Anonymous during peer-review: **Yes** No
Anonymous in acknowledgements of published article: **Yes** No

Recommendation to the editor

1) Scientific significance Does the manuscript represent a substantial contribution to scientific progress within the scope of this journal (substantial new concepts, ideas, methods, or data)?	Outstanding **Excellent** Good Fair Low
2) Scientific quality Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)?	Outstanding Excellent **Good** Fair Low
3) Presentation quality Are the scientific results and conclusions presented in a clear, concise, and well structured way (number and quality of figures/tables, appropriate use of English language)?	Outstanding Excellent **Good** Fair Low

For final publication, the manuscript should be

accepted as is

accepted subject to **technical corrections**

accepted subject to minor revisions

reconsidered after **major revisions**

rejected

Were a revised manuscript to be sent for another round of reviews:

I would be willing to review the revised manuscript.

I would not be willing to review the revised manuscript.

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The authors have replied to many of my remaining comments, albeit not always in a very systematic fashion. Moreover, some of the replies to the comments are not reflected in the main text. Many of the questions that a reviewer has may also be shared by readers, and there is limited benefit from the peer-review process if additional analyses or discussion only appear in the replies to reviewers. I provide some further suggestions below, including highlighting a number of previous comments that have not been fully addressed.

1. In reference to my previous comment number 3, I do not find this to be a particularly

satisfactory answer. Draxler (1987) writes that “Tests […] showed that increased temporal resolution did increase trajectory accuracy” and that: ”temporal enhancement of the meteorology is better than spatial enhancement”. I struggle to see how, based on these results, the authors can claim that. ”the increase of only temporal or spatial resolution does not necessarily improve the accuracy of trajectory calculation”. It is true that the benefit of improving one of the two depends on the resolution of the other (e.g. improved temporal resolution is less effective if the spatial resolution is very low), but the statement the authors make is at best twisting the results of the study they quote. I would also encourage the authors to dedicate some more time to a literature search on the topic of spatio-temporal resolution and air-parcel tracking, rather than relying on myself conducting this on their behalf. A quick check returns more recent studies by Scheele and colleagues and by Hoffmann and colleagues, the latter stating for example that: ”Part of the differences between [trajectories in] ERA5 and ERA-Interim is attributed to the better spatial and temporal resolution of the ERA5 reanalysis”. I would (again) also encourage the authors to justify their choice of data resolution with reference to past studies. For example, one of Stohl’s studies states that “a minimum resolution of 6 h is necessary if any diurnal variations in the flow field are to be resolved”, which would support the authors’ choice of 6-h data.

Thanks for the comments. We have considered the reviewer’s opinion and modified the corresponding paragraph in the paper lines 131-137.

It is to be noted that both the temporal and spatial resolutions may increase the accuracy of the trajectory calculation (Draxler, 1987; Kahl and Samson, 1986; Stohl et al., 1995). However, it is less effective to only improve the temporal resolution, if the spatial resolution is very low (Stohl et al., 1995). Minimally, a 6 h temporal resolution is needed to resolve diurnal variations in the wind field (Stohl et al., 1995), supporting the temporal resolution used in this paper. As the error of trajectory calculation increases exponentially with time, in this study, we calculate the trajectories 2 days forward and backward, instead of calculating 4 days trajectory at once.

2. In reference to my previous comment no. 5. I appreciate the authors having conducted this additional check, which is very relevant for readers wishing to study the same topic and evaluating what requirements to place on the trajectories. As a note, 45+28 = 73, the latter being over 13% of the total number of trajectories the authors retain. I do not deem this to be a negligible amount. I would encourage the authors to include the numbers they have computed in the text, to give the readers a feel for the importance of the different criteria. If they really wanted to be thorough, the authors could provide a table in an Appendix or Supplement with these numbers for all sectors.

Thanks for the review’s opinions here. We have added the information into lines 117-124.

Taking Barents Sea sector as an example, we have checked how strict requirement 1 is, by counting how many trajectories turn southward before they reach 85°N. According to our calculation, there are 45 (8.2%) trajectories tracking southward before they reach 85°N. Similarly, we have also checked how strict requirement 2 is, by calculating the number of trajectories which are all the way north and reach 85°N but the terminal point is less than 5 ° north of the sea ice edge. The results show that only 28 (5%) trajectories are in this case. Actually the strictest requirements is requirement 3. Around 59% trajectories cannot meet this requirement but this requirement is necessary since we want to look at how the air column evolve on its way to the central Arctic over the sea ice.

3. ll. 265-270 I appreciate this addition, but would recommend that the authors proof-read it carefully. As it stands, it is poorly formulated and with numerous grammatical errors (number agreement, adverbs in place of adjectives, punctuation etc.).

Thanks for this opinion. We have read through the text again and tried to improve the language.

The boundary-layer energy-budget pattern is very variable from case to case, mainly because the northward component of the advection is differently from case to case. Additionally, the location of the ice edge is also different from case to case. Some trajectories are long but reach less far north while others are shorter but still reach further north. In the vertical, the cases are also subject to different subsidence, affecting the boundary-layer growth. We therefore have not yet come up with a workable idea that would allow an ensemble average of all the cases.

4. In reference to my previous comment no. 6, once again I would like to see a more thought-out reply. There is literature on the role of vertical interpolation in trajectory schemes, which could be used to support the methodology adopted by the authors.

Thanks for the input here. The vertical interpolation does introduce new errors. We have clarify this in the paper.

Errors are also introduced by the vertical interpolation from pressure level to geometric height. The vertical interpolation of vertical velocity produces larger errors than the vertical interpolation of horizontal components (Stohl et al., 1995).

5. l. 123 ”would”. ”May”, or even “will” would be more appropriate here.

Thanks for the comments. We have replaced ‘would’ by ‘may’.

6. ll. 186-188 This is a rather lax definition of statistical significance. What is the theoretical justification for such an approach, versus using one the widely adopted statistical significance tests with a given significance level.

Thanks for the comment. We have clarified the definition of ‘significantly positive’ at lines 198-200.

Here, if the mean values of these surface energy-budget terms are positive and they are still greater than 0 after deducting their standard deviations, then we consider they are statistically significantly positive. This definition is quite lax, since it only passes 0.32 student significance test.

7. ll. 282-283 Table S1 is an excellent addition, but only lists the trajectories for the Barents Sea. Would it not be appropriate to list all sectors, and especially the Kara Sea sector which, as the authors themselves note, displays characteristics of both “ocean sector with land-locked sea ice and open ocean”?

Thanks for the comment here. We have added another table (table S2) for Sea sectors with land-locked sea ice.

8. Captions to Figures 2 and 8. In general, I would not call what you show a regression, but rather a regression coefficient. I would suggest updating the captions accordingly, including where you refer to the units of the regression. Also, please correct the typo in the caption to fig. 2: ”regrssion”

Thanks for the comment. We have replaced ‘regression coefficient’ to ‘regression’.