Review notes for the manuscript to ACP: "Wintertime Extreme Warming Events in the High Arctic: Characteristics, Drivers, Trends, and the Role of Atmospheric Rivers" by Ma et al.

General comments

This paper presents characteristics of extreme wintertime Arctic extreme events for 1979 – 2021 utilizing hourly ERA5 data. Extreme warm events are defined per grid-point when the hourly near-surface temperature (T2m) reaches or exceeds a threshold of 0 ℃ in the high Arctic (north of 80°N). The strict spatial and threshold criteria restrict the events only to be found in the Atlantic sector. The authors find that these events are rare and short-lived, usually lasting less than one day. The highest frequency and longest duration of these events are found to be confined to a region close to 80°N and within 0-30°E. These events are further associated with positive anomalies in sensible heat fluxes (largest contribution closest to the lower latitude boundary) and positive anomalies in both IWV and DLW, with largest contributions especially further towards the Arctic interior. Large-scale circulation anomalies with a well-known dipole pattern in the SLP anomalies are found to be favourable for warm and moist air advection into the affected grid-points. A 100% match is found between warming events at northerly latitudes and the co-occurrence of atmospheric rivers (ARs). The grid-point wise defined warming events are then elaborated to regions of several grid-points with temperatures at or above 0 ℃, i.e., "concurrent warming events", a method taking into account the spatial extent of such events. Three different large-scale circulation patterns are found to be associated with these regionally defined warming events: a strong dipole in SLP, a blocking-like surface anticyclone or a strong negative SLP anomaly over Greenland. The authors finalize the paper by a trend analysis of grid-point wise defined warming events, where a positive trend in both duration, magnitude and frequency of the events are found. This paper has well-done figures and gives a nice overview of extreme warming events (per grid-point) in the high Arctic, and discusses their drivers with respect to the changes in e.g., SEB anomalies and their relation to large-scale circulation and AR frequency. Most of the results presented here agree and follow results from previous studies. I still have a few concerns regarding the event definition and interpretation of the results presented here. Thus, I suggest minor revision before any possible acceptation of the paper. See comments below for clarifications.

Specific comments

- The authors aim to make a climatological record of warm extreme events and associate these to large-scale circulation patterns and ARs. But this is something that has already been done by previous studies, however, with different dataset (ERA-Interim) and event definitions. For example, Graham et al. (2017) (https://doi.org/10.1002/2017GL073395), a study that is also cited in this paper, do not only use in-situ observations, but also look at the historical record winter warming events in the Arctic using different temperature thresholds. They also find a positive trend in duration and occurrence of warming events. Also, your Fig. 1b is similar to their Fig. 2a, showing maximum 2m temperatures, however they use 6-hourly ERA-Interim data compared to yours hourly ERA5 data. Another study that is completely left out of this paper is Messori et al. (2018) (https://doi.org/10.1175/JCLI-D-17-0386.1), where the drivers behind warm extreme events in the high Arctic are also examined. Please refer to these two studies in your paper for a comparison to similar studies. What is novel in your study compared to theirs? What is the reason for an event definition with a fixed threshold in your study?

- *Warming event definition*:
    - I also find it a bit concerning, as raised by the first reviewer regarding the event definition, that warming events, as the authors name these extremes, are defined based on a *fixed temperature threshold* (2-m temperature >= 0°C) and for *each grid-point* separately. What if grid-points close to each other actually belong to the same event that is affected by the same synoptic weather system? Grid-point defined warming events, using a Eulerian approach, don't say that much about the real nature of the synoptic weather event, as this event, let's call it an AR guided by the large-scale circulation, will move and affect grid-points further away in the direction of the AR, leaving the previously affected grid-point to cool. As the authors point out, there can be temperature fluctuations in hourly data (so that the temperature shortly drops below 0°C), but still be part of the same synoptic event (which has a different impact on the surface temperature depending on the time scale and location). The extent of these intrusions can then also

affect adjacent grid-points at the same time step, however, in your study, these two would be counted as two separate events? The authors explain on L98 the method "interval requirement", and state that your results are not affected by the temperature fluctuations. Could you please re-explain how this was done? Could you think of using a longer time span of 4-6 days (representing the synoptic scale) for your requirement? Thus, having less than two days between two warming events (as stated in L168-170) would most likely just be the same event. How about the close-by grids? How often do the authors find that several grid-points experience temperatures above zero degrees at the same time? One way to avoid these issues would be to maybe increase the time from hourly to daily and check for days where the temperature exceeds a threshold in the grid for at least once or up to a certain percentage within the time? The authors could also consider looking at pre-defined regions instead of per grid-point. In Messori et al. (2018), warm extreme events were defined over the polar cap (as in this study) as daily T2m anomalies computed against a transient climatology (long-term trends and seasonality removed) and area-weighted over the study domain. Furthermore, anomalies were smoothed to remove fluctuations and only events at least one week apart were chosen in order to avoid double counting of the same event. Despite this, when considering large-scale drivers, in a study from Murto et al. (2022; also referred to in the current paper), they could associate up to 3 of the warm extreme events from Messori et al. (2018) to one or two consecutive blocking events, suggesting that these warm events were actually one event but affected by a similar large-scale setting. Did the authors of this current paper try to define events using temperature anomalies? The event definition as it is now is for me a bit problematic, and caution must be taken when interpreting the results of your study. Please also point out in the abstract that events are defined per grid point.

- Another issue that arises when defining events by a Eulerian perspective is to refer to the "lifecycle" of the event, i.e., with an *onset and decay*. As written above, ARs tend to move and thereby at a future timestep will affect nearby grid-points. When talking about "onset" and "decay" of an event, it refers to some phenomenon that usually moves and which can be tracked. Thus, the same AR can continue to the next point, but it does not mean that the event has "decayed" when the temperature drops back below zero degrees at another grid-point. I would maybe consider using another word usage here instead of "onset" and "decay" when referring to the warming events at a point. Maybe use "time when a grid-point's T2m exceeds zero degrees"? Close-in-time warming events at one (or nearby) grid-point and the risk of double counting events is then further accumulated in lead/lag composite plots, as in Figs. 6-8. For example, as stated in L230, anomalies 6-days prior to the onset of a warming event could just be a result of a warming event that was present at that lag-time. A timestep at the onset of one event could be same time as the decay of another event. In Fig. 6 it is also clear that the warm air is advected to other regions, which shows that heat is transported further into the Arctic, leading to a temperature drop (and as you name it, decay of the event) at a grid-point when the source of the heat is moved (Woods and Caballero (2016) found, for example, that it takes five days for a moist intrusion to cross the Arctic interior, which is within the timeframe of the event duration here).
- *Spatial restriction*: Why did the authors decide to restrict to the polar cap (north of 80N), when events are defined from absolute temperatures? Studies found that warm extremes are mainly located in the Atlantic sector (Graham et al. 2017) and associated with strong moist-air intrusion from the Atlantic that penetrate into the high Arctic (Messori et al. 2018). With a more relaxed temperature threshold, as done in Graham et al. 2017, Pacific warming events can also occur. I would like to see some discussion to why Pacific warming events are not included in this study. For example, sea ice extends further south at the Pacific side and thus airmasses from southerly latitudes have a longer path to cool before reaching into the polar cap compared to Atlantic pathways. A northerly ice edge at the Atlantic side, on the other side, allows the air to collect moisture and heat for a longer time and distance before losing them while traveling northwards. Storm tracks might also be more active over the Atlantic side. In Graham et al. (2017), a southerly latitude band was also chosen for the Pacific side compared to Atlantic to include both sectors with warming sources. Studies also show that ARs from the Pacific side, such as in 2007, can

have an important impact on the temperatures in the Arctic. Southerly winds promoted ice-export and the warm and moist air transport over the Beaufort Sea enhanced surface temperatures, and led to anomalous SEB fluxes and the onset of sea ice melt (e.g., Graversen et al. 2011 https://doi.org/10.1007/s00382-010-0809-z and Stroeve et al. 2008 https://doi.org/10.1029/2008EO020001). Please add a bit more discussion around the reason for the chosen study area (distance to sea ice, importance of both transport pathways...) in the introduction and around L147.

- *Concurrent warming events:* (Sect. 3.3). For me, this method seems much better, as, to my understanding, these events are defined by finding areas in the Atlantic sector where several grid-points at one timestep satisfy the criterion for warming events. What is the region where these events occur (maybe show on a map figure)? Are these areas coherent, i.e., that the grid points are adjacent to each other? What are the main characteristics for these events (number of points included in one area, duration, trends, seasonal evolution, spatial frequency…)? The lifecycle perspective of these events is again utilized here (onset and decay), which rises similar questions as stated above. How is the decay defined (when the temperature for all or at least one grid point within the area drops below zero)? Again, as stated in L258, close-by areas can also be affected by the same weather event, and here the authors decide to use a 5-day temporal limit, which is good. How is the temporal criterion applied here (on warming events (grid-points) within a region or for separate regions)? The definition of the peak of these events is somewhat unclear: on L263 it is referred to as the hourly time when the area of the grid-points with T2m>=0℃ is the largest, but at L281 the authors refer to the peak time of the T2m anomaly when referring to Fig.10. Does the time at maximum area of grids satisfying the criterion always correspond with maximum T2m anomalies? Please clarify. If this is not the case, maybe an intensity measure where both the spatial extent (area) and the average temperature anomaly (magnitude) could be used to define the peak.
- How much does the *surface type* (ocean, sea ice, leads in sea ice) affect the spatial location of the warming events, and the anomalies in e.g., SEB? Did the authors consider to divide the warming events into ocean and sea ice (with different SIC's) – this could improve the quality of the paper and the interpretations of the results. L185 states that the climatological SHF is upward in winter, which is true over warm ocean or openings in the sea ice. But over sea ice, I think that the climatological values of turbulent fluxes are almost negligible, if not slightly positive. The surface type influence can nicely be seen in S2, with a sharp division between negative and near neutral values at the sea ice edge. Thus, L187 "suppression of the upward THF" is true if the surface is relatively warm, but ARs could also enhance downward THF if the surface is relatively cold (e.g., over sea ice). Discussion about the impact of different surface types on the results shown here would be nice to see and adding the sea ice edge on the climatological plots would be helpful.

- *Wording* and terminology for the warming events: The authors are not always consistent with the wording of their events: warming events or extreme warming events for the grid-point wise defined events, and concurrent warming events or large-scale events (L378) for a larger region. Please be consistent with the terminology, and always refer to which events (region or grid-point) the analysis shown in a figure or discussed in the text is referring to. For example, add it to the abstract and conclusions, as well as in figure captions (such as in Fig. 5 for the duration categories). In the abstract it is, for example, not clearly stated that the match between ARs and warming events are done on the concurrent events.

- The word "driving" or "driver" is used several times in the paper. I would rather write "associated with" or "related to" instead of driving. For example, in L300, I would rather write that the SLP pattern are "guiding" the ARs or "making a pathway".

- Figure 2c and Fig. 3a: remove the value shown in the upper left corner but keep it in the captions. Also, I assume the caption for Fig. 3 for the red line should refer to panel (a), not c. I would also suggest to change the color for zero in Fig. 3b to another color to be clearer (blue according to the colorbar means no days or actually more than 0?).

- I suggest the authors to add more references, at least when discussing the results of this paper and when comparing directly to other observations or studies (e.g., in L153).

- At L158: I assume the authors mean that the duration drops gradually away from the region at 80N, 0-30W to less than 5h north of 85N OR east of 60E (there are no data in the corner north of 85N and east of 60E, right?)

- Figure 4: Would be interesting to see how many grid-points satisfy the temperature criterion over all seasons within the study period, to know if the average e.g., DLW is a result of only a few or several events. The average occurrence of these events (S1) gives some idea about this, but maybe a relative frequency plot would better demonstrate this (so that 100% would refer to the max occurrence of events per grid point over all periods). I also find the Figure S1a to be relevant for the main paper, as main characteristics for events include spatial distribution.

- I find it interesting to see that you find a relationship between the duration of the events, their locations and associated SEB anomalies. Have you performed any correlation analysis to make your statement stronger? I assume that we would then find higher correlation between longer-lasting events (longer duration) and SHF anomalies, whereas shorter-lived events would be correlated with DLW, especially at higher latitudes. Intuitively, I would think longer lasting events are those that penetrate further into the Arctic, but this confusing comes only because of your different way of defining events and their duration (at a fixed point; your longer-lasting events are closer to the warm and moist air source). Maybe worth reminding the reader again what your definition is to avoid possible confusion.

- When talking about anomalies in SLP or SEB components, for example, we are mainly interested in the sign of the anomaly. Therefore, I would suggest writing “positive/negative” instead of “high/low” when writing about anomalies.

- Figure 5: what is the absolute number of events within each sub-category based on duration? A map showing the locations of these grid-points would be nice to see. Further point out in Fig. 5 caption that the durations are defined per grid-point. Could the stronger anomalies shown in Fig. 7 compared to the other ones be a result of less events included in the composite?

- Figs 6-8: these figures are nice but also a bit messy with so many black lines. I would suggest changing the color for the SLP anomaly contours to purple (as in Fig. 13a).

- How do the authors think that the spatial patterns of the events (or occurrence and duration) might directly be affected by cyclones or blocking? Adjacent grid-points could be captured by the same cyclone but affected by either the cold advection or the warm sector (L206). Have you looked at cyclone tracks or used blocking detection algorithms to associate your warming events with them (instead of using ARs and SLP anomalies to represent the large-scale setting)? The authors also write “blocking” (L380) in the conclusions despite not using a blocking detection algorithm (?). I would suggest rewriting to “blocking-like structure” and maybe refer to other studies (e.g., Woods and Caballero 2016, Messori et al. 2018 and Murto et al. 2022).

- In sect. 3.4, why do the authors return back to the grid-point defined warming events to relate them to ARs? Have you looked at how your results would change when using the concurrent warming events?

- L325: are these nine days of T2m above zero degrees found here in your study? Please mark their locations on a map, e.g., in Fig. 13a. Have you studied the origin of the ARs or utilized the tracking algorithm of the AR shapes, if it was provided with the AR algorithm?

- Figure 10: Are the T2m anomalies shown for all grid-points independent of if the grids are part of a concurrent warming event (or only that the time is at the peak of the event)? Maybe some density lines would be helpful to show where these concurrent warming events are spatially located and how these anomalies extend wrt the originally defined events.

- One additional concern is the statement at L318: “ARs are the *only* weather system capable of triggering the occurrence of the warming events”. ARs are definitely important! That grid-points are co-occurring in time and space with ARs (even though 100% overlap), does not, however, directly imply that these warm anomalies can only result from an AR. In a recently published paper (Papritz et al. 2023 https://doi.org/10.1175/JCLI-D-22-0883.1), the relative contributions to the warm potential temperature anomalies extending in the whole tropospheric column (associated with extreme positive SEB anomalies over wintertime Arctic sea ice) were investigated. They found (using backward trajectories) that only airmasses ending up in the middle troposphere had an AR-like evolution, whereas airmasses making up the positive anomalies closest to the surface actually had an Arctic origin! These airmasses were either warmed diabatically while crossing over warmer oceans or when airmasses descend from higher altitudes, but all within the Arctic. It was these two airmasses together that could give rise to these anomalous positive vertically extending potential temperature anomalies. Local processes are thus also important, so I would suggest rewriting this strong sentence and add “likely” or “strong impact” instead of stating ARs are the only driver.

- Are the trends discussed in Sect. 3.5 based on grid-point defined warming events? Referring here to an absolute number of warming events (2150) does not tell the reader so much, as your events (and the number) are dependent on the grid-size, the temporal resolution of the data etc. Maybe more informative would be to show seasonal trends of days with atleast one warming event per grid point? Or rather use the concurrent warming events here and calculate their trends per decade. I also think that the event definition (a fixed temperature threshold) and with a rapidly warming Arctic, warmer temperatures become more common by default (compared to the first decade of your study period).

- Figure 13c is not referred to in the main text

- I am also lacking a final concluding statement. The authors nicely summarize the findings of the paper in the final section, and list some limitations. What potential of further studies would your study contribute? One way to tie the final section to the rest of the paper is to answer the questions raised in the introduction in the conclusions.

- Please add some discussion about ERA5 warm bias to the limitations of this study (relate to the representation of snow and sea ice in ERA5, see e.g., Batrak and Müller 2019 https://doi.org/10.1038/s41467-019-11975-3).

Technical corrections

I agree on the typing errors pointed out by the first reviewer. Below some minor correction suggestions in addition to them:

- Add a reference to panel (f) for the temperature advection in the figure (Fig. 11) caption
- I would suggest to add the latitude threshold in the caption for the long-lasting events in Fig. 5g, h
- Be consistent with the terminology used in the text and in the figures. In Figs. 5a and 14a, “TS” is used instead of “T2m” (I assume), it is always referred to as T2m in the main text.