We thank both referees and the associated editor for very constructive and helpful comments. There were several points raised by both referees that addressed similar or equivalent points. We listed the common points of criticism first before individual comments of each referee are considered separately. Minor changes such as typos have been incorporated in the MS without listing them here. In order to improve readability, comments by the respective referee are listed in italic, while responses and modifications in the MS are written in regular typesetting. Sentences and paragraphs being incorporated in the manuscript are listed in bold letters here and in the manuscript. To keep the manuscript up to date, we checked for recent publications and included some wherever appropriate. Within the introduction, we included Mottram et al. (2019) as another source for changes in mass loss processes and added Lewis et al. (2019) as another example of extensive ground-based radar campaigns. In addition, we exchanged the previously referenced Lewis et al. (2019) discussion paper in TCD to the now published Lewis et al. (2019) TC paper.

Common points of criticism:

- Both referees suggest to change the title of the manuscript. We decided to use the suggestion by Lynn Montgomery and changed the title to: **Relating regional and point measurements of accumulation in southwest Greenland.**

- Another point both referees criticize is the inconsistent/ interchangeable usage of SWE and snow accumulation within the manuscript. Surface mass balance (SMB) is solely used (and properly introduced, L27) within the introduction. Here, SMB is defined as …sum of snow accumulation and lateral redistribution by sublimation, wind and runoff…. This specifies the usage of the term "accumulation" and the importance of determining its spatial representativeness. In the revised manuscript, we consistently have changed the terminology to snow accumulation with symbol b_s and units [kg/m^2].

- In addition, it has been suggested to simplify especially the section 2.3 dealing with spatial extrapolation. We now introduce terms such as variogram, nugget and anisotropy to facilitate readability of Section 2.3. Some radar terms are additionally explained as well.

- We modified the respective paragraphs in the introduction, which deal with objectives and scientific questions this work tries to answer. We fully agree that the main purpose of this manuscript is the relation of point measurements to regional accumulation. As stated by referee #1, the raised question (i) is a prerequisite to assess spatial representativeness and, hence, is removed from this listing. Since commonly applied in situ measurements of snow accumulations represent only a snapshot in time, it remains open whether accumulation patterns change with summer melt processes and are similar for two different winter accumulation season. We agree that the assessment of seasonal persistency cannot be properly determined with the available field data. However, since temporally continuous determinations of changes in accumulation are available and feasible in Greenland nowadays (upGPR, neutron probes), a relation of two consecutive years of data with point measurements is valuable and consequently is addressed in the results and discussion section. In addition, liquid water percolation has an effect on accumulation resulting in seasonal mass fluxes from the surface into deeper firn especially for

the investigated sites within the deep percolation zone of the Greenland Ice Sheet. We changed the respective paragraph to the following statement:
The aim of this work is to relate point scales to regional scales of one to several square kilometers in area to improve our understanding of the representativeness of point measurements. For this purpose, we examine snow-pit and GPR data from two sites within the percolation zone of the GrIS and one site at the equilibrium line gathered over several field seasons. For each site, we investigate density variability between measurements from up to six snow pits within an area of 4 km^2 made in a single season, process radar transects of up to 25 km recorded in close proximity to those snow pits, and spatially extrapolate the radar-derived accumulation to estimate area-wide accumulation variability. For temporal comparisons, we use continuous observations of accumulation and melt recorded by upGPR \citep{Heilig2018}. Our results show that spatial representativeness of snow accumulation for a point measurement (snow pit) is high but values can be affected by local wind-induced surface roughness. We recommend to apply multiple snow depth measurements at the vicinity of the pits to better assess accumulation on regional scales.

Reply to referee #2:

We highly appreciate comments raised by the referee and present a point-to-point reply for all issues raised by the referee. For an improved readability and to facilitate direct response, we sometimes subdivided comments into several paragraphs referring to similar issues Please also note our general response at the top of this document.

This paper tries to answer the question of how representative point measurements of snow accumulation are for the larger regional scale. This subject is important, urgently needs attention, and this paper fills a void in our knowledge on the connection between the observation scale and the (regional) climate modelling scale. Scientifically, the paper is solid, and I have few methodological remarks. In terms of presentation however, I have quite a few remarks. Changing a word or sentence here or there won't fix the fact that the paper is quite tough to read.
We thank the referee for the evaluation and the overall positive assessment.

General

- Various terms are used interchangably, without a proper definition. Snow accumulation, SMB, SWE, snowfall, snow depth. Please have a critical look at the terminology, simplify, and make uniform.
Please see the common comments above. We agree that snow accumulation and SWE were used inconsistently. Snowfall and snow depth are standing terms all described in Fierz et al. (2009). We do not consider it being necessary to introduce these terms. SMB is only used within the introduction where it is properly introduced.

- I had to dig quite deep in my memory to connect the dots between variograms,nuggets, space-invariance, isotropy and stationarity. Would it be feasible to ease the text in section 2.3?.
A criticism raised by referee #1 as well. We now introduce each geostatistical term within this section. Range in the sense of correlation range is consistently used as correlation range from now on.

- The abstract is particularly awkward in grammar and style, as if it was the last part that was written and not checked before submission. I'll give three example sentences and how to make this readable:.

We sincerely apologize for the sloppiness of the abstract and carefully revised the entire abstract. We included all recommendations and now hope it is significantly simplified.

In recent decades, the Greenland ice sheet (GrIS) has frequently experienced record melt events, which significantly affected surface mass balance (SMB) and estimates thereof. SMB data are derived from remote sensing, regional climate models (RCMs), firn cores and automatic weather stations (AWSs). While remote sensing and RCMs cover regional scales with extents ranging from 1--10~km, AWS data and firn cores are point observations. To link regional scales with point measurements, we investigate the spatial variability of snow accumulation (b_s) within areas of approximately 1—4 km^2 and its temporal changes within two years of measurements. At three different sites of the southwestern GrIS (Swiss Camp, KAN-U, Dye-2), we performed extensive ground-penetrating radar (GPR) transects and recorded multiple snow pits. If the density is known and the snowpack dry, radar-measured two-way travel time can be converted to snow depth and b_s. We spatially filtered GPR transect data to remove small scale noise related to surface characteristics. The combined uncertainty of b_s from density variations and spatial filtering of radar transects is at 7--8\% per regional scale of 1—4 km^2. Snow accumulation from a randomly selected snow pit is very likely representative of the regional scale (with probability p=0.8 for a value within 10\% of the regional mean for KAN-U, and p>0.95 for Swiss Camp and Dye-2). However, to achieve such high representativeness of snow pits, it is required to determine the average snow depth within the vicinity of the pits. At Dye-2, the spatial pattern of snow accumulation was very similar for two consecutive years. Using target reflectors placed at respective end-of-summer-melt horizons, we additionally investigated the occurrences of lateral redistribution within one melt season. We found no evidence of lateral flow of meltwater in the current climate at Dye-2. Such studies of spatial representativeness and temporal changes in accumulation are necessary to assess uncertainties of the linkages of point measurements and regional scale data, which are used for validation and calibration of remote sensing data and RCM outputs.

- In several parts, you claim that snow accumulation should be established for an area of at least 20 x 20 m. This seems a very important implication for future field work. However, I miss the quantitative underpinning of these numbers. Why not 10 x 10, or25 x 25 m? And how should this be done if no GPR is available? This is such a crucial part of the manuscript that I expect some more discussion of the implications on field practice.

We included an analysis on benefits from multiple snow probings on the assessment of the mean snow depth per area. This changed a large fraction of the respective section:

The above results imply that a point measurement of b_s (snow pit, upGPR value, neutron probe, etc.) is representative for an area of roughly 4x4 km^2 at Dye-2 with a probability of p ≥ 0.9 and an uncertainty of ±10% in case snow depth is averaged. For KAN-U, the spatial variability is slightly higher and, consequently, there is less certainty about how well a single measurement represents the surrounding area. However, we consider a probability of p ≥ 0:8 with uncertainty of ±10% for both study sites as a resilient estimate.

To quantitatively assess the benefit of snow depth measurements in addition to a snow pit, we numerically assume a sinusoidal snow depth variation with wavelengths of 56 m (arithmetic mean of the previously presented range in wavelength for the GPR transects) and average amplitude of ±6.8 cm (the fluctuations in snow depth from arithmetic mean). Averaging multiple snow depths (with a sampling distance of 1 m) from a 20 m long probing transect, result in a maximum possibly measured offset in snow depth of -20\% (amplitude decreases to 5.4 cm). A 10 m long probing line reduces the maximum offset by -6\% compared to single point measurements (6.4 cm amplitude). A 30 m long snow probing line, however, result in a decrease of maximum possible offsets by -44\% (3.8 cm

amplitude). An additional cross line of probings will further decrease offsets. Only if the surface features are aligned symmetrically in both probing directions, the maximum offset derived from both lines will theoretically remain stable. For a measured snow pit with ρ_s=350 kg/m^2 and L_s=1 m, the combined regional uncertainty (±5\% density uncertainty, ±6.8 cm snow depth variation) reduces from a single point measurement with b_s = 350±42 kg/m^2 to a maximum possible uncertainty of b_s = 350±35 kg/m^2 for just a single 20 m probing line. These numerical results confirm values for representativeness derived from geostatistical extrapolation. Hence, we recommend to combine a larger number of snow-depth probings within an area of at least 20 m by 20 m in the vicinity of the pits to increase the regional representativeness. Regional snow density variations of ±5\% can be accepted if snow depth uncertainty is minimized. Snow probing lines can easily be performed with respectively low time consumption compared to multiple snow pits. In particular, the wind-induced surface roughness has to be accounted for to provide spatially-representative b_s values.

- The title is inappropriate. My suggestion would be: "Representation of point measurements for regional-scale snow accumulation in/on the southwestern Greenland Ice Sheet."
See above, the title has been modified in accordance to Lynn Montgomery's suggestion and we believe, it addresses your concerns as well.

- Throughout the paper, you seem to use rho_s mostly as a bulk parameter: a mean over a certain depth. Can you more clearly distinguish between the actual snow density and this vertically integrated bulk density, and define the bulk density clearly?
We now specify "bulk snow density" when it first appeared in the methodology section: **In dry snow and firn (with two contributing volume fractions θ_a+θ_i=1), the wave propagation depends solely on the relation of air (θ_a) to ice volume fraction (θ_i) (e.g., \citealp{Kovacs1995,Maetzler1996}). Hence, with the bulk snow density (ρ_s, the average density of the entire snow column) measured in snow pits, we can convert from TWT to snow depth (L_s) and the amounts of bulk accumulation b_s with unit kg/m^2) using the equation**

Specific remarks
- Title: "South-Western". I looked it up but this should be either "southwestern Green-land" or "Southwest Greenland".
Title has been changed - see above.

- L1: significant changes. In what?
See above the abstract has been changed significantly.

- L3: remove the sentence "Data sources ... coverage"
Sentence has been removed.

- L3: remove "at least"
Changed accordingly.

- L11: "per regional scale"?
Changed to: **The combined uncertainty of b_s from density variations and spatial filtering of radar transects is at 7--8\% per regional scale of 1—4 km^2.**

- L11: "to analyze for". To analyze is a transitive verb. Suggestion "we investigate". This recurs frequently in the text (e.g. P1L17)

We thank the referee for highlighting this. We haven't been aware that analyze is a transitive verb. We consistently substituted "to analyze" with "to investigate" or verbs with similar meaning, where appropriate.

- L18: "cannot be evidenced" -> "At Dye-2, we found no evidence of..."
Changed accordingly.

- L27: SMB related processes -> SMB
- L35: "certainly contribute" -> "may also contribute"
- L52: remove "at the GrIS"
- L57: point -> location
All changed accordingly.

- L70: suggest "To improve our understanding of the representativeness of ..."
Has been changed and rephrased to:
Point observations, such as snow pits and ice cores are usually performed once a year at most. Such temporal snapshots limit the evaluation of spatial representativeness as they can be influenced by recent weather conditions. Hence, it is necessary to clarify whether regional accumulation patterns are consistent over more than one accumulation season to investigate if temporally continuous point measurements such as AWS data, upGPR and neutron probes remain representative.

- L71: "in area for two sites" -> " in areas around two sites"
We have removed the respective sentence.

- L72: I have once been taught that one paper answers one major question. Your one major question is about representativeness of point measurements. All other questions are hurdles that you come across while answering that question. My suggestions would be to rephrase, and to formulate L70-80 such that you introduce the different steps needed to answer your "major question" with associated sections.
We rephrased the respective paragraph as listed above.

- L72: "equilibrium line of altitude" -> "equilibrium line"
- L125: numerous -> several
- L144: range -> ranges
- L146: constantly -> always
- L172: more constraint -> better constrained
- L181: remove "a realm of"
All changed accordingly.

- L197: Do not start a sentence with a mathematical symbol.
Modified to: The term $b_{s,N}$ is simply…

- L196 - 206: past and present tense are used here interchangably. Please unify.
We apologize for this. It is now unified.

- L217: snow depths -> snow depth
Changed accordingly.

- L217: why call this ice volume fraction? Suggestion to simplify these sentences:

"We investigate the error that we introduce by assuming a single bulk density in the conversion from TWT to snow depth for an entire GPR transect. For that, we use a collection of snow pits, several from each of five locations, that were collected in a period of three years (table 2)."
Modified to: **We investigate the error that we introduce by assuming a single bulk density in the conversion from TWT to snow depth for an entire GPR transect. Hence, we determine the spatial variability in density within the respective area. Table 2 presents snow-pit data from our three study sites and two additional sites.**

- L243: above average -> above-average (idem below-average)
- L327: larger scale -> larger-scale
- L341: north to south direction -> north-to-south direction
We have not yet inserted dashes for all such phrases. Such details are very specific and treated differently depending on the style and language of each journal. If applicable such phrases will be corrected within the final editing phase with the journal directly.

- L261: equals to -> equals
Changed accordingly.

- Figure 4: consider inverting the color scale. Blue = low accumulation, yellow is high accumulation
Here, we respectfully disagree. We consider it being more intuitive to have high accumulation associated with blue color and low accumulation with yellow.

- L293: awkward construction. Rephrase
Modified to: **The unfiltered data, however, show a decreased representativeness with p=0.89 in 2015/16 and p=0.77 in 2016/17 for the same uncertainty range of ±10\%. Here snow depth is solely derived from the snow pit. Such values demonstrate that b_s data derived simply from a snow pit without averaging snow depth around the pit location will decrease the area-wide representativeness at Dye-2.**

- L316: Not all of the collected radar transect patterns (grids?) ...
Sentence has been changed to: **Not all of the recorded radar transect grids are ideal for the applied geostatistical analyses.**

- L347: this sentence is not complete
The whole paragraph has been modified as suggested by referee#1. KAN-U is no longer used for this analysis.
To quantitatively assess agreement in accumulation patterns, we used the respective normalized accumulation data and calculated the quotient. The cumulative data distribution of the quotients is presented in Figure 8. A constant area-wide quotient of 1 would imply that the scaled accumulation patterns are exactly equal. For Dye-2, the probability of data being equally distributed in May 2016 and 2017 with a given uncertainty of ±10\% is p≥0.95, meaning all intersecting locations of the accumulation pattern in two consecutive years at Dye-2 are similar.

- L368: LaTeX error
Corrected.
- L387: 40% -> 40% lower
Corrected.

- L408: The conclusion about persistence is unsatisfactory, and you seem to be shifting goal posts in the manuscript. In the abstract you write that interannual accumulation patterns "are very persistent". In section 3.3, the 2016 and 2017 data are "very similar". Then in L408 you say that "results suggest persistence". I think you should refrain at all from inferring persistence based on two data points. It's ok to mention that the patterns were similar in both 2016 and 2017, but I don't think there is enough argument here to start discussing persistence.

We fully agree and weakened consistently throughout the manuscript pattern persistence to changes in accumulation pattern for 2016 and 2017 or agreement in accumulation patterns. In the conclusion, we state:

Our results suggest that there is only little change of accumulation patterns at Dye-2 for spring 2016 and 2017.