This paper demonstrates mosaics of winter ice surface velocities for the 1990's over the Eastern Arctic (Novaya Zemlya, Franz-Josef-Land, Severnaya Zemlya and Svalbard) through using the offset tracking approach on historical SAR data. Both the JERS-1 SAR data (primary) and the ERS-1/2 SAR/InSAR data (secondary) are used to generate the 1990's velocities. The authors also studied the long-term variability of winter ice surface velocity from the 1990's by comparing to mosaics derived from ALOS PALSAR in 2008-2011 and Sentinel-1 in 2020-2021. The paper generally reads well and compensate the existing ice velocity products on the knowledge of the ice surface velocity in 1990's. However, we found a few fundamental problems and also suspect the paper in its current form is not fitting the scope of ESSD well. Please refer to our following comments.

Major comments:

1. ESSD mainly requests the description, processing methods and validation of the dataset, rather than the data interpretation or analysis. Please refer to the website: https://www.earth-system-science-data.net/about/aims_and_scope.html. The third paragraph is copied here (with the specific line highlighted):
   "Articles in the data section may pertain to the planning, instrumentation, and execution of experiments or collection of data. Any interpretation of data is outside the scope of regular articles. Articles on methods describe nontrivial statistical and other methods employed (e.g. to filter, normalize, or convert raw data to primary published data) as well as nontrivial instrumentation or operational methods. Any comparison to other methods is beyond the scope of regular articles."
   This paper in its current form only has Section 3 on the data description, which is not very clear on its complex data structure either. However, the majority of the paper is on the data interpretation or analysis in the field of glacier changes (e.g. frontal line retreat/advance, surge, instability, etc) over the four test sites, respectively, e.g. almost the entire Section 4 and Section 5 are on such glacier analysis. We thus recommend the authors to consider publishing this paper in a regular research journal such as Cryosphere not in a data journal like ESSD.

2. The key idea in this paper is to publish velocity mosaics in 1990's by grouping 7-year (1991-1998) velocity products derived from JERS-1/ERS SAR and InSAR data, and consider the winter velocities do not change much compared to the annual mean velocity, which is claimed by the authors to be justified by using Sentinel-1 time-series velocity products (since 2014). In our opinion, this is not a convincing assumption. Not to mention the paper shows the Svalbard site tend to violate the assumption when including more surging glaciers, even for the other three sites, the Sentinel-1 data that were used to justify the assumption were collected almost two decades later than the 1990's cases. It is thus risky to make this assumption. In contrast, why not just publish a time series of velocity mosaics (rather than a multi-year mean) for each of the four sites, which could have a temporal resolution of 1 year or shorter depending on the JERS-1/ERS data availability? In our opinion, this would be more meaningful to the cryosphere community facilitating more flexible needs of the researchers in this field.

3. Data description section (Section 3) is not clear on the complex structure of the dataset. It would be great to rewrite it using graphs or tables for the readers' convenience.

4. The processing methods as described in Section 2.1 and 2.2 are too general. As shown in Table 1, the JERS-1 and ERS data products have different setting of processing parameters, as well as compared to the ALOS and Sentinel-1 data products. You need to clarify why the processing parameters were chosen as such and what impact these different parameters (e.g. template size) would have on the accuracy in the final comparison.

5. Inter-comparison of the dataset with other similar products is required by ESSD. Please see Sect. 3.5 of https://essd.copernicus.org/articles/10/2275/2018/. You seem to compare the 1990's mosaic with ALOS and Sentinel-1 a lot in this paper, but that comparison was only to show the long-term variation. Even though there was a lack of other satellite data back to the 1990's, you still need to validate the product with some contemporary measurements (e.g. GPS) or over static flow and/or rocks. Without such validation, readers do not know how accurate or confident to use this data product.

Below we list the detailed comments:

1. Line 14: in Appendix A, you only used the Sentinel-1 data at one or two epochs as the 2020-2021 maps. Unlike the abundant acquisitions from JERS/ERS and ALOS, why not include more data spanning the whole year of 2020-2021?

2. Line 72-77: better to tabulate these with the numbers for each sensor

3. Line 81-82: the sequence you mentioned the steps sounds like you have geocoding done before coregistration and offset tracking. Do you run offset tracking over already-geocoded images or the other way around?

4. Line 88: what is the effect of using various DEM with differing resolution?

5. Section 2.1: you should clearly mention the product is temporally averaged across few years, e.g. 1991-1998.

6. Table 1: refer to the major comment #4. What is the reason behind the selection of processing parameters? Given the various processing parameters (e.g. template size), how do you analyze the results with different effective spatial resolution?

7. Line 110: when to use each of 1/2-pass and 3/4-pass?

8. Line 111: the TanDEM-X DEM is almost two decades later than the JERS/ERS missions. What is the effect of using such DEM in your data product? Please quantify the possible error source for the readers' convenience.

9. Line 124: median and standard deviation of what? Velocity over stable terrain? How do you guarantee their result over other regions is applicable to yours without actual error analysis or validation?

10. Section 3: refer to the major comment #3. This section needs to be rewritten with graphs and tables. Since this data product is based on historical not operational satellite data, for each site, you can provide a table of the statistics of how many pairs used for JERS and ERS (how many InSAR pairs were used), temporal baselines, etc. Basically, you could analyze the tables in Appendix A for the readers' convenience.

11. Line 141: why saturation at 300 m/a?

12. Line 142: what is phase coherence? Is it just the InSAR coherence?

13. Line 144: missing ")" ?

14. Line 145: the grid spacing of 100 m is not matching the smallest template size in Table 1
15. Line 147: how to define "best"? Please mark those ones in the tables of Appendix A
16. Fig.1: what does the "red" color mean?
17. Line 161 and Line 166: why do you give priority to one data over the other? Please clarify
18. Line 162: any suggestion how to automate the manual adjustment of the outline?
19. Fig. 2,3,4,5,6: those figures should move to the relevant places
20. Line 171: as in detailed comment #1, only 1 or 2 pairs of Sentinel-1 data were considered to be 2020-2021 maps, which is not representative of the whole year. Please consider averaging multiple products throughout the year 2020. It is problematic to compare a multi-year mean (sampled at a few epochs) in 1990's with one or two epochs in 2020. Better to compare the results sampled at the same time of the year, otherwise, it is not clear if seasonal variation plays a role or not.
21. Line 178: why using 50 m/a as a threshold?
22. Line 180: please do not use "unpublished" as citations
23. Line 183: the previous ALOS product has different parameter setting. See the major comment #4.
24. Line 186: "masked out" means "removed". Please reword it.
25. Line 202: better to quantify the retreat
26. Line 205: could the seasonal variation play some role? Comparing the products at the same time of year would be more informative. Also refer to the above detailed comment #20
27. Line 214: better to quantify the retreat
28. Line 217: please provide a citation for this statement
29. Line 222-223: as mentioned a few times above, you only used one or two pairs of the Sentinel-1 data and the 1990's are grouped by averaging a few years' data. How do you guarantee it is not seasonal variation in 1990's and/or in 2020/2021? What do you exactly mean by "not detect clear sign of destabilisation"?
30. Line 228-231: not clear what you exactly meant? Please clarify what type of errors are you referring to and what methods (offset-tracking or InSAR) are you talking about?
31. Line 233-234: any reason for higher maximum speeds in 1990's? Also, the difference maps in Fig. 5 have the higher maximum speeds in 1990's masked out, due to retreat of the glacier's frontal line. In our opinion, this difference map is not a good graphical representation of the difference in ice flow flux. You might need to consider another graph for better representing the difference. Same problem to the difference maps in other figures.
32. Line 245-248: same problem as detailed comment #29
33. Line 268: would be great to show a map of advance/retreat in meters
34. Line 278-281: this statement by the authors relate to our detailed comment #1, #20 and #29
35. Fig. 7 and 8: Is Fig. 7c same as Fig. 4c? Similarly, is Fig. 8c same as Fig. 6c? As mentioned above a few times, difference maps are problematic both in space and in time, i.e. masking out speed changes due to retreat, and sampling too few Sentinel-1 pairs as the 2020-2021 maps.

36. Line 329-331: this should be elaborated as the primary content in an ESSD section on error sources. You seem to omit it by only citing a few citations. Questions: 1) are you using the same parameter setting as the citations? 2) even though you might have used the same setting as the citations, the settings across various sensor data are very different as Table 1 shows. How would this error due to different parameter setting propagate to the final differential map? Also refer to the major comment #4.
37. Line 331: QA-IV-1 seems too subjective to be included in a scientific dataset
38. Line 334: QA-IV-1/2/3 are not found in the tables of Appendix A as the authors stated in the text. Please fix
39. Line 339-340: here you seem to have much more Sentinel-1 data for the short-term variability analysis. Why not including them for the Sentinel-1 2020-2021 mosaic that is used to compare with the 1990's and the ALOS PALSAR one? Instead, you only used one or two time epochs, which is too few.
40. Line 418-419: should be Fig. 13 not 14 (typo). If more (not two) surging glaciers are included, the winter data variability could further increase with respect to the annual mean. Given the 20-year time difference, it is risky to assume the same behavior of winter data variability especially when including more surging glaciers with several surging events. Also refer to the major comment #2
41. Line 463: "submitted" is an invalid citation
42. Line 469: please refer to the major comment #2
43. Line 473-475: please rewrite the sentence to make it clearer