*Thank you to both reviewers and the editor for such careful and constructive comments. They are much appreciated! We respond to comments point-by-point in blue italics, and have used red text in the manuscript to indicate revisions.*

18-Dec-2017   
  
Dear Ms. Pulwicki,   
  
JOG-17-0123 entitled "Uncertainties in estimating winter balance from direct measurements of snow depth and density on alpine glaciers" which you submitted to the Journal of Glaciology, has been reviewed.  The comments of the reviewer(s) and editor are included at the bottom of this letter.   
  
While the reviewer(s) and editor recognise the potential of your manuscript as a valuable contribution to the journal, they also suggest some major revisions to your manuscript. Therefore, I invite you to respond to their comments and revise your manuscript.   
  
To start your revision now, click the link below:   
  
\*\*\* PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. \*\*\*   
  
https://mc.manuscriptcentral.com/jog?URL\_MASK=c4f6bc6939974504b4e71a0271d844e7   
  
Alternatively, you may log into your Author Centre at https://mc.manuscriptcentral.com/jog, where you will find your manuscript under "Manuscripts Awaiting Revision". Upon submission of the revised version, your manuscript number will be appended to denote a revision.   
  
When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the space provided.  Please use this space to document any changes you make to the original manuscript.  In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).   
  
Because we are trying to facilitate timely publication of manuscripts submitted to the Journal of Glaciology, your revised manuscript should be uploaded as soon as possible. We expect to receive your revision by 17-Jan-2018. If it is not possible for you to submit your revision by this date, please contact the Editorial Office to rearrange the due date. Otherwise we may have to consider your paper as a new submission.   
  
Once again, thank you for submitting your manuscript to the Journal of Glaciology and I look forward to receiving your revision.   
  
Best regards,   
Dr. Nicolas Eckert   
Scientific Editor, Journal of Glaciology   
nicolas.eckert@irstea.fr   
  
Scientific Editor: Eckert, Nicolas   
Comments to the Author:   
(There are no comments.) 

Reviewer: 1   
  
Comments to the Author   
General comments to the authors   
  
This paper presents a large number of winter balance measurements over three glaciers. Through the statistical analysis of these measurements, the study estimates glacier-wide winter mass balance and the associated uncertainties. This study is a valuable contribution since understanding the uncertainties and modeling winter mass balance variability remains challenge within the scientific community. The large number of measurements is impressive and an important contribution in itself. With some minor revisions, I recommend that this study for publication.   
  
This study is very thorough and contains a lot of information. While this information is useful, it the main objective is often lost in the details. I would suggest reorganizing the structure and splitting the “Result and Discussion” section in two distinct sections. For instance, the result part could include only the glacier-wide mass balance computed with LR and KG methods based on data measurements with the associated uncertainties and the comparison of these two approaches. This seems to be the main objective of the study. The second part of the discussion would include the spatial variably of winter balance at glacier scale (with the influence of topographic parameters), as well as the difference between glaciers and a discussion on the complexity of obtaining a transferable relationship, as mentioned in this paper. The regional winter balance gradient could also be mentioned here. The section “Limitations and future work” could also be included in the second part of the discussion. Please, refer to my comments below for more details.

*We agree with the reviewer’s statement that the main point of the paper is lost in the details of the results section. We originally chose to merge the Result and Discussion sections because each subsection builds on the previous one (e.g. the assumptions and results of density assignment are needed to understand the results and discussion of the interpolated winter balance values). Therefore, we do not want to reorganize these sections in the way suggested by the reviewer because then the main interpolation results would not be built on the results of the previous steps. Further, the suggested arrangement would mean that there would be a large proportion of results presented in the Discussion section.*

*We propose an altered restructuring of the Results and Discussion. In the Results sections we now present the results from Field measurements, Density assignment and Gridcell-averaged winter balance, as before, and present a brief overview of the results for Distributed winter balance and Uncertainty analysis using a Monte Carlo approach. The Discussion now contains only the interpretation of Distributed winter balance and Uncertainty analysis using a Monte Carlo approach, which includes a discussion of important topographic parameters, transfer of regression coefficients and regional context. We hope that since the Discussion is focused on the interpolation and uncertainty analysis, the main ideas of the paper will be clearer. We note that in this arrangement, a small amount of interpretation of the results from Field measurements, Density assignment and Gridcell-averaged winter balance is present in the Results section but we believe this interpretation is short enough that it fits best in the Results section.*

My second general comment is that the method used to calculate the LR is interesting but complex. Thus this approach could be explained in more detail (by providing a scheme and/or some equations).

We have expanded on the LR methods section in a way that we hope makes the LR process more clear and reproducible. Although we did not provide any additional equations or schemes, we divided the cross-validation and model-averaging descriptions into two paragraphs that build on each other, indicating that cross validation is nested within model averaging.

Regarding the formatting, each section should be assigned a section number.

*We used the LaTex formatting provided by IGS and it did not include section titles. Not sure how to add them…*

Detailed comments   
  
Introduction   
- Line 33: Please add a reference

“Winter balance (WB) is notoriously difficult to estimate *(e.g. Dadi ́c and others, 2010; Cogley and others, 2011)* “

- Lines 34 to 39: Are there the only reason? Isn’t it also because the access on these glaciers during the winter period is more difficult?

“Simultaneously extensive, high resolution and accurate snow distribution measurements on glaciers are therefore difficult to obtain (e.g. Cogley and others, 2011; McGrath and others, 2015) *and is further complicated by the inaccessibility of many glacierized regions during the winter.*”

- Lines 62-63: A reference to the work made by Cullen et al., (2017) based on a co-kriging method could be mentioned here.

“…as well as geospatial interpolation (e.g. Erxleben and others, 2002; *Cullen and others, 2017)* including various forms of kriging.”

- Line 64: The SnowTran-3D model (Liston and Sturm) could be mentioned here.

“Physical snow models such as *SnowTran-3D (Liston and Sturm, 1998),* Alpine3D (Lehning and others, 2006), and SnowDrift3D(Schneiderbauer and Prokop, 2011) are widely used,…”

Study Site   
- Fig.1: Grey squares are not visible. Please change color or/and extend size.

*Grey squares have been changed to red squares*

- Table 1: It is surprising that glacier 2 and glacier 4 have the same ELA as there are oriented in the opposite direction!   
It will be interesting and useful for future discussions to add in this Table the elevation range of the measurements, and the % of the glacier covered by these measurements (e.g. using the % of grid cell of the DEM containing at least one measurement or something similar).

*Since both reviewers have asked for additional information about the survey, we have split up Table 1 into two tables. Table 1 now contains only information about the study glaciers and Table 2 only contains information about the survey. The elevation range and percent area covered (for the entire glacier and of the ablation area only) of measurements have been added to Table 2.*

Methods   
- Line 102: To facilitate the reading, add a paragraph (before “sampling design”) with general information (mentioned below) about the field campaign (e.g. Line 119, lines 122-123, see my comments below).

*The first paragraph of the Field Methods section is now a general description of the field campaign. The second paragraph describes the accumulation and melt events experienced during the campaign.*

- Line 114: Is there a reference for this method?

“To capture variability at the grid scale, we densely sampled up to four gridcells on each glacier using a linear-random sampling design *(Shea and Jamieson, 2010)* we term a ‘zigzag’.”

- Line 115: Fig. 1f (not 1e)

*…(Fig.1f)…*

- Line 119: This is not only about the snow depth. You could add this sentence in the “introduction paragraph” of the field measurements.

*Moved (see first paragraph of Field Measurements)*

- Lines 122-124: It is also about the campaign in general. Add this in the “intro paragraph”.

*Moved (see first paragraph of Field Measurements)*

- Lines 135-136: There is here mixing information between snow depth and density. Please mention here only the snow depth measurement method as the density method is explained below.

“Successful snow depth *~~and density~~* measurements within the accumulation area were made either in snow pits or using a Federal Sampler (described below) to unambiguously identify the snow--firn transition.”

- Lines 138-139: This has already been said in the section “Sampling design”. So you can delete this sentence and add the reference “Shea and Jamieson, 2010” on line 114.

*Sentence deleted and reference added*

- Lines 143-144: Is there a specific reason for this fourth zigzag measurement?

*“Extra time in the field allowed us to measure a* fourth zigzag on Glacier 13 in the central ablation area at ~2200 m a.s.l.“

- Line 146: This is quite confusing. I suggest to delete “as well as…Federal Sampler”; or add the number of density measurements made using this method, and start the next sentence (snow pit description) in a new paragraph to make it clearer.

“Snow density was measured using a Snowmetrics wedge cutter in three snow pits on each glacier*~~, as well as with a Geo Scientific Ltd. metric Federal Sampler~~*.”

- Line 146: Please define the acronym (SP) here (i.e. the first time that you mention it) and then, only use the acronym (e.g. line 147, 158…) to be consistent.

*“Snow pit” replaced with “SP” for all instances after defining the acronym*

- Line 147: Same remark for Federal Sampler acronym (SF).

*“Federal Sampler” replaced with “FS” for all instances after defining the acronym*

- Lines 149-155: To be consistent, if the density measurement uncertainty is mentioned here, the snow depth measurement uncertainty should also be quantified in the previous section (i.e. uncertainty due to probe and snow pits methods).

*The depth measurement uncertainty is quantified as the standard deviation of point-scale depth measurements at a single measurement location. The glacier-wide mean of those standard deviations is taken to be the depth uncertainty for the given glacier. The following text has been added to address the reviewer’s comment:*

*“The 3-4 point-scale depth measurements are averaged to obtain a single depth measurement at each transect measurement location. When considering snow variability at the point scale as a source of uncertainty in snow depth measurements, we find that the mean standard deviation of point-scale snow depth measurements is found to be <7% of the mean snow depth for all study glaciers.”*

- Line 165: FS method uncertainty should also be evaluated to remain consistent.

*The FS density uncertainty is evaluated as the standard deviation of the FS measurements at a single location and the average of those values for a given glacier is taken as the overall uncertainty. The following text has been added to address the reviewer’s comment:*

*“The mean standard deviation of FS-derived density was <4% of the mean density for all glaciers.”*

- Lines 166-172: This is general information about the field campaign, and concern also the snow depth. This paragraph could be in the “introduction paragraph” mentioned above.

*Moved (see first paragraph of Field Measurements)*

- Lines 171-172: It could be evaluated using a modeling approach but this is not the aim of the study. So you could at least provide an estimation of the uncertainty related to these events.

*The reviewer has asked for an estimate of the amount of accumulation related to two precipitation events that occurred on Glaciers 4 and 2 during the field campaign and the amount of melt that occurred on Glacier 13 at the end of the field campaign. To address the estimation the accumulation event we have added the following text:*

*“During the field campaign there were two small accumulation events. The first, on 6 May 2016, also involved high winds so accumulation could not be determined. The second, on 10 May 2016, resulted in 0.01 m w.e accumulation measured at one location on Glacier 2. Assuming both accumulation events contributed a uniform 0.01 m w.e accumulation to all study glaciers then our survey did not capture ~3% and ~2% of estimated Bw on Glaciers 4 and 2, respectively. We therefore assume that these accumulation events were negligible and apply no correction.”*

*To address the amount of melt…*

- Line 180: These acronyms have already been defined before. So please use only the acronym.

“Densities derived from *SP and FS* measurements are treated separately…”

- Line 183: Information about DEM is only available in the supplementary material. Please mention here the year and the source of the DEM.

*The following sentence has been added to the paragraph that first mentions the DEM (first paragraph of the Methods section, rather than Line 183 as suggested):*

*“We use the SPIRIT SPOT-5 DEM (40x40 m) from 2005 (Korona and others, 2009)*

*throughout this study.”*

- Line 189: The uncertainty that you mention here is not clear. The uncertainty concerns the snow depth difference between two observers distant from 10m, or is it the difference in measurement when the next observer arrives at the location of the one before him? And how the uncertainty is evaluated?

*We have changed ‘uncertainty’ to ‘error’ to better reflect the analysis that was conducted and also further explain the analysis. The sentences now read as follows:*

*“Error due to having multiple observers is also evaluated by conducting an analysis of variance (ANOVA) of snow-depth measurement along a transect and testing for differences between observers. We find no significant differences between snow-depth measurements made by observers along any transect (p>0.05)…“*

- Line 191: Can you quantify this difference? Is it taken into account?

*We have added additional information regarding significant differences between observers for one transect:*

*“We find no significant differences between snow-depth measurements made by observers along any transect (p>0.05), with the exception of the first transect on Glacier 4 (51 measurements), where snow depth values collected by one observer were, on average, greater than the snow depth measurements taken by the other two observers (p<0.01). Since this was the first transect completed and the only one to show differences by observer, this difference can be considered an anomaly. This result shows that observer bias is likely to not affect the results of this study and no corrections to the data based on observer were applied.”*

- Line 197: Acronym OK already defined line 194. So please use only the acronym here and above (e.g. lines 224, 226, 232…)

*OK acronym has been implemented where recommended*

- Line 194: “Regression” or “Linear regression”, please be consistent in the article and use the acronym once defined.

*LR acronym has been implemented where recommended*

- Line 201: The wind-redistribution parameter should be defined (and not only in the supplementary material). Please add “Sx” and the reference to Winstral and Marks’study.

*“…and a wind-redistribution parameter (Sx from Winstral and others (2002));…”*

- Lines 207-208: Refer to my comment on the Supplementary material.

*See response to comments on Supplementary Material*

- Lines 199-222: This is an interesting and complete method which have been used here, but also quite complex. This paragraph mentions many references which need to be read to properly understand the method. I suggest providing more information and detail about the approach used. Some equations and/or a scheme could help the understanding. I think this is important, as it is one of the main objectives of this study.

*We have provided more details about the LR methodology used. We hope that procedure is now more clear and reproducible. See Methods: Distributed winter balance: Linear regression for changes.*

- Line 224: Use only the acronym.

*OK acronym has been implemented where recommended*

- Line 232: “topographic regression” has not been defined, even if we can guess what is it. Please be consistent when using “Linear regression”, “regression” and “topographic regression” after defined it to avoid confusions.

*LR acronym has been implemented where recommended*

- Lines 232-234: So why this method has been applied? Please explain and justify why you choose to use this approach.

*We chose to use OK because it allows for an interpretation-free interpolation of data. We have added the following sentence to address this point:*

“Unlike LR, OK is not useful for generating hypotheses to explain the physical controls on snow distribution, nor can it be used to estimate winter balance on unmeasured glaciers. *However, we chose to use OK because it does not require external inputs and is therefore an interpretation-free method of obtaining Bw.”*

- Line 235: The specific section “uncertainty analysis” is confusing. In reading the article we expect to find all information about uncertainties in this section (including an evaluation of measurements uncertainties). I suggest renaming the section (e.g. “Uncertainty evaluation on glacier-wide WB” or “Uncertainty analysis using a Monte Carlo approach”, or something similar).

*All of these sections have been renamed to “Uncertainty analysis using a Monte Carlo approach”*

Results and discussion   
- Fig 2.: Should be in this section, not before.

*Moved, see comments regarding Discussion at the beginning of this document*

Fig. 2b: Why the point G13\_ASP doesn’t have an x-error bar?

*x-error for G13\_ASP is too small to be seen in the figure.*

In the legend, change “Labels indicate snow pit locations” by “Labels indicate SP locations”

“Labels indicate *SP* locations…”

- Line 278: Add “a” as following: “Fig. 2a”.

“(Fig. 2*a*)”

- Fig. 3: Please use more contrasted colors. For instance, it is almost impossible to see the difference between M2 and U in Fig 3c.

*Fig. 3c colours have been changed to provide more contrast (greater difference in shades of purple)*

- Lines 288, 291,297,302,307… Change Feredal Sampler by FS.

*FS acronym has been implemented where recommended*

- Line 289: The diameter size information should be in the method section.

*Moved to the “Snow density” subsection:*

“Therefore, a Geo Scientific Ltd. metric Federal Sampler (FS) (Clyde 1932) *with a 3.2-3.8 cm diameter sampling tube*, which…”

- Line 302-303: Can you be more specific and provide more information regarding the “the conditions at the time of the sampling”.

*The following sentence is added to specify the snow pack conditions at the time of sampling:*

*“At the time of sampling the snow pack had little fresh snow, which confounds the low density values, and was not yet saturated and had few ice lenses, which confounds the high density values.”*

- Line 311: Please justify why do you use the FS density measurements, while you said that there is a bias that you cannot correct.

*The following sentence is added at the end of the paragraph to justify using FS values:*

*“Despite this bias, we use FS-derived densities to generate a range of possible WB estimates and to provide a generous estimate of uncertainty arising from density assignment.”*

- Line 323: Where are these results? Are they “not shown”? As you have a supplementary material, it could be used to insert these results.

*A plot of the distributions of gridcell-averaged WB values for individual glaciers has been added to the supplementary material and the referenced sentence has been changed accordingly:*

“…are similar to those in Fig. 2a but with fewer outliers *(see Supplementary Material, Fig. S3)*.”

- Lines 324-326: Same remark than above.

*A plot of the standard deviation of point-scale WB values within a gridcell measured along transects is now provided in the supplementary material and the referenced sentenced changed accordingly:*

“…gridcell measured along transects *(see Supplementary Material, Fig. S4)*.”

- Line 327-328: “We nevertheless… gridcells”: can be deleted as it has already been said in the method part.

*~~We nevertheless assume that the gridcell uncertainty is captured with dense sampling in zigzag gridcells.~~*

- Fig. 4: This figure could be larger. “Topographic parameters” could be added for the x-axis.

*These changes have been made*

- Fig. 5: Information about the date (winter 2016, measurements performed in may) could be added in the legend.

“Locations of snow-depth measurements *taken in May 2016* are shown as black dots.”

In this figure the LR method used remains unclear. Have topographic parameters been used?  I also don’t really understand how this figure indicates that elevation and Sx are the most significant predictors (mentioned in line 333).

*The LR and OK methods have both been clarified in the figure caption:*

*“The linear regression method involves multiplying regression coefficients, found using cross validation and model averaging, by topographic parameters for each gridcell. Ordinary kriging uses the covariance of measured values to find a set of optimal weights for estimating values at unmeasured locations.”*

*The reference to Fig 5 in Line 333 has been removed.*

- Lines 331-335: We can also see that the larger correlation is obtained for the glacier covering the largest elevation ranges (and vice versa). This could be mentioned. In addition melting events don't seem to affect the glacier 4 as stronger as for the two other one (also mention below).   
So do you think that the low correlation with the elevation for glacier 4 is due to the absence of snow melt event or that the glacier covers low elevation ranges, or a combination of both? To provide better precision, the correlation could be computed in the upper part of the glacier less affected by melting events. I think it is important and relevant to discuss this point, as most of the studies indicate the elevation as the main predictor of the WB variability, but it could be principally due to winter melting events...

*From field observations of the snow pack on all three glaciers, we hypothesize that the presence/absence of an elevation gradient in bw result from different physical processes for each glacier. The snow on Glacier 13 was isothermal and showed clear signs of radiation melt at the surface (as mentioned in the Field Methods section). No signs of melt were observed within the snow packs of Glaciers 2 and 4. Glacier 2 had a strong elevation gradient within the ablation area due to minimal snow accumulation at the terminus, which is steep and wind-affected. We hypothesize that Glacier 4 did not have an elevation gradient because the snow has experienced a considerable amount of wind redistribution. The snow surface was relatively uniform but depths varied in a way that appeared to follow the underlying ice surface, indicating that the snow had likely smoothed the large-scale ice roughness. We do not think that the differences in elevation gradient between our study glaciers are a result of elevation range differences. Glacier 4 has a maximum elevation that is ~300 m less than that of Glacier 2, which would not result in a melt effect as large as the difference in elevation gradient observed between these glaciers. We hypothesize that differences in elevation ranges would have a detectible effect on large glaciers because the elevation effect on snow distribution would be more prominent. Unfortunately, we cannot compute a correlation in the upper portion of the glacier because we have too few data in these areas. To address the reviewer’s comment and add our field observations of snow distribution, we have added the following sentences to the Discussion:*

“As expected, gridcell-averaged bw is positively correlated with elevation where the correlation is significant. It is possible that the elevation correlation was accentuated due to melt onset for Glacier 13 in particular. *Glacier 2 had little snow at the terminus likely due steep ice and wind-scouring but the snow did not appear to have been affected by melt.* … The low of correlation between $b\_\mathrm{w}$ and elevation on Glacier 4 is consistent with \cite{Grabiec2011} and \cite{Lopez2011}, who conclude that highly variable distributions of snow can be attributed to complex interactions between topography and the atmosphere that cannot be easily quantified. *The snow on Glacier 4 also did not appear to have been affected by melt and it is hypothesized that significant wind-redistribution processes, that were not captured by the Sx parameter, covered ice-topography and produced a relatively uniform snow depth across the glacier*.”

- Line 335-342: This sentence provides general information and should be mentioned in the introduction. References to other studies should be mentioned as a comparison with your results/ to discuss your results (e.g. “Our results are in good agreement with previous studies (references) which have found the elevation to be the…”   
Same remark regarding the variability of WB-gradients between glaciers: your findings should be mentioned here to be compared with the reference mentioned.

*The first paragraph in the Results: Linear Regression section now contains the following text, which we hope addresses the reviewer’s comments:*

*“Our results are consistent with many studies that have found elevation to be the most significant predictor of winter-balance data (e.g. Machguth and others, 2006; McGrath and others, 2015). The WB–elevation gradient is 13 mm/100 m on Glacier 2 and 7 mm/100 m on Glacier 13. These gradients are consistent with those reported for a few glaciers in Svalbard (Winther and others, 1998) but considerably smaller than many reported WB–elevation gradients, which range from about 60–240 mm/100 m (e.g. Hagen and Liestøl, 1990; Tveit and Killingtveit, 1994; Winther and others, 1998). Extrapolating linear relationships to unmeasured locations typically results in large uncertainties, as seen by the large WB values (Fig. 5) and large relative uncertainty (Fig. 6) in the high-elevation regions of the accumulation areas of Glaciers 2 and 13. The low correlation between WB and elevation on Glacier 4 is consistent with Grabiec and others (2011) and L ́opez-Moreno and others (2011), who conclude that highly variable distributions of snow are attributed to complex interactions between topography and the atmosphere that could not be easily quantified.”*

Same remark regarding slope, wind (mentioned then at lines 346-361), ect…

*The second paragraph in the Results: Linear Regression section now contains the following text, which we hope addresses the reviewer’s comments:*

*“Our results corroborate those of McGrath and others (2015) in a study of six glaciers in Alaska (DEM resolutions of 5 m) where elevation and Sx were the only significant parameters for all glaciers; Sx regression coefficients were smaller than elevation regression coefficients, and in some cases, negative. While our results point to wind having an impact on snow distribution, the wind redistribution parameter (Sx) may not adequately capture these effects at our study sites. For example, Glacier 4 is located in a curved valley with steep side walls, so specifying a single cardinal direction for wind may not be adequate. Further, the scale of deposition may be smaller than the resolution of the Sx parameter estimated from the DEM. Creation of a parametrization for sublimation from blowing snow, which has been shown to be an important mechanism of mass loss from ridges (e.g. Musselman and others, 2015), may also improve explanatory power of LR for our study sites.*

I think you really have interesting results, but it is quite hard to follow the discussion. I therefore suggest reorganizing this paragraph (from line 335 to 370) and splitting it into a result and a discussion section (as mention in my main comment). Here, only mention the main results of linear regression should be included. In the discussion part, discuss the spatial variability of WB (at glacier scale / between glaciers / at regional scale) and compare your result with previous studies. Note that the section “Regional winter balance gradient” as well as Fig.9 could be insert in this section.

*See comments at the beginning of the document regarding restructuring.*

Lines 360-361: Yes, but how? Using models?

*This sentence has been deleted as part of the changes in the previous comment.*

Line 380: The kriging method always gives very bad extrapolations!

“*As expected,* extrapolation using OK leads to large uncertainty.”

Line 386-387: I understand that the 1/3 kept was only necessary to compute LR uncertainty, so the LR method is based on 2/3 of the data. So to make the comparison more consistent, the KG uncertainty could also be computed using only 2/3 of the data.

*We have computed the OK estimate with 2/3 of the data and the results are consistent with results from using the full data set. We have made the following change to indicate this consistency:*

“This comparability is interesting, given that all of the data were used to generate the OK model, while only 2/3 were used in the LR *(consistent with the best OK model estimated with 2/3 of the data)*.”

Line 393: Probably because there is less data measurements in the accumulation area. As KG extrapolations usually lead to large uncertainties, maybe the comparison between the 2 approaches could be reduced over the measurement area (i.e. only based on interpolations).

*The OK and LR estimates are compared for just the ablation area within this sentence. Since our measurements span the majority of the ablation area there is limited extrapolation. Further, the two methods show good agreement (estimates differ by <7%). Therefore, we do not further constrain the area of comparison.*

Line 397: interpolation uncertainties or interpolations and extrapolations?

“…and interpolating*/extrapolating* WB values across the domain…”

Line 397-398: So why not computing topographic regression only over the measurement area?

*The goal of most winter-balance studies is to estimate a glacier-wide winter balance value, which requires extrapolation of winter balance values beyond the measurement area. We consider interpolation/extrapolation uncertainty over the entire glacier because it influences the uncertainty in glacier-wide winter balance. We therefore make no changes pertaining to this comment.*

Figure 8: Some colors are really difficult to see. Please use more contrasted colors.

*We use yellow, purple, and green to indicate the three study glaciers and these three colours are well contrasted. There are multiple curves for each glacier, which result from different density assignment methods, which are all the same colour but with transparency. These curves cannot be contrasted because they are the same colour. As a result, no change has been made.*

Line 405: Same remark than above: the comparison could be only based on the measurement area using the interpolation (and not the extrapolation).

*See reply above*

Table 4: In the legend please add (WB) to define the acronym.

*Added*

Line 415: “…from a number of sources” Please indicate what are these sources (e.g. snow depth measurements?)

*The referenced paragraph has been changed to be as follows:*

*“Our analysis did not include uncertainty arising from density measurement errors associated with the FS, wedge cutters and spring scales, from vertical and horizontal errors in the DEM or from error associated with estimating measurement locations based on the GPS position of the lead observer. We assume that these sources of uncertainty are either encompassed by the sources investigated or negligible.”*

Line 418: It is unclear according me. Could you provide more information?

“…or from error associated with estimating measurement locations *based on the GPS position of the lead observer*.”

Line 422: What are here the “combined sources of uncertainty”?

“The glacier-wide WB uncertainty from *the three investigated* sources of uncertainty ranges from…”

Line 425-437: I would recommend to include this paragraph in a section discussing the spatial variability of WB (please refer to my remark above and my main comment.

*This section has been moved to the Discussion (see comments regarding restricting at the beginning of this document.*

Lines 454-455: This could be mentioned in introduction.

*The methods referred to in these lines are mentioned in the introduction:*

“…a wide range of observational  techniques, including direct measurement of snow depth and density (e.g. Cullen and others, 2017), lidar  or photogrammerty (e.g. Sold and others, 2013) and ground-penetrating radar (e.g. Machguth and others,  2006; Gusmeroli and others, 2014; McGrath and others, 2015).

*We therefore make no changes.*

Conclusion   
Line 507: I suggest to mention the principal limit of the study as well as future works here.   
 *The follow sentence has been added to the end of the conclusion paragraph:*

*“The major limitations of our work include the restriction of our data to a single year and minimal sampling in the accumulation area.”*

Supplementary material   
General comment: The method is here briefly described. As you are not limited, you could provide more information.

*We have provided two additional sections in the supplementary material to address these comments. Within the text we have also references the MSc thesis written by the first co-author, which has extensive details on topographic parameter calculation.*

More detailed comments:   
- The method to obtain the smoothed DEM is unclear. Could you provide more information?

*An additional section called “DEM smoothing” has been added.*

- Can you provide more information about the Sx index computation? For instance, is different angle sizes and distances have been tried?

*An additional section called “Wind redistribution parameter” has been added.*

Reviewer: 2 

Report on ”Estimating winter balance and its uncertainty from direct measurements of snow depth and density on alpine glaciers”

Alexandra Pulwicki, Gwenn E. Flowers, Valentina Radic

The aim of this work is to estimate accurately winter surface mass balance and evaluate its uncertainty on three glaciers in the St. Elias Mountains, Yukon, Canada by statistical methods and from direct measurements. 9000 manual measurements have been collected through a sophisticated sampling design aiming at catching both the variability at different scales and the heterogeneity across the glaciers. The paper is well written and well organized, the figures and tables are relevant and well documented.

*Thank you!*

I am not specialist in glaciers study, rather in environmental statistics, I will not comment extensively the part of the paper concerning the measurements and the discussion on the interpretation of the results, just to say that as far as I can judge the goal of the study is an important point to assess, the measurements collected form a very valuable dataset for this domain of research and that this paper and its results deserve being published. I will focus mainly on the statistical methods, for which I have several concerns.

1. First it is likely obvious for those who work on glaciers, but a definition of glacier-wide winter balance and how it is derived from the distributed estimates of winter balance is lacking. I guess it is the average of the distributed values, but it should be precised.

*This is a good observation. We have added the following sentence in the first paragraph of the Methods section to specify that we took the average of all gridcells.*

“Glacier-wide WB is then calculated by taking the average of all gridcell-averaged WB values for each glacier.”

*Further, we have decided to use Bw to indicate glacier-wide winter balance throughout the manuscript to be consistent with the Glossary of glacier mass balance and related terms (Cogley and others, 2011). This change in notion was informally suggested by colleagues after submission. We hope that using Bw will improve clarity of the process used to calculate glacier-wide winter balance. Point- and gridcell-scale winter balance are now denoted as bw to further comply with the Glossary of glacier mass balance and related terms.*

*Cogley J, Hock R, Rasmussen L, Arendt A, Bauder A, Braithwaite R, Jansson P, Kaser G, Moller M, Nicholson L and others (2011) Glossary of glacier mass balance and related terms. UNESCO-IHP, Paris*

1. I did not find an information that is crucial for me: in how much gridcells were there point-scale values to average and what is the total size of the 40 x 40m DEM grid? It is stated line 183 that the point-scale values are averaged in each 40 x 40m DEM gridcell, but I suppose that there is not point-scale values in each gridcell.

*We agree that this is important information to include. We have added the requested information to Table 2.*

1. About the Linear Regression procedure, I am not sure on how it is really performed. It is stated that first, cross validation is used to obtain a set of βi and second model averaging is used to account for uncertainty when selecting the predictors, but how are estimated the βi in the cross validation step? If it is by model averaging, then the two step are nested and the presentation of the procedure needs to be reformulated.

*We have provided more details about the LR methodology used. We hope that the procedure is now more clear and reproducible. See Methods: Distributed winter balance: Linear regression for changes.*

1. I have more concerns with the kriging part.

(a) Simple kriging needs the knowledge of the expectation of the random field in all the sites to be predicted and which are measured. How is it estimated? If it is assumed constant over the glacier its value should be (in my opinion) the glacier- wide winter balance and this is just what is intended to be estimated from the kriged values.

*See response to 4(c)*

(b) No model is given for the covariance function, as the fitted parameters are the range and the nugget, I guess that it would be an exponential or a Gaussian covariance function, but this needs to be clarified and justified.

*We have added details about the covariance function.*

“A Matere covariance function with ν=5/2 is used to define a stationary and isotropic covariance and covariance kernels are parameterized as in Rasmussen and Williams (2006).”

(c) For that kind of data it is usual to use rather ordinary kriging with a fitted variogram, the advantage is that the expectation can be unknown and needs not being estimated, and the stationarity assumption is relaxed to an assumption of stationary increments.

*The reviewer is correct that simple kriging requires defining the expectation (or mean) of the field that is to be estimated and since the goal of our work is to obtain this mean, simple kriging is inappropriate. To address this comment, we further investigated how we were implementing the kriging program (DiceKriging) by enlisting additional collaborators in the Department of Statistics at SFU. Our collaborators clarified that DiceKriging is a two stes process: (1) Define or estimate a set of kriging parameters (e.g. nugget, theta, covariance, mean, trend parameters) and (2) Apply the kriging parameters to a grid using either simple kriging (i.e. no trend) or universal kriging (i.e. with a trend). In our original analysis, we had allowed the program to estimate all parameters (i.e. not define the mean) and then apply those parameters to a grid using simple kriging. Since we had not defined a mean for the model, we had in fact been using ordinary kriging to estimate winter balance (Roustant and others, 2012). Therefore, we had simply used the incorrect terminology for a correct kriging procedure.*

*However, our collaborators did inform us that our set-up for step (1) did not allow for stable solutions to be estimated. To create stable solutions, we needed to increase the number of fitting attempts. Therefore, we have re-run the analysis in step (1) to provide better estimates of distributed bw and the results have been updates accordingly. This change in analysis did not produce any large changes in the results. The only noticeable change is that the uncertainty resulting from gridcell-scale variability is smaller (Fig xx and Table xx).*

(d) Even more sophisticated kriging may be performed accounting for a trend, for in- stance universal kriging that use the geographical coordinates, or regression kriging which kriges the residuals of a regression, it could be for instance the LR previously introduced.

*Prior to submitting this manuscript, we had already completed regression kriging. We did not include the results in our manuscript because the document already contained a large volume of information and we felt that the regression kriging results did not provide additional insight to the estimates of winter balance. We now include the regression kriging results in the Supplementary Material.*

5. The uncertainty analysis is driven by different means according to the steps of the procedure. It is indicated at the beginning of the section that repeated random sampling is performed according to different input variables. It is the case for the uncertainty linked to the grid scale variability, 1000 values are drawn from a zero mean normal distribution with standard deviation equal to the mean standard distribution of zigzag.

(a) The inputs to evaluate the density assignment uncertainty are the eight density interpolation methods hence eight values, is then the density assignment uncertainty the standard deviation of the eight resulting values for the glacier wide winter balance? This would be very far from the announced 1000 sample values.

*This comment is correct, we use eight values to calculate the standard deviation for density assignment uncertainty. To clarify this confusion, we have rearranged the information presented in the first paragraph of the Methods: Uncertainty analysis section and added the following sentence:*

“Density assignment uncertainty is calculated as the standard deviation of the eight resulting values of glacier-wide winter balance.“

(b) The way that is calculated the interpolation uncertainty for the simple kriging is weird. A standard deviation is derived from the confidence interval given by the package DiceKriging, but usually the confidence interval is calculated from the kriging standard deviation, why this value is not used directly? I don’t know if it is an output for this package, but it is without any doubt for most standard kriging packages (gstat, geoR, spatial, fields, RandomFields ...).

*Upon investigation, DiceKriging does directly output the standard deviation of estimated values. The results for simple kriging are equivalent when the standard deviation is used directly and when the standard deviation is derived from the confidence interval. However, we now use the standard deviation values directly to simply analaysis. The section that details the calculation of simple kriging uncertainty now reads as follows:*

“OK interpolation uncertainty is represented by the standard deviation for each gridcell-estimated value of WB generated by the DiceKriging package. “

The most serious concern on this work, is the use of simple kriging, which seems not relevant in this framework. Some points need to be clarified (number of gridcell averaged values, steps of the LR procedure, calculation of the uncertainty of the density assignment and the interpolation assignment).

I recommend a major revision, as this work develop many interesting points and deserves being published.