

Dear Simon Gascoin,

Thank you for your constructive feedback. We have revised the manuscript as outlined below.

I appreciate the responses by the authors, which helped me better understand the scope of this study. I hope it will help other readers too. I noted a few additional questions/suggestions below.

1) The semivariograms could be included in the manuscript (or in Supplement). This analysis can be used to strengthen this sentence in conclusion ("The strong agreement in non-mountain areas occurs because the reference SWE measured by gamma or snow courses agrees up to the scale of most gridded products evaluated")

We have added the semivariograms to the Supplement and referenced these figures in the main text.

2) I cannot agree with the implicit logic of this new sentence: "Importantly, despite these limitations, our analysis is able to capture inter-product differences in mountain regions". The fact that there are inter-product differences in the evaluation of the products is not a proof that the reference data were correct.

Sentence removed

3) In Supplement, I still struggle to understand the method of aggregation.

3.1) What is a "base site"?

3.2) This sentence confused me "Sites included in a search pool were dropped from the list and the window moved to the next site on the list." Does it mean that the final product is different if the first site in the list is different?

3.3) It would be useful to explicit what is meant by "within the same native product grid". Is it defined according to a radius or to the actual shape of the grid cell (I ask because both options were mentioned in the response letter "a given search radius (100km, equivalent to a 200km aggregation window)"). Some aggregation algorithms even consider points that are not within the target cell (e.g. inverse distance weighting, or kriging). Sorry to be picky here but it is a key algorithm of the final dataset (otherwise sharing the code would help).

#### Revised supplement text:

The following describes how we performed the data aggregation generalized in Section 3.1. As outlined in Section 3.1, we averaged all reference-product pairs within the same product grid at its native resolution and then averaged the reference-product pairs within a larger area. Because the product grids do not overlay perfectly we did the following:

Starting with a single reference site all reference sites within 100 km of this site were identified. If, within a given pool of matched reference sites, there were multiple reference-product data pairs within the same native product grid, these pairs were averaged. The mean product and reference SWE within each pool of data was then calculated. This

process was repeated sequentially, starting with site ALE-05AA805 and ALE-05FA802 for snow course mountain and non-mountain, respectively, and AK101 and AB101 for gamma mountain and non-mountain, respectively. Sites included in a search pool were dropped from the list and the window moved to the next site on the list. Snow course and gamma SWE were considered separately, and mountain measurements were separated from non-mountain. **Due to the nature of the aggregation method, the result will differ slightly if starting with site other than that listed above.**

4) In the conclusion, it may be useful to clarify or rephrase this statement "targeted approaches are required to validate SWE products in mountain regions" since mountain SWE products were actually evaluated and ranked in the companion paper.

**Text removed.**