Review of Hartmuth et al.
Anonymous Referee #1

Referee comment on "Identification, characteristics, and dynamics of Arctic extreme seasons" by Katharina Hartmuth et al., Weather Clim. Dynam. Discuss., https://doi.org/10.5194/wcd-2021-18-RC1, 2021

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.

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

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.).

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.

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).
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.
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'?
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.
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.
6) The justification of how an extreme season is chosen on Lines 310-314 should be moved up to Section 2.3.
7) Line 315: Which season does Figure 2 show? This could also be referenced here along with Table 2.
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.
9) Line 367: How are blocking, cyclone, and CAO frequencies computed exactly? Need references and a short description.
10) Line 389: "Several episodic precipitation events..." But wouldn't Fig. 5h suggest consistent precipitation events?
11) Line 431: Remove "it is obvious that"
12) Line 440: Please also label JJA 2016 in Figs. 6 and 8.
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.
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?
15) Lines 120, 541: Is this approach really novel given that (Graf et al. 2017) first introduced it in a similar application?
Technical corrections:
1) Table 1: 2 m temperature --> 2-m temperature
2) Table S1: Caption states standardized values are in brackets, but they are instead in parentheses.
3) Section 2 should be "Data and methods" given that there is more than one method used to complete the analysis.
4) Figure 1 caption: State what the green and red boxes denote.
5) Line 135: There does not need to be a space between the number and the "%" symbol
6) Lines 140-141: What is the sign convention for the surface energy balance?
7) Line 183: There should be a period at the end of the equation.
8) Lines 352, 465: normal --> average
9) Line 393: This --> These
10) Line 424: Remove "of"
11) Line 453: Insert "of" after "Comparison"
12) Line 463: got --> became

References:
