

## Referee 2 Craig Smith

We thank Craig Smith for his helpful comments. Please find our detailed answers (in blue) below.

*The paper provides some insight into the potential variability in local-area climate indices that users can expect due to measurement station location and local-scale variability in snow cover properties, and the potential pitfalls of extrapolating point-measurement-derived indices to the regional or landscape scale. The paper is interesting and relatively well written. I do have a couple of concerns that should be addressed before this paper can be published in TC. My major concerns are as follows:*

- 1. I have concerns about the use of the term “uncertainty”, largely in the title, abstract, and conclusions. I don’t consider myself a metrology expert, but to me, “uncertainty” is a metric attached to a measurement to inform the user of the range of values to be expected when the measurement is made with respect to what the true value actually is. Therefore, each manual measurement presented in this paper would have an attached uncertainty, and that uncertainty would contribute to the overall uncertainty in the calculation of seasonal climate indices. However, the more appropriate terminology for what is actually being assessed here is “variability”, or specifically, the impact of spatial variability on the indices. This suggested revision doesn’t impact the interpretation of the results (in fact, the term “uncertainty” is really only used in the title, abstract, and conclusion, and not in the results) so updating these sections with more appropriate terminology should be a relatively easy revision.*

*[Answer]: We agree and updated title, abstract, and conclusion accordingly; using variability or local bias instead of uncertainty.*

- 2. The paper presents some insight into the impact of local-scale variability of snow cover measurements on seasonal climate-related indices and offers some explanation as to why snow cover measurements can be quite variable in space. I believe that it is implied, but both the authors and the readers need to understand that it is highly unlikely that two measurements can adequately assess local-scale variability. I suggest that this point be clearly made (with references where appropriate) so as not to accidentally mislead the reader.*

*[Answer]: We clarified that point in the conclusion by adding the following sentence: Our term variability or “local bias” is only valid for the parallel analysis (two point measurements). But even so, the results give an indication of possible variations for various indices.*

Specific comments from annotated pdf:

*Title: I'm not a very good metrologist, but in my opinion, this is not the most appropriate use of the term "uncertainty" given the context of this paper. Uncertainty is the range of values expected when you make a measurement as compared to what the "true" value actually is. You would certainly have an uncertainty associated with each manual snow depth measurement at each station, and that would contribute to the uncertainty in estimating the land scale mean snow depth (for example). However, what you are assessing is the local-scale variability, and it's impact on deriving seasonal*

*indices for the region. Having said that, the local scale variability is likely not assessable with only a pair of measurements, but that does not diminish the value of this assessment provided that both the author and the reader understand that (which means that it should be pointed out in the discussion or conclusions).*

*As an example, you state in the abstract that "there is hardly any difference between DJF and NDJFMA which show median uncertainties of less than 5% for all indicators." but it's not the uncertainty that is less than 5%, but merely the difference in the indicators. For the most part, the only place that you talk about "uncertainty" is in the abstract and the conclusion, so this should be a relatively easy fix.*

[Answer]: Thank you for pointing that out. We agree and changed the term to variability.

*23-24: I found this sentence somewhat difficult to interpret. Maybe it's just me. For better clarity, perhaps make this two sentences:*

*The highest percentage of stations....*

*The lowest percentage of stations...*

[Answer]: true and done

*27: Can you clarify what you mean by "prevention measures"?*

[Answer]: We added avalanche prevention measures. Could be the closure of a road, artificial triggering or evacuations.

*36-37: this sentence structure is not quite right. I think this can be joined with the previous sentence for better structure and flow.*

[Answer]: Rephrased to:

All measurements of snow cover are dependent on the local characteristics of the site: i.e. exposure to wind or solar radiation, as well as nearby buildings or trees may have an impact on the measured snow quantities.

*38: "explain" is a better word*

[Answer]: changed to recommend

*40: typo*

[Answer]: done

*40: availability of observers? Not sure what you mean here*

[Answer]: Rephrased

*49: comma*

[Answer]: done

*55: the investigation of*

[Answer]: done

*65: their*

[Answer]: done

71: You should use the entire name of the institute here, defining the acronym

[Answer]: done

74: As a geographer, I'm a little partial to maps. Could you add a map of your station pairs, perhaps colourizing the markers to indicate length of overlap?

[Answer]: We didn't want to overload the manuscript with figures and decided in favour of a table in the appendix (Table A1) instead of a map in the manuscript.

78: Could you use a couple of sentences to outline what the regular measurement procedures are? E.g., time of daily measurement, static stakes vs snow probes, snow boards, etc.

[Answer]: We added a sentence with references in L76-77.

Measurements are taken every morning at 6:00 UTC at least between November and April (for details refer to Haberkorn (2019) and Buchmann et al. (2021)).

87: perhaps "lack of awareness" would be a better choice of words

[Answer]: done

166: Is this speculation or do you have information to support this? You can soften this statement by saying "likely due to..."

[Answer]: softened accordingly

167: what exactly is "hardly any"? Rather, you should say "fewer than...".

[Answer]: done

189: Figure 2; Is there a suitable compromise such that the y-axis scales can be made the same in the left and right panels?

[Answer]: Unfortunately not. We tried several variations, but decided against it due to information loss.

191: This would be better stated as: "Snow depth (HS) and snowfall (HN) indicators are based on 30 and 24 station pairs, respectively."

[Answer]: done

198: You can drop the "respective" wording, if 20 out of 30 station pair differences are larger, then 10 are assumed to be smaller (unless some are equal, then you have to revise further).

[Answer]: done

212: "huge" is subjective. You are comparing variability in decline vs. accumulation periods, so simply saying "larger" would be more appropriate.

[Answer]: done

217-222: I'm not sure why this matters for your discussion. It would only matter if you are trying to draw conclusions about the climatological significance of the changes in Dstart and Dstop, and that is not the focus of this paper. You have noted how you define Dstart and Dstop for this paper and that's good enough. If anything, include the discussion about Klein et al when you discuss your methodology.

*[Answer]: We agree and decided to drop the entire paragraph. We moved the references for the various definitions to our method section but kept the comparison with Klein et al. to put our values into context.*

*Methods L114ff:*

There are various definitions for snow onset (Dstart) and disappearance dates (Dend) depending on the application in hand (see e.g. Foster (1989), Kirilyanov et al. (2003), Peng et al. (2013), Stoebe et al. (2002), and Klein et al. (2016)). However, as none of them suits our purpose and for sake of simplicity, we defined them as the [...]

*4.3 L234ff:*

Our values of temporal changes in Dstart and Dend correspond to values obtained by Klein et al. (2016). Although the time periods are not exactly the same, the absolute changes in Dstart and Dend are similar for the few stations analysed by both studies (see Table A2). This suggests that the absolute changes of Dstart and Dend are in general quite robust.

*229: ... in Dstop. I think since this section is only about Dstop, you should include it in the header.*

*[Answer]: done. As pointed out by referee 1, we changed Dstop to Dend.*

*238: above, you use site abbreviations in the discussion. I appreciate when the site names are used. I think this is appropriate given the limited number of times that you refer to a specific site. However, since the sites are abbreviated in the plots, you really should use both: site(abbreviation).*

*[Answer]: done*

*241: ...just 3 m away from the large turbine house at the power station.*

*[Answer]: done*

*255: I don't think this is defined anywhere*

*[Answer]: We changed it to MeteoSwiss and SLF.*

*258: IMO, this would read ok by just changing this to "variability"*

*[Answer]: done*

*263: this is a bold subjective statement, and I don't think that you have proven this at all. In fact, you showed that the return values for your station pairs are all within your prescribed confidence intervals.*

*[Answer]: We disagree, as Fig. 7 demonstrated that for HSmax one station pair and for HN3max three stations are outside each other's 95% confidence intervals. However, this sentence has been moved and rephrased as the conclusion was shortened on the request of one reviewer.*

*274: You like the word "huge" but IMO, it is a word better suited to telling fishing stories in the pub than in a scientific paper (maybe it's just me) but I would rather you use a less colloquial term.*

*[Answer]: done*

*291: better to use a few words here to indicate why the stations are "problematic".*

*e.g. The problem is that stations that exhibit "local-bias" are not easily detectable in the record without the availability of parallel measurements.*

*[Answer]: Rephrased*

298: Shown in green are the...and shown in red are the...

[Answer]: done

Figure A2: You use red and green above to represent  $D_{start}$  and  $D_{stop}$  and then switch to red and green here to represent each station of the pair. It's not a big issue, but can I suggest using a different colour pair here to avoid confusion?

[Answer]: We agree and changed the colours of Fig 5 (formerly Fig A2).

Table A1: These should be specified in the caption

[Answer]: We changed it to station pairs