Responses to Reviewer 2 (Anonymous)

Thank you very much for the detailed comments and constructive suggestions. We agreed with all the comments and made significant changes in the manuscript accordingly. We would note that the line numbers correspond to the manuscript without comments/changes.

General Comments:
Note that we grouped general comments into five groups: (1) data specification, (2) GPS positioning, (3) GPR processing, (4) PhoDAR terminology, (5) topographic measures. We responded to each comment below.

(1) The authors compare many different methods for measuring and estimating snow depth in a tundra environment over several years….. More detail should be provided to describe which in situ snow depth measurements are being discussed. Please be specific about which snow depths, from which years, from which sites, etc.
[Answer] We agreed and added the data specification (including year, data type and sites) at seventeen locations indicated by "Comment R2-(1)" in the comments

Line 436: "data collected in May 2012"
Line 438, 452: "in May 2012"
Line 464: "in Plots A–D (May 2012)"
Line 474: collected in May 2012
Line 484: "May-2012 snow depth"
Line 495: "In the region of the 500-meter transects, …. snow-free DSMs collected in July 2013 and August 2014"
Line 509: "around the 500-meter transects in May 2015"
Line 520: "in each plot measured in May 2012"
Line 538: "in May 2012 (measured by the snow depth probe)"
Line 560: "in May 2012"
Line 573: "snow variability in May 2012"
Line 581: "over the full study domain in May 2012"
Line 609: "(Plot A-D and the 500-meter transects"
Line 611, 677, 690: "May 2012"

(2) The authors describe the controlling influence of the microtopography on snow depth distribution and highlight the variability that is present at the meter-scale. However, there is minimal discussion concerning the errors/uncertainty associated with each of the GPS positioning information… More detail needs to be provided on both the horizontal and vertical positional uncertainty of all datasets used in this analysis.
[A] We added several sentences on the positioning errors and calibration processes linked to various snow thickness measurements in section 2.2. For GPS, we included (Line 192), "the precision of the RTK DGPS is 2 cm in the horizontal direction and 3 cm in the vertical direction." For the snow depth probe measurements, we added (Line 204), "This was done using a GPS snow depth probe (Magnaprobe by Snow-Hydro) which had a reported vertical precision of < 0.01 m and horizontal precision of around 0.5 m. The corner coordinates for each grid were

surveyed with the RTK DGPS, while each snow depth point measurement was represented by the built-in GPS unit that was programmed to automatically record locations" for the fine-grid sampling, and (Line 212) "The coarse-grid snow data were collected using a tile probe, which had a precision of approximately 0.01 m. … The start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference the measurement locations" for the coarse-grid sampling. For GPR, we included (Line 223), "the start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference the measurement locations. The precision in horizontal positioning of the measurements is estimated to be about 0.1 m."

A discussion needs to be added to the manuscript on how GPS positional errors, or lack of GPS information (tile-probe, GPR measurements), were taken into account and how these errors can affect the agreement (RMSE) between measured/estimated snow depth products in an environment where meter-scale microtopography controls the snow distribution.
[A] We included the discussion on the GPS and other positioning errors (Line 664). We also mentioned that our accuracy assessment was all based on the snow depth probe data, which had some uncertainty itself (Berezovskaya and Kane, 2007). "For all the types of measurements, accurate positioning was critical in the polygonal tundra due to microtopography. The GPS snow depth probe (Snow-Hydro), for example, had the positioning error larger than 50 cm, and required extra post-processing to correct the locations. On the other hand, measuring the RTK DGPS at all the snow depth measurement locations would not be realistic since it would take time. We found that having a tape measure and measuring the start and end points by the DGPS were a reasonable approach, when the snow surface is smooth and hard. In this study, we used the snow depth probe data as the true snow depth to compare with other measurements (i.e., GPR, PhoDAR, and Bayesian estimation). To improve the accuracy further, the uncertainty in the snow depth probe would need to be measured and included (Berezovskaya and Kane, 2007)."

(3) The authors use an extensive survey of Ground Penetrating Radar (GPR) measurements to produce interpolated snow maps …..More details should be provided on the GPR system and associated post-processing including:
- Near field effects: The literature suggest that GPR does not collect good data near the surface down to a depth of about 1.5 times the centre wavelength - 500 MHz wavelength = 60cm x 1.5 = 90cm snow depth!
[A] We added a few references that provide GPR background information (Annan, 2005; Jol, 2009) in Line 241. The vertical and lateral spatial resolution in GPR data depends on the GPR frequency and soil properties. Vertical resolution is often approximated at ¼ (up to ½) of the wavelength (velocity/frequency = 0.25e9/500e6), in this case about 12.5 cm (up to 25 cm). This number is related to the thickness that a layer needs to be to be identified, not the positioning of this layer in the subsurface. Although the top ~5 cm of the ground can be partly obscured by the groundwave, snow-ground interface located deeper than 10-15 cm depth is easily imaged. The number given by the reviewer are not consistent with the literature cited below. We think that these numbers are actually related to the use of GPR at a given height above ground surface. Here the GPR is located on the ground and the coupling is occurring between the GPR and the ground.

Additional references:
Annan, A.P., 2005, Ground penetrating radar, in near surface geophysics, in D.K. Butler (eds), Society of Exploration Geophysicists, Tulsa, OK, USA, Investigations in Geophysics No. 13, pp. 357–438.
Jol., H.M. 2009, Ground Penetrating Radar Theory and Applications, Elsevier Science.

- What about echoes from snow layering

[A] Such echoes are not visible in the data because (1) the antenna frequency was relatively low (500 MHz), (2) the snow layer was relatively thin (if present), and (3) the low contrast between various snow layers in this environment. We included a sentence (Line 243) to avoid the confusion: “Differing from previous studies (e.g., Harper and Bradford, 2003), we did not observe echoes from snow layering. This is possibly because of the low antenna frequency (500 MHz), relatively thin snow layers (if present), and the low contrast between various snow layers.”

- ice crusts/depth hoar layers?

[A] We included one sentence to clarify this question (Line 246). “Hoar layers or ice layers were not visible in our data or sensed using the probe. Although ice may form at the ground surface, causing the uncertainty of a few centimeters, we did not consider this effect in this study.”

- What methods/software are used to pick the top/bottom of the snowpack

[A] We used ProMAX® software (included in Line 237). The system used in this study is a ground-based system that is pulled on the snow. We do not pick the top of the snow pack. Coupling occurs directly between the GPR antennas and the snow. This is different from other studies that use high-frequency GPR system often positioned a few decimeters or meters above the ground to evaluate snow-layer properties (e.g., Snow stratigraphy measurements with high-frequency FMCW radar: Comparison with snow micro-penetrometer, Marshall et al., 2007).

- Discuss post-processing for: changing topography?

[A] The GPR data processing (to determine the travel time) and travel-time picking is done before accounting for topography. The effect of topography was evaluated during the velocity estimation. To avoid the confusion, we included one sentence “we processed the GPR data including travel-time picking before accounting for topography (Line 237).” We evaluated the effect of topography, when we determined the radar velocity.

- need to warp the data in the vertical direction so that reflection horizons are not distorted?

[A] We only have one main reflection in the GPR data which correspond to the bottom of the snowpack. The distortion in the data is not significant.

- forward velocity variations?

[A] In our data collection, the GPR is moved much slower than the time needed for one trace measurement. This is not a significant source of uncertainty.

- need to dynamically stretch or compress in the horizontal direction after data collection to correct for any significant variation in forward velocity
[A] We used a wheel to trigger the GPR, which means that our measurements are relatively independent of the velocity (speed) at which the field crew walks.

- Typically corrected via attached wheel odometer, however wheel odometers have issues over uneven snow and ice features
[A] This depends on snow quality and wheel. We compared the distance from wheel with the distance on tape and found that the difference is generally very small at this site. The start and end coordinates of each transect were also surveyed with a RTK DGPS and used to georeference point measurement locations in respective transect. The associated final uncertainty is about 0.1 m in horizontal direction. To clarify, we added a sentence: "We compared the distance from wheel with the distance on tape and confirmed that the difference is generally very small at this site (Line 224)."

- Provide more detail on how the GPR data were linked to in situ snow depth probes for validation. Both datasets did not have GPS positional information. What methods were used?
[A] We measured in situ snow depth probes along the tape measures. The start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference point measurement locations in respective transect. The associated uncertainty is about 0.1 m in horizontal direction. We also evaluated the effect of positioning errors when we calculated the radar velocity (described in Section 3.1). To avoid the confusion, we added a sentence "Due to microtopography at this site, the positioning errors between in situ measurements and GPR data could lead to an error in snow depth estimation. We evaluate the effect of such positioning errors extensively, as described in Section 3.1 (Line 229)."

(4) The authors describe a unique method of using the Photogrammetric Detection and Ranging (PhoDAR) technique with an unmanned aerial system (UAS) to derive high resolution snow depths in a tundra environment. However, the authors needs to be more consistent in the terminology they use when describing these data…. The authors should use the term PhoDAR, rather than UAS when describing these data….. Please choose a terminology and make changes throughout the text.
[A] We have changed "UAS" to "UAS-based PhoDAR" or "PhoDAR" throughout the manuscript (Line 251, 262, 267, 299, 300, 308, 392, 308, 495, 501, 503, 509, 511, 645, 646, 647, 654, 749)

(5) One of the main results of this study confirms that changes in complex surface topography influence snow distribution in a tundra environment. They identify regions with complex topography (ice-wedge polygons), as areas with high topographic variability using two different methods: 1) calculated elevation difference (range in elevation) using a 0.5 m DEM, in a 1 m radius surrounding snow depth measurements (in situ and GPR). 2) A wavelet transform method is applied to the 0.5 m DEM. The output high-pass image from this transform produces what the authors call 'microtopographic elevation'. The wavelet transform method for identifying topographic variability in a tundra environment is a novel approach that to my knowledge has not been done before. However, very little detail is provided by the authors on

what exactly the microtopographic elevation units represent. Both positive and negative numbers are listed on the y-axis of Figure 5c. There are also no comparisons between these two measures of topographic variability for identifying complex terrain or discussion on the use of two distinct methods. - Can the authors provide a direct comparison of these two measures of topographic variability? Can they provide justification for using two different methods? Should only one method be used for consistency?
[A] The elevation difference within a 1-m radius was only used to represent the potential error of the GPR and GPS positioning. If the topography is highly variable in a short distance, such a positioning error is more likely to affect the calculation of GPR velocity. On the other hand, the microtopography and macrotopography – derived from the wavelet transform – were used to relate snow depth itself and topography. Although the topographic gradient is small over this site (the elevation difference in the domain is 3.1 m), assuming the flat snow surface would create a large error in the snow depth estimate. Separating the macrotopography and microtopography is the key when we estimate the snow depth over this site.

To avoid the confusion, we specified the first topographic variability by "the submeter-scale variability of topography (Line 633, 637, 737)" throughout the manuscript. In addition, the unit of microtopography is meter, and zero-microtopography means that the topography is equal to macrotopography that represents the large-scale topographic variation. To quantify the importance of microtopography, removing such large-scale topography in the LiDAR DEM has been done in recent studies (Wainwright et al., 2015; Gillin et al., 2015). To further clarify, we added this sentence, "Removing the large-scale topography has been done in the previous studies in order to capture or quantify the effect of microtopography on carbon fluxes (Wainwright et al., 2015) or soil properties (Gillin et al., 2015) (Line 326)."

Additional references:
Gillin, C., S. Bailey, K. McGuire, and J. Gannon (2015), Mapping of Hydropedologic Spatial Patterns in a Steep Headwater Catchment, Soil Science Society of America Journal, 79(2), 440, doi:10.2136/sssaj2014.05.0189.

**Specific Comments**
Line 76-77 – Please check the definition of microtopography. It is a term used to describe the study of surface features on a very small scale. Therefore, the soil freeze/thaw processes do not lead to microtopography. The authors need to qualify this term with an adjective such as complex. Please check all references to microtopography and macrotopography throughout the paper to ensure proper use of the term with appropriate adjectives.
[A] The definition of microtopography is consistent with previous studies such that microtopography was created by the ice-wedge development. We modify the sentence and include more references to clarify this point (Line 75): "The ice wedges develop when frost cracks occur in the ground, and ice grows laterally over years (Leffingwell, 1915; MacKay, 2000). Soil movement associates with ice-wedge development creates small-scale topographic variation – *microtopography* – where the ground surface elevation can vary significantly over lateral length distances of several meters (e.g., Brown, 1967; MacKay, 2000; Engstrom et al.,

2005; Zona et al., 2011).” In addition, we added Engstrom et al. (2005) in Line 88, which quantified the effect of both microtopography and macrotopography on soil moisture.

Additional reference:

Engstrom, R., Hope, A., Kwon, H., Stow, D., & Zamolodchikov, D. (2005). Spatial distribution of near surface soil moisture and its relationship to microtopography in the Alaskan Arctic coastal plain. Hydrology Research, 36(3), 219-234.

Line 83-86 – Fragmented sentence – Please put this sentence in context, i.e. related to investigating controls of snow on ecosystem properties.

[A] We fixed this sentence, and included the relationship to snow and ecosystem properties (Line 84): “In addition, there are large-scale topographic variability at the scale of several hundred meters to kilometers – *macrotopography* –which is often associated with drained thaw lake basins or drainage features (Hinkel et al., 2003). Although the effect of macrotopography on snow depths has not been studied, Engstrom et al. (2005) quantified that both macrotopography and microtopography have a significant effect on soil moisture distribution.”

Line 86-89 – Awkward/vague sentence – please rephrase

[A] We rephrased this sentence (Line 89): “The snow representation of the Arctic tundra needs to be refined to account for the effect of such multiscale terrain heterogeneities on hydrology and ecosystem functioning, by bridging the gap between finer geographical scales (sub-meter) and large areal coverage (several hundred meters to kilometers).”

Line 91-92 – The authors should differentiate between alpine and Arctic tundra, as both LiDAR and GPR have been used in alpine environments. Additional references should also be added, as much work has been published on characterizing Arctic tundra snow. Suggestion: ‘Snow depth characterization in Arctic tundra environments have traditionally been observed using snow depth probes (Benson and Sturm, 1993; Hirashima et al., 2004; Derksen et al., 2009; Rees et al., 2014; Dvornikov et al., 2015), or modeled using terrain and vegetation information (Sturm and Wagner, 2010; Liston et al., 1998; Pomeroy et al., 1997).’

[A] We added the references and replaced this sentence accordingly (Line 94).

Line98 – Spelling error

[A] Corrected from “show” to “snow” (Line 103)

Line 100 – Replace ‘at site or’ with ‘local and’

[A] Replaced (Line 105).

Line 98-103- These references are for LiDAR-based studies only. Can you provide references for the structure-from-motion algorithm and use of PhoDAR for snow depth estimation?

[A] We included the reference (Line 106): Nolan et al., (2015).

Line 104 – Reword – Suggestion: 'While there is potential for providing detailed information about local-scale snow variability using LiDAR and PhoDAR snow depth estimates, these techniques have not been extensively tested in ice-wedge-polygonal tundra environments.'
[A] Re-worded (Line 108).

Line 108 – Please state 'increased snow depth uncertainty'
[A] Corrected (Line 114).

Line 109 – If you are going to use the term point probe to refer to all ground-based snow depth probe measurements – be consistent throughout the text.
[A] We changed all to "snow depth probe" (Line 115, 116, 125, 138, 144, 199, 202, 204, 224, 227, 266, 278, 282, 286, 290, 292, 308, 309, 368, 393, 397, 402, 453, 463, 473, 481)

Line 108-117 - The authors discuss snow depth uncertainties associated with indirect measurement techniques. This is a good paragraph to acknowledge and provide details on the uncertainties associated with snow depth probe measurements (all types of probes), especially in a tundra environment due to over-probing into the organic subsurface. (Berezovskaya S, Kane DL. 2007).
[A] We included a sentence uncertainty in the snow depth probe measurements, and added the reference (Line 116) as "The uncertainty of the snow depth probe measurements is sub-centimeter to several centimeters depending on the surface vegetation (Berezovskaya and Kane, 2007)."

Line 121 – insert comma after multiscale, - This correction should be repeated at Line 134 – check entire paper.
[A] Inserted (Line 129), and the entire manuscript was checked.

Line 127-128 – The term 'integration methods' is vague. Distributed blowing snow models could be considered an integration method. – Reword Suggestion: 'To our knowledge, Bayesian methods have not been developed. . .'
[A] Corrected to "Bayesian data-integration methods" (Line 135)

Line 130 – Change 'UAS' to 'PhoDAR'. UAS can also be used to record LiDAR snow depths.
[A] Changed it to "UAS-based PhoDAR' (Line 139)

Line 132 – are objectives 2 and 3 the same?
[A] (2) is to quantify the variability of snow depths themselves (e.g., boxplots and median of snow depths in Plot A-D, shown in Figure 5), and (3) is to quantify the correlations between snow depths and topography. To avoid such confusion, we changed (Line 140), "characterize the spatial heterogeneity of snow depths" to "quantify the spatial variability of snow depths"

Line 135 – Change 'the LiDAR domain' to 'a LiDAR DEM covering an ice-wedge polygonal tundra landscape'
[A] Changed (Line 143).

Line 146 – Can you provide a reason/explanation for defining a study site of 750m x 700m? Extent of DEM?
[A] We included the explanation (Line 156): "This study domain has been characterized intensively in the NGEE-Arctic project, and produced various ecosystem and subsurface datasets, including snow depth measurements (Wainwright et al., 2015; Dafflon et al., 2016)".

Line 148 – Insert NGEE-Arctic acronym
[A] Inserted (Line 156).

Line 156 – Vague sentence. Current references (Sturm and Dvornikov) imply that these works confirm that no such plants exist within this study site – Suggestion: There are currently no tall shrubs or woody plants established within the study site, therefore complex topography is most likely to control the snow depth distribution within the study domain (provide references that complex topography can control snow depth).
[A] Changed accordingly (Line 168).

Line 171-172 – Change 'providing' to ', and were used to produce a high-resolution. . .' Additional LiDAR DEM details are provided in Hubbard et al., 2013 – horizontal accuracy 0.3 m and vertical accuracy 0.15 m should be stated in this paper, as this relates to aligning/matching the various geospatial data used in this paper
[A] Fixed, and included the sentence (Line 188): "The original reported accuracy is 0.3 m in the horizontal direction and 0.15 m in the vertical direction."

Line 173 – Change 'resolved' to 'resolves'
[A] Changed (Line 187).

Line 175-176 – How many RTK DGPS measurements were used to evaluate the LiDAR DEM? Were they obtained across the whole study domain? What is the positional accuracy of the RTK DGPS (vertical and horizontal)? Was the LiDAR horizontal accuracy 0.3m stated in Hubbard et al., 2013 considered when computing the RMSE between the RTK DGPS and LiDAR DEM? The RMSE of 6.0cm is within the 0.15m vertical accuracy of the LiDAR DEM – which is good!
[A] We had 1286 points of RTK DGPS measurements around the 500-meter transects to compare with the LiDAR DEM (added in Line 190). The precision of the RTK DGPS is about 2 cm horizontal and about 3 cm vertical (added in Line 192). We did not consider the reported accuracy when we computed the RMSE. We agree that the agreement between LiDAR and RTK DGPS was excellent.

Line 180-181 – To avoid confusion with the transects measured within the study plots (A,B,C,D)– please refer to the three long transects that traverse the study domain as the '500 m transects'. Please make these changes throughout the paper. It is stated that snow depths were measured along transects, but no details on which probes were used, and which sampling methods were employed (intervals) and the total number of depths recorded along these lines are provided.

[A] Changed to “the 500-meter transects”, and included the probe and the number of measurements (Line 191, 215, 221, 250, 267, 301, 495, 509, 609, 612).

Line 185 – The term accuracy should be changed to precision. Also a reference should be provided. What about the horizontal positional accuracy of the GPS snow probe measurements (references)?
[A] Changed to “precision” (Line 203). Since the horizontal positional accuracy was larger than 0.5 m, the locations were later corrected, using the RTK DGPS data at the end of lines. We included the procedure as “The start and end coordinates of each transect within the grid were surveyed with the RTK DGPS and used to correct the GPS snow depth probe measurement locations in respective transect, since the comparison to the RTK DGPS confirmed that the errors were a constant horizontal shift within each plot.”

Line 186-187 – Which transects? Even the 500 m transects? Fine Grid? Coarse Grid? GPR transects? All of them?
[A] Added “transect within the grid” (Line 204)

How did you ‘correct point measurements’? A simple XY shift for each individual transects? Was it a single shift for all depths across the study domain? What are the horizontal uncertainties associated with the ‘correction’?
[A] The corner coordinates for each grid were surveyed with the RTK DGPS, while each snow depth point measurement was represented by the built-in GPS unit that was programmed to automatically record locations. All the snow depth point measurements were made along regularly spaced transects. Comparisons between coordinates surveyed with both the RTK DGPS and the built-in GPS confirmed constant biases in the horizontal directions, which allowed a constant bias adjustment for all GPS surveyed snow depth point measurements. We included this description in Line 204.

Line 189-190 – Please change the term accuracy to precision of 0.01m.
[A] Changed (Line 213).

Line 191-192 – The ‘fine-grid’ total snow depths were reported as single dataset =~7200. For consistency, report the ‘coarse-grid’ snow depths as a grand total as well, rather than a per/grid number. I also did some rough calculations for estimating the total coarse-grid measurements: Five ~160m survey transects at 5m intervals only equals ~ 160 snow depths per grid. The authors state 380 snow depths were sampled per grid. If at equal intervals the distance would be ~2.1 m between depths? Was the spacing of the tile probe snow depths measured by survey tape as hinted at by Line 200, or only roughly estimated? Please provide further details on the accuracy of the GPS positioning of the tile probe/GPR measurements to provide context on how they were linked to the LiDAR DEM.
[A] We found a mistake; the total number was 380 (19*5=95 in each plot). The spacing along each line was 8 m. The start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference the measurement locations. We estimated that the horizontal errors are around 0.1 m. We corrected this in Line 215.

Line 194-195 – See comment for Line 180-181 – Please refer to the three transects as the '500 m transects'.
[A] Corrected (Line 215).

Line 197 – How were they co-located? GPS? RTK DGPS? Were the unique ID numbers associated with particular GPR measurements noted and related to snow depth measurements? Please provide more details, and a measure of uncertainty associated with this co-locating.
[A] The start and end coordinates of each GPR transect were surveyed with a RTK DGPS and used to georeference the GPR measurement locations in respective transect. The RTK DGPS coordinates and the length between the two RTK GPS coordinates were used to calculate the absolute position of each measurement location. By comparing this approach with surveying each measurement location with RTK GPS, we found that the error is about 0.1m. We added this explanation in Line 223.

Line 199-200 – What does 'exact location of each measurement was within 1 m' mean? Were the GPR survey lines conducted along survey measuring tape laid on the ground? If so, how were the measuring tapes fixed to the ground? Did they bend/blow in the wind creating potential positional uncertainty? Were the start and end coordinates of the GPR transects marked with RTK DGPS? Does the GPR data have GPS coordinates linked to each measurement?
[A] We agreed that this sentence was confusing, and removed this sentence. We included the exact procedure (Line 222), "In each plot, we acquired the GPR data at 0.1-m intervals (marked by an odometer wheel) along 37 lines of 4-m spacing. The start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference the measurement locations. We compared the distance from wheel with the distance on a tape and confirmed that the difference is generally very small at this site. The error of horizontal positioning is estimated to be about 0.1 m." The measuring tapes worked well at this site due to the smooth snow surface.

Line 200-202 – It should be noted that the use of GPR in Hubbard et al., 2013 was not for snow depth extraction. Please change text to 'a detailed explanation of the use and processing of GPR data for the purposes of identifying the active layer depth at the end of the growing season at this study site, are provided in Hubbard et al., 2013.' Are the Hubbard et al., 2013 thaw season GPR measurements for active layer thickness even relevant to the GPR measurements obtained in the frozen winter-period for snow depth estimates? The GPR measurements are conducted over two very different media. Hubbard et al 2013 describe a different GPR system entirely with two different antennas (450 MHz and 900 MHz vs the 500 MHz antenna used in this study). Can the authors please provide specific details on the software used and the steps taken to maximize signal-to-noise ratio, and correct for change in topography, change in speed of data collection (comment on the utility of the odometer in a sastrugi environment). In-addition, can the authors comment on the presence of any snow layering (ice crusts/wind crusts) that may or may not have caused echoing and how this affected the ability to identify the air/snow and snow/ground interfaces? Can the authors also comment on how sensitive the GPR data was to the thick organic layer and mosses, lichens and vascular plants that can make it

difficult to ascertain where exactly the snow/ground interface actually begins? What/where exactly is the GPR return signal bouncing off?
[A] We added more details on GPR data processing in the manuscript, corresponding to the general comment (3) above (Line 234). We also specified the processing software ProMAX ®. The odometer wheel worked very well in the type of snow encountered at the site. We tested this approach in various environments and the type of snow strongly influences the precision. By comparing this approach with surveying each measurement location with RTK GPS, we found that the precision in the used approach is on the order of 0.1 m. We did not observe any snow layering, and the reflection from the ground surface was clear. In addition, we did not observe the dependency on different vegetation types on the ground.

Line 202-206 – The pre-processing routine is explained. Can the authors explain for steps 1 and 2 how this was done? What automated algorithms were used? Was there human-visual intervention/training involved in identifying these two interfaces?
[A] GPR data were process using the software ProMAX®. The processing step are explained in the manuscript (Line 234). The travel time have been picked manually and automatically snapped. Please note that the snow-ground interface represented a strong reflector and that picking was relatively straightforward.

Line 206-207 – Please justify how the use of GPR for identifying the tundra activelayer thickness at the end of September (thawed) in Hubbard et al (2013), provides context for the use of GPR for snow depth estimation in the same tundra location in May (frozen).
[A] The GPR background theory and data processing is very similar to what has been done in Hubbard et al (2013) even if the target is different. The reflection from the ground surface was clear, which allowed us to measure the travel time between the snow surface and ground surface. We included "The GPR reflection signal from the bottom of snowpack (i.e., the ground surface) was clear, which allowed us to measure the travel time between the top and bottom of snowpack (Line 234)."

Line 209 – It is not clear whether the UAS and PhoDAR were measured over all four plot sites (A-D), or the three long 500 transects? Or entirely new transects in the same study area.
[A] To clarify, we included "only along the 500-meter transects" (Line 250)

Line 211-214 – What was the resolution of the PhoDAR elevation/snow height products?
[A] The reconstructed PhoDAR surface elevation models at this site show a resolution of 4 cm by 4 cm. We included this explanation (Line 263)

Line 218-220 – How many control points? What distribution? How did you assess this using the RTK DGPS? What kind of horizontal and vertical accuracy did you achieve with the PhoDAR products? How different were these accuracies from year to year for the snow-free ground measurements (were the same snow-free ground areas surveyed in both 2013 and 2014? Are there any comparisons between estimated PhoDAR elevation of the same area between 2013 and 2014? What were the vertical accuracies any different for the estimation of the snow surface elevation compared to the estimation of the snow-free ground?

[A] We compared the accuracy using the RTK DGPS measurements (1296 points) acquired in 2011. We included (Line 300), "We first evaluate the accuracy of the PhoDAR-derived digital surface model (DSM) by comparing it to the GPS elevation measurements along the 500-meter transects acquired in 2011." The accuracy was evaluated in the procedure described in Section 3.2, and the results were shown in Table 2. We did not see a significant change between 2013 and 2014.

Line 222-223 – Were the snow probe locations recorded at the RTK DGPS control points? If not, what GPS positional information was associated with these 183 snow depths, and what horizontal positional uncertainty did they have?
[A] To clarify, we included (Line 268) "The locations were marked on a tape measure, the start and end coordinates of which were surveyed with a RTK DGPS and used to georeference the measurement locations."

Line 229-230 – snow density can be variable in both a vertical and horizontal direction (Proksch et al., 2015)
[A] We added the sentence (Line 274): "The snow density could be variable in both and horizontal directions (Proksch et al., 2015)."

Line 241-242 – How did you link the DEM elevation data to the tile probe snow depth measurements?
[A] We selected the DEM elevation (0.5 m by 0.5 m resolution) at the nearest locations to the title probe measurements. We added this sentence (Line 285): "To link the DEM elevation data to the snow depth probe and GPR data, we selected the DEM elevation (0.5 m by 0.5 m resolution) and GPR measurement at the nearest locations to the title probe measurements."

Line 251 – Change 'GPS elevation measurements' to 'RTK DGPS control point elevation measurements'
[A] We specified the GPR measurements as "RTK DGPS measurements around the 500-meter transects acquired in 2011" (Line 301).

Line 253-254 – change 'co-location' (implies horizontal positional accuracy), to vertical agreement.
[A] Changed (Line 304).

Line 255-256 – All snow depth measurements? Or just RTK adjusted magnaprobe measurements? Do you use the three same schemes for exploring co-location of the two snow depth estimates? Do you use the average UAS snow depth within 0.5m radius in the vicinity of the snow depth probe (as described with the elevation datasets at Line 417-418)?
[A] We used the snow depth probe data collected in May 2015 (Line 308). We included this sentence "we used the same scheme for determining the snow-free and snow surface elevation at the co-located points" (Line 306).

Line 262 – remove term 'recently'

[A] Removed (Line 315).

Line 266-267 – Do all four images produced in the wavelet transform provide topographic information? What information do the two images not used in this analysis produce?
[A] The others two images (high-pass horizontal and high-pass vertical) contains the topographic information only in a single direction (either horizontal or vertical). Therefore, we did not used in this study.

What are the different 'scales' for the wavelet transforms, and how are they determined?
[A] The scale is the width of the filter, and a parameter in the wavelet transform. To clarify, we added in this sentence (Line 320): "The scale is a parameter in the wavelet transform, representing the width of the filter and the scale of topographic variability (Kalbermatten et al., 2012)."

Line 270-271 – Which high-pass image is used: horizontal or vertical (refer to Line 266-267)
[A] We added "high-pass diagonal image" to clarify (Line 325).

Line 278-279 – Please be specific on what the dependency was relating to the DEM resolution? Decreases with increasing resolution?
[A] We included (Line 337) "the lower resolution led to lower correlation coefficients"

Line 297 – -change to – 'interpolating the sparse in situ snow depth measurements.'
[A] Changed (Line 357).

Line 334-335 – Change 'accuracy' to 'precision'.
[A] Changed (Line 401)

How would the inclusion of a larger error associated with each probe measurement affect the results? AS mentioned in my comment for Line 108, there are common over-estimation errors associated with snow probe measurements in a tundra environment that could be added on to any potential error associated with the precision of the depth probing instrument.
[A] We could include such an error in this model if the error is quantified, although we did not consider it in this study. We added this sentence (Line 405): "Although it is not considered this study, we could include a systematic bias of snow probe measurements as added shift"

Line 362 – Please define MCMC before using in a sentence (Line 360).
[A] Added (Line 428).

Line 369-372 – These two sentences contradict one another. First sentence states no systematic dependency. Second sentence states radar velocity depends on snow depth and topographic variability.
[A] The first sentence states the *radar velocity itself*, and the second sentence states the *variability of radar velocity*. To clarify, we included "itself" in the first sentence (Line 437). We

also added the sentence (Line 438), “The correlation coefficient between the radar velocity and snow depth is 0.11, and the one between the radar velocity and topographic variability is 0.15.”

Line 372-373 – Please provide more quantitative interpretation of Figure 2 variability. Based on the Figure 5 boxplots, the median snow depth for the site is somewhere between 50 cm and 45 cm = ~47cm. If you insert a vertical line in figure 2a to highlight the median snow depth for all sites, areas of shallow vs deep snow are better highlighted.
– Could also insert vertical lines at +/- 1 STDEV from the median to really highlight shallow vs deep snow. With these lines inserted It becomes more obvious that higher radar velocity is more frequent in deeper snow areas (provide average and STDEV), and low radar velocity is more common in shallow snow (provide average and STDEV). Overall variability is higher in shallow snow vs deep (compare STDEV), but this may be a function of the frequency of occurrence – i.e. smaller n in deep snow areas results in lower variability. The addition of a vertical line at the study domain average elevation difference could also help highlight the characteristics of radar velocity in complex vs smooth terrain.
[A] We added the lines of the median and +/- STDEV snow depth, and added the quantitative representation of the variability (Line 442): “The standard deviation (STDEV) of the radar velocity is 0.039 m/ns at the snow depth smaller than one STDEV minus the median snow depth, and 0.019 m/ns at the one larger than one STDEV plus the median.”

Please also provide more quantitative interpretation of Figure 2b, with average and STDEV velocities for complex vs smooth terrain. Why was 0.05 m elevation difference chosen as a cut-off to estimate the average radar velocity for smooth terrain?
[A] We include the cut-off topographic variability in Figure 2b. We added the explanation of the cut-off topographic variability (0.05 m) as (Line 446) “The STDV of the radar velocity is 0.015 m/ns at the topographic variability (i.e. elevation difference) smaller than 0.05 m, and 0.036 m/ns at the one larger than 0.05m.”

Line 378-379 – Can you provide the correlation values for both all depths, and only those depths in regions of low topographic variability?
[A] We added the correlation coefficients (Line 453) “The correlation between the measured and estimated snow depth is high (the correlation coefficient is 0.88)…”, and (Line 458) “The correlation coefficient between the GPR-based and probe-based snow depth was increased to 0.94.”

Line 377-382 – In section 2 the authors indicate that snow depths were located within 1m of GPR measurements. GPR measurements were recorded every 10cm. Were the GPR snow depths averaged in a radius surrounding each snow depth probe measurement? Was the closest snow depth/GPR measurement taken? if so – how was that determined using non-georeferenced tile-probe depths? Can the authors please provide more details on how the GPR vs probe snow depths were analyzed?
[A] We selected the closest GPR measurement. To clarify, we added the explanation (Line 285): “To link the DEM elevation data to the snow depth probe and GPR data, we selected the DEM elevation (0.5 m by 0.5 m resolution) and GPR measurement at the nearest locations to the title

probe measurements." For the tile probe measurements and the GPR measurements, the start and end coordinates of each transect were surveyed with a RTK DGPS and used to georeference the local grid.

Line 399-400 – They are not true mirror images of each other. Please re-phrase.
We modified this sentence to (Line 477) "Comparison of the fine-grid snow data with the DEM reveals the microtopographic effect such that the troughs and center of the polygon have larger snow depth."

Line 407 – Insert 'snow depth' in-front of 'estimates'
[A] Inserted (Line 486).

Line 408-409 – Reword. Suggestion: 'The high resolution GPR snow depth estimates are useful for determining if macrotopographic features can influence the distribution of snow depths across each study plot.'
[A] Reworded (Line 487).

Line 416 – Be specific about which transects. 'In the region of the 500 m long transects, the UAS. . .
[A] Specified (Line 495).

Line 416-417 – Change '. . .was first compared with GPS data in Table 2,' to '. . .was first compared with the RTK DGPS elevation control points measured in July 2013 and 2014. The RMSE values from this comparison are listed in Table 2.' – Please add the total number of control points measured for these comparisons to Table 2. The authors indicate that taking the average UAS derived elevation in the vicinity of each RTK DGPS elevation measurement provides the lowest RMSE, however all values listed in Table 2 are within 0.69 cm! Therefore, these results show that it doesn't matter what method is used to co-locate the RTK DGPS and the UAS derived DSM as these are less than the precision of all available datasets.
[A] We changed the sentence accordingly (Line 495), "the PhoDAR-derived snow-free DSMs (Figure 4a) collected in July 2013 and August 2014 were first compared with the RTK DGPS data (acquired in 2011) in Table 2". The number of points were specified earlier in the data description section. We agreed that the all the methods provide small RMSE. We add "Although all the scheme yielded an excellent accuracy (the RMSE less than 7.0 cm), taking the average provides the lowest RMSE."

Line 418 – The value for the RMSE=6.0 cm does not match any of the values presented in Table 2. Can you provide justification on why you separate the results in Table 2 by year of survey if you are trying to compare the three different schemes for assessing the vertical agreement between the RTK DGPS and PhoDAR elevations? If you have multiple years of PhoDAR elevation measurements of the same terrain, can you provide a comparison between the two snow-free elevation products from PhoDAR alone?
[A] The value was changed to (Line 500) "6.41 cm in 2013 and 6.19 cm in 2014" corresponding to the values in Table 2. Table 2 includes the two snow-free elevation products from PhoDAR

alone at the same terrain. We included the results from the two years to confirm the consistency of the snow-free elevation products at the same locations. To clarify, we added the sentence (Line 497): "We included the results of both years to confirm the consistency between the two snow-free DSM products at the same terrain."

Line 421-422 – It is unclear whether the same three schemes tested and presented in Table 2 are also used to co-locate the UAS derived snow depth estimates with the 183 in situ snow probe measurements. Should the same methods be used? i.e. average UAS snow depth within 1m of the snow depth probe?
[A] As noted earlier, in the methodology section, we added "We used the same scheme (the best scheme among the three) for determining the snow-free and snow surface elevation at the co-located points" (Line 306).

Line 422-424 – Did you use the same cut-off of elevation difference 0.05 m? How did you decide on the threshold between large/small topographic variability?
[A] Yes. To clarify, we added (Line 506) "the same cut-off values as the GPR snow depth analysis".

Line 426 – Can you include the location of the snow depths used to assess the UAS product in Figure 4b?
[A] Added in the Figure 4b. The black line in Figure 4b represents the snow depth probe measurements every 3 meter along the 500-meter transect.

Line 429 – Replace 'Figure 3' with 'Not shown' – The scale of Figures 3 and 4 do not illustrate the added detail found in the UAS data.
[A] 'Figure 3' is removed (Line 513).

Line 430 – Do you mean to say 'the minimal effect of macrotopography?'
[A] Changed (Line 513)

Line 446-447 – What do the units in 'microtopographic elevation' from the wavelet transform represent? How does it differ from elevation difference calculated for a 1 m radius of all probe/GPR measurements using the LiDAR DEM used to infer topographic variability in Figures 2b and 4c? Please justify why the authors use two different measures of topographic variability? Why is a scale of 32m chosen?
[A] The unit of microtopographic elevation is meter. The methodology to determine the microtopography and macrotopography is described in Section 3.3. The elevation difference within 1-m radius was primarily used for interpreting the GPR and PhoDAR measurements (particularly errors induced by the topographic variability at the short length scale), while microtopography and macrotopography were used to describe the topographic variability and its influence on snow depth. The scale of 32 m was selected to yield the best correlation between snow depth and microtopographic elevation. To clarify, we added (Line 530) "The microtopographic elevation is computed based on the wavelet transform with the scale of 32 m

as described in Section 3.3 (Figure 5b). The scale of 32 m was selected to yield the best correlation between snow depth and microtopographic elevation."

Line 450 – Change to - In contrast, Plot B has the largest variability in both microtopography and snow depth.
[A] Changed (Line 534).

Line 467-469 – Can you provide a possible explanation why there was no correlation of snow depth with curvature in this study vs previous studies? Is this a significant result?
[A] We added a possible explanation (Line 554), "This is possibly because the microtopography at our site was completely filled by snow, and the overall elevation gradient at our site (the elevation difference in the domain is 3.1 m) is much smaller than the one that Dvornikov et al. (2015) reported (the elevation difference was more than 60 m)."

Line 475-477 – Suggest removing the line ' Such a large correlation length is consistent with the field observation that the snow surface is smooth across the site.' This sentence is very subjective. Do you have any measure of smoothness or photographs perhaps? – Replace with something like: The estimation of a snow surface height (elevation + snow depth), effectively removes the influence of microtopography resulting in much larger correlation lengths (250 m).
[A] Replaced (Line 564).

Line 483 – add comma after 'Equation (1),'
[A] Added (Line 572).

Line 489 – Change to – 'The Bayesian estimated mean snow-depth field over the full study domain (Or NGEE-Arctic study site). . .'
[A] Changed (Line 581).

Line 492-493 – Change to – 'The snow depth does not have any large-scale trends over the full study domain, which is different from the LiDAR DEM in Figure 1b, but consistent with the interpolated GPR snow depths depicted in Figure 3 (right column), and the measured UAS snow depth measurements depicted in Figure 4b.
[A] Changed (Line 583).

Line 494-496 – I don't understand how exactly Figure 7b was produced? Please rephrase 'the mean field based on the kriging-based interpolation of the snow surface elevation.' What snow depth inputs were used to produce Figure 7b? And what method of kriging was applied (there are several), and why? It is unclear how the work of Diggle and Ribeiro, 2007 relates to Figure 7b.
[A] We agreed that the explanation was not adequate. We included more detailed and specific explanation (Line 590), "In this estimation, we used the same Bayesian algorithm one described in Section 3.4, except that we removed the topographic correlations and assumed a standard geostatistical model for snow surface (Diggle and Ribeiro, 2007). In other words, we had the

same algorithm except that we modified Equation (1) to $\boldsymbol{y} = -\boldsymbol{z} + \tau$, where $\boldsymbol{y} + \boldsymbol{z}$ represents the surface elevation."

Line 496-497 – The two figures looks very similar – however I have difficulty identifying which regions are considered 'central', and what type of measurements are considered 'many'? Can the authors also provide a more quantitative assessment on how similar?
Perhaps produce a third figure of the snow depth difference between the two rasters providing a quantitative comparison/statistics?
[A] We realized that we were using the duplicated images from the same estimation. Thank you very much for pointing this out. We fixed the image, and also changed the explanation in Line 594, "Although the two mean fields (Figure 7) are similar in the central regions that have many measurements, the regions without any measures have a significant deviation. This is because the snow surface estimation did not capture the change in macrotopography (e.g. the drainage feature in the southern part of the domain)." We would note that the quantitative comparison are given in the tests as RMSE (Line 717) and also in the cross validation (Figure 8).

Line 499 – change to – 'The estimated standard deviation of the Bayesian derived snow depths. . .'
[A] Changed (Line 600).

Line 509 – Please provide more details on how the independent snow depths used in the validation were 'randomly selected', and how many were kept independent for validation purposes vs those used in the production of the Bayesian and Interpolated snow depth maps.
[A] We added the explanation (Line 611), "The 100 points of the snow depth probe data were randomly selected from all the locations (Plot A-D and the 500-meter transects), using uniform distribution."

Line 514-515 – Suggest changing to – 'The RMSE for the Bayesian method of estimating snow depth including the topographic correlation is 6.0 cm, while the RMSE for the interpolated snow surface is 8.8 cm.' (I think a more representative/descriptive word should be used to describe this 'interpolated' snow depth data – i.e. what input data was used and what type of interpolation – then update all references to this data throughout the text)
[A] Changed (Line 617).

Line 520 – This RMSE value is for only the smooth terrain (limited to areas of low topographic variability). Please update stats for all tundra terrain.
This RMSE is for all the tundra terrain, since we selected the 100 validation points from all the locations (Plot A-D and the 500-meter transects). Please see the response to Line 509.

Line 525-529 – Firstly, thick snow is very limited in this type of tundra environment, especially in areas with low topographic variability (where GPR estimates are more reliable). The small sample size for these types of GPR measurements is the most likely reason for the observed smaller range in radar velocity variability.Secondly, - I agree. Thirdly, I feel the authors are contradicting the Second result concerning the influence of topography, by stating that there

are no trends in radar velocity related to topographic position. See my comments for Line 372-373. I believe there is a trend for radar velocity related to snow depth.
[A] We added the explanation according to this comment (Line 631): “The relatively low topopographic variability (compared to mountainous terrains) would have contributed to this fairly uniform radar velocity.” We would note that we had more than 50,000 points of the GPR measurements. To responding to the comments for Line 372-373, we added “The correlation coefficient between the snow depth and snow depth is 0.11”

Line 530 – This methodology was never fully explained.
[A] We included “described in Section 3.1” (Line 636)

Line 532 – see comment for Line 229.
[A] In our site, we found that the depth-averaged radar velocity was fairly uniform. To avoid the confusion, we added (Line 638) “We note that – even though the depth-averaged radar velocity and hence the depth-averaged snow density have little variability over the space –the snow density could be variable vertically along the depth.”

Line 553 – Change ‘option’ to ‘alternative’
[A] Changed (Line 645).

Line 569-577 – Why was a hundred-meter scale used? Please support the statements in this paragraph with quantitative anlaysis (average and STDEV values). Be consistent with the use of the average snow depth vs median snow depth throughout paper.
[A] We added (Line 689), “a hundred-meter scale (i.e., the size of Plot A-D)”

Line 585 – replace ‘the improved’ with ‘an improved’
[A] Changed (Line 706).

Line 596 – The authors never quantify whether the improvement is significant. Visually in Figure 8, one can see an improvement in lowering the STDEV when the topographic correlation is taken into account.
[A] We included the RMSE for both cases (Line 716), “Taking into account the topographic correlation explicitly improved the accuracy of estimation significantly (RMSE 6.0 cm), compared to interpolating the snow surface and subtracting the DEM (RMSE 8.8 cm).”

Line 603-605 – Reword – Suggestion: ‘It would be possible to quantify the seasonal changes in the topography-snow correlations by designing a full season ground-based measurement campaign and acquisition of remote sensing snow depth measurements (by PhoDAR or LiDAR), that monitored the same site over several years to account for inter-annual variability.’
[A] Changed (Line 724).

Line 613 – Change to ‘ice-wedge’ (check text throughout paper for consistency)
[A] Changed (Line 736).

Line 617 – Reword – Suggestion: 'The PhoDAR derived snow depth estimates have great potential for accurately characterizing snow depth over large regions, with an RMSE of () relative to the in situ snow depth measurements. The GPR snow depth estimates were slightly more accurate, with an RMSE of () relative to the in situ snow depth measurements, but required considerable more effort to obtain, and require complex post-processing to minimize errors associated with radar positioning and changing snow density.
[A] Changed (Line 739).

Line 620-621 - Please provide the quantitative results: correlation values and correlation lengths.
[A] We included (Line 747), '(the correlation coefficient of up to –0.8)'and '(the correlation range of 12.3 m)'

Line 622 – smooth snow surface is subjective (are there photos perhaps?) I thought the results showed that most of the snow depth distribution did not follow macrotopography (with the exception of the DTLB)
[A] We removed "smooth", and changed this sentence to (Line 749) "It is considered that the wind redistribution filled the microtopography by snow, and created a snow surface following macrotopography at the site."

Line 630 – Change 'extensible' to 'applicable for estimating both spatial and temporal variability of snow depth at other sites, and in other landscapes (prairie),'
[A] Changed (Line 757).

Line 860 – Figure 1 caption – The acronym NGEE is not explained in the text until the
[A] 'Next Generation Ecosystem Experiment' was added in Section 2.1 (Line 156).

Acknowledgement sections. Be consistent on whether to use the description 'NGEEArctic site' as the name for the full study domain throughout the text/figures. It would also be helpful to always refer to the long transects that traverse the study domain, as the '~500 m transects' to differentiate them from the transects surveyed at each individual plot – Please update all references to these '~500 m transects' throughout the paper.
[A] We added the acronym of NGEE-Arctic earlier in Section 2.1 (Line 156). We also added "500-meter transects" throughout the manuscript to clarify.

It would also be useful to use the same color ramp and scale for all snow depth maps and elevation maps (Figure 1, 3, 4, 7), with preferably different color ramps for snow depth vs elevation.
[A] We changed the color scale for all the snow depth maps. The scale for the elevation in Figure 3 was different from Figure 1. This was because it was necessary to visualize the microtopography in Figure 3, while Figure 1 needed to capture the overall elevation difference across the site.

Line 872 – Figure 3 – It would be helpful to delineate the extent of the 'fine grid' sample box with a black outline in all three figures for each plot to compare polygon features.
[A] The extent of the fine-grid measurements was included as black boxes in Figure 3.

Line 878 – Figure 4 – Can the authors overlay symbols of snow depth measurement location used to assess/validate the UAS product on the snow depth map in Figure 4b?
[A] Included.

It would also be useful to use the same color ramp and scale for all snow depth maps and elevation maps (Figure 1, 3, 4, 7), with preferably different color ramps for snow depth vs elevation.
[A] We changed the color scale for all the snow depth maps. The scale for the elevation in Figure 3 was different from Figure 1. This was because it was necessary to visualize the microtopography in Figure 3, while Figure 1 needed to capture the overall elevation difference across the site.

Figure 4c – state the elevation difference threshold used to define areas of low variability (red circles).
[A] We included, "the sub-meter elevation variability less than 0.05 m"

Line 884 – Figure 5 - boxplot 1 – change y-axis title from 'snow thickness' to 'snow depth' for consistency.
[A] Changed.

Line 889 – Figure 6 – describe in the caption what the dashed vertical lines represent and why only some colors are repsented.
[A] Added two sentences in the caption: "The different colors represent different plots (Plot A–D) or all the data (All). Each dash line represents the scale that maximize the magnitude of the correlation coefficient."

Line 894 – Figure 7 caption – Use 'NGEE-Arctic site' or 'full study domain' when describing the spatial extent of these figures.
[A] Changed.

It would also be useful to use the same color ramp and scale for all snow depth maps and elevation maps (Figure 1, 3, 4, 7), with preferably different color ramps for snow depth vs elevation.
[A] See the answers above (the repeated comments)

Line 900 – Figure 8 - Matching the color ramp scales would be helpful for comparing the STDEV results between the two methods of snow depth estimation.
[A] Changed.

Line 906 – Figure 9 - Use the same symbol size and line width for each plot.

[A] Changed. We would note that the points are different, since a new set of validation locations (100 points) were chosen randomly.