**Manuscript Title:** Benchmarking of SWE products based on outcomes of the SnowPEx+ Intercomparison Project

**General review:**

Mudryk et al. aim to evaluate 23 different SWE products based on how well they represent SWE climatology, variability, and trends across mountainous and non-mountainous regions in North America and Eurasia. Using existing and newly created reference datasets, the gridded products are scored using skill target diagrams, resulting in a series of Taylor and target plots, eventually leading to an average ranking of the SWE products. The methodology and technical approach to this evaluation is clear. However, the presentation of the datasets, methodology, and results is notably convoluted or disorganized at times, dampening the impact that this thorough analysis could have. In particular, a clear workflow figure could aid in the introduction of the overall evaluation strategy, where related text often references several other sections, causing for much back-and-forth within the manuscript. While repetition in stating the methodology is appreciated, sometimes the methods, results and associated discussion appear in a single section, making the information challenging to process. Thus, I include no major analysis comments and suggest the authors primarily focus on restructuring the manuscript for a clearer portrayal of meaningful results. My more pointed comments below should help with these review items.

**Detailed/line-by-line comments:**

Suggest writing out each abbreviation first (e.g., ILAMB, AMBER, IMS)

Table 1: Suggest placing a reference column and stating, prior to the table, that each "family" of SWE product will be discussed in more detail following the table and/or placing the text prior to the table.

Section 2.2

This section is generally challenging to follow, yet it is intended to set up the pertinent evaluation strategy. There is reference to various sections ahead of the current (e.g., full details referenced in section 2.4, reference to 7 products chosen for an ensemble). Perhaps a schematic of the workflow/evaluation scheme would be helpful. As such, it is also unclear when a point system was introduced (line 153-154).

Line 156: Unclear what exactly expert judgement is considered in this case

Table 2: Suggest including some justification as to how regions were selected. Unclear at what spatial scale these variables are evaluated.

Line 172: It would benefit the readership if the text stated the spatial extent in addition to the figure.

Figure 2 c-d: It is unclear what the scale bar is referencing, as it pertains to climatology. If this is "SWE climatology," or peak SWE magnitude, labeling it and stating so in the figure caption, as opposed to "climatology," would be helpful.

Line 219: It is unclear what "broad a range of meteorological analysis fields as possible" looks like for the 7 selected products, which are spelled out later in this paragraph. Could an example be provided compared to a product that was not selected?

Line 221: Consider removing the following sentence, as it again jumps ahead to several sections from the current and causes confusion: "It will be demonstrated in Sect. 4 that the reanalysis-type products which employ assimilation of surface snow information all have seasonal incongruities with one another." Suggest revising line 124 for similar reasons.

Section 2.4: Please number and format equations similarly, as there are many and some build off of one another. Reference to a figure here or in one of the citations may bring additional intuition to this methodology here, prior to seeing the results.

Line 249: Suggest a ½ to full sentence on why this approach was taken to rank similarity across products (were there other approaches in consideration?).

Line 353: Additional annotations on this figure would be helpful. For example, placing notation near Crocus-ERA5 and ERA5-Land on the upper right panel would aid in the necessary scanning between text and figure (especially since the numbers/rankings obviously change between panels). This comment extends to Figure 4 and 5. Suggest also reiterating what is represented on each axis in each figure and/or across panels, particularly in the Taylor plots and in reference to pattern statistics.

Section 3.3: Many of these beginning sentences/paragraphs, aside from the sentences explicitly referring to Figure 7, read as though they belong in the methods or discussion sections, which decreases the impact of the following results. There is a lot of information to unpack in Figure 7. The paragraph structure for each row is appreciated. Perhaps the authors would consider beginning each paragraph with the intended take home point, particularly for the bottom row of results.

Section 4: The final result, presented quite clearly, does seem as though it could live in the results section. Are there comparable results specific to the select regions (NA and Eu mountain and non-mountain)? This question also pertains to my next comment.

Line 535: Can the authors expand and be more explicit about "The relative overall rankings shown in Fig. 8 are meant to function as a guideline only"? The following conclusion also states that "user needs and objectives must guide product selection," however a lot of technical and thorough work went into the culminating Figure 8. Are there thus product recommendations and takeaways for users broadly and by region?

Discussion: There lacks a discussion on limitations to this assessment and consideration of other gridded SWE products.