thresholds chosen?

The results are sensitive to the choice of the SIC thresholds when defining ice, mixed and sea, because obviously the resulting regions get larger or smaller depending on how the thresholds are changed. For instance, if for ice, the threshold SIC_{clim} was lowered from 0.9 to 0.8 then this would increase the size of the ice regions (and decrease the size of the mixed regions) and therefore the results for ice and mixed would be slightly less distinct. We decided to use relatively strict thresholds for ice and sea to ensure that these regions are indeed almost completely ice-covered and ice-free, respectively.

7) L174: Why choose just the first 2 PCs? This seems arbitrary, although I see later you mention that these explain a very large part of the overall variance.

See answer to first general comment of Reviewer 1. And yes, indeed the first two Principal Components explain usually 80-90% of the overall variance (in more detail: in 88% of the cases its >80% explained variance, in 53% of the cases they explain even >85% of the overall variance) and they are - for almost all regions and seasons - statistically distinct.

8) L178: Why are these rescaled by their respective SDs to give equal weight to each PC? Do you not wish to identify the extremeness of a season rather than the extremeness of a season with respect to these two PCs? (ref L114) If you don't do this rescaling do you still identify the same seasons as being extreme seasons?

We decided to use the scaled Euclidean distance (= Mahalanobis distance) in the PCA phase space to define our extreme seasons for the following reason:

With this approach, outliers in both, PC1 and PC2, are considered equally (without the rescaling, there would be more weight on the PC1 outliers). Thus, outliers in both PCs are treated similarly, independent of the individual variance explained by each PC.

9) Fig 8 and elsewhere: why was 10^5 km^2 chosen as the size threshold for a region?

This is a very pragmatic and subjective choice. Results from a PCA might be less reliable for very small regions. With this threshold, each region comprises at least 40 model grid points.

10) L393: grammar – "This periods typically are…"

Thank you, changed as suggested.

11) L408: “exemplarily shows” -> “exemplifies”

Thank you, changed as suggested.

12) L510: “as is the case”

Thank you, changed as suggested.

References:

Gabriel, K. R.: The biplot graphic display of matrices with application to principal component analysis, Biometrika, 58, 453–467,https://doi.org/https://doi.org/10.2307/2334381, 1971.

Gabriel, K. R.: Analysis of Meteorological Data by Means of Canonical Decomposition, J. Appl. Meteorol., 11, 1071–1077, https://doi.org/https://doi.org/10.1175/1520-0450(1972)011<1071:AOMDBM>2.0.CO;2, 1972.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, Sampling errors in the estimation of empirical orthogonal functions. Mon. Wea. Rev., 110, 699-706, 1982.

Röthlisberger, M., Hermann, M., Frei, C., Lehner, F., Fischer, E. M., Knutti, R., and Wernli, H.,A new framework for identifying and investigating seasonal climate extremes, J. Clim. (published online ahead of print 2021). Retrieved Jul 9, 2021, from https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-20-0953.1/JCLI-D-20-0953.1.xml, 2021.

Sprenger, M., Fragkoulidis, G., Binder, H., Croci-Maspoli, M., Graf, P., Grams, C. M., Knippertz, P., Madonna, E., Schemm, S., Škerlak,B., and Wernli, H.: Global climatologies of Eulerian and Lagrangian flow features based on ERA-Interim, B. Am. Meteorol. Soc., 98,1739–1748, https://doi.org/10.1175/BAMS-D-15-00299.1, 2017.

Tyrlis, E., Manzini, E., Bader, J., Ukita, J., Nakamura, H., and Matei, D.: Ural blocking driving extreme Arctic sea ice loss, cold Eurasia, and stratospheric vortex weakening in autumn and early winter 2016–2017, J. Geophys. Res.: Atmos., 124, 11 313–11 329, https://doi.org/10.1029/2019JD031085, 2019.