Dear Editor,
We revised the manuscript in accordance with the comments from the reviewers. We now feel that the paper is more clear and understandable. We wish to thank the reviewers for the helpful suggestions, which we believe strongly improved our manuscript.

Please find below the detailed comments and responses to the Reviewers' suggestions. In the revised manuscript, the changes are colored in yellow.

We hope that the revised manuscript can meet your and the Referees' expectations, and be accepted for publication on the journal; otherwise, we are open to new improvements.

Many thanks for your help.
Best regards,

Antonella Senese and co-authors

Referee #1

Dear Referee

Thanks for helping us in improving the manuscript.

We have considered all your comments in the revised version of the manuscript. The title, as you suggest and in agreement also with the second Referee, is: Estimating the snow water equivalent on remote and glacierized high elevation areas (Forni Glacier, Italy).

We have therefore modified the "Method" section in order to better explain the approach applied for estimating the site-average-new-snow density ($\rho_{new\,snow}$) and the sonic-ranger-depth-derived *SWE*. In the previous version of the manuscript, we performed the analyses using the mean new snow density (140 kg m^{-3}) that was obtained by Senese et al. (2012) considering the 2005-2009 dataset, and then we discussed to what degree this value is able to describe the data from the years to come. In the revised version of the manuscript, we have started with updating the site average new snow density estimation exploiting all the available datasets and we have performed all the subsequent analyses using this value.

The text concerning this issue in "Method" section is:

> In addition to the measurements recorded by the AWSs, since winter 2005-2006, personnel from the Centro Nivo-Meteorologico (namely CNM Bormio-ARPA Lombardia) of the Lombardy Regional Agency for the Environment have periodically used snow pits (performed according to the AINEVA protocol, see also Senese et al., 2014) in order to estimate snow depth and *SWE*. In particular, for each snow pit *j*, the thickness (h_{ij}) and the density (ρ_{ij}) of each snow layer (*i*) are measured for determining its snow water equivalent, and then the total $SWE_{snow\text{-}pit\text{-}j}$ of the whole snow cover (*n* layers) is obtained:
>
> $$SWE_{snow-pit-j} = \sum_{i=1}^{n} h_{ij} \cdot \frac{\rho_{ij}}{\rho_{water}} \qquad (1)$$
>
> where ρ_{water} is water density. As noted in a previous study (Senese et al., 2014), the date when the snow pit is dug is very important for not underestimating the actual accumulation. For this reason, we considered only the snow pits excavated before the beginning of snow ablation. In fact, whenever ablation occurs, successive *SWE* values derived from snow pits show a decreasing trend (i.e., they are affected by mass losses).
>
> The snow pit *SWE* data were then used, together with the corresponding total new snow derived from sonic ranger readings, to estimate the site average $\rho_{new\,snow}$, in order to update the value of 140 kg m^{-3} that was found in a previous study of data of the same site covering the 2005-2009 period (Senese et al., 2012a). We need to update our figures for $\rho_{new\,snow}$ as it is the key variable for estimating *SWE* from the sonic ranger's new snow data. Specifically, for each snow pit *j*, the corresponding total new snow was first determined by:
>
> $$\Delta h_{snow-pit-j} = \sum_{t=1}^{m}(\Delta h_{tj}) \qquad (2)$$
>
> where *m* is the total number of days with snowfall in the period corresponding to snow pit *j* and Δh_{tj} corresponds to the depth of new snow on day *t*. In particular, we considered the hourly snow depth values recorded by the sonic ranger in a day and we calculated the difference between the last and the first reading. Whenever this difference is positive (at least 1 cm), it corresponds to a new snowfall. All data are subject to a strict control to avoid under- or over-measurements, to remove outliers and nonsense values, and to filter possible noises. $\sum_{t=1}^{m}(\Delta h_{tj})$ is therefore the total new snow measured by the Campbell SR50 from the beginning of the accumulation period to the date of the snow pit survey. The average site $\rho_{new\,snow}$ was then determined as:
>
> $$\rho_{new\,snow} = \frac{\sum_{j=1}^{k} SWE_{snow-pit-j}}{\sum_{j=1}^{k}(\Delta h_{snow-pit-j})} \qquad (3)$$
>
> where *j* identifies a given snow pit and the corresponding total new snow and the sum extends over all *k* available snow pits. Instead of a mere average of $\rho_{new\,snow}$ values obtained from individual snow pit surveys, this relation gives more weight to snow pits with a higher $SWE_{snow\text{-}pit}$ amount.
>
> The *SWE* of each day (*t*) was then estimated by:
>
> $$SWE_{SR-t} = \begin{cases} \Delta h_t \dfrac{\rho_{new\,snow}}{\rho_{water}} & if\ \Delta h_t \geq 1\ cm \\ 0 & if\ \Delta h_t < 1\ cm \end{cases} \qquad (4)$$

We have also applied the leave-one-out cross-validation technique (LOOCV, a particular case of leave-p-out cross-validation with p = 1) to ensure independence between the data we use to estimate $\rho_{new\,snow}$ and the data we use to assess the corresponding estimation error. Specifically, we applied Eq. (3) (see answer above) once for

each snow pit (j), using all other snow pits in the calculation (*LOOCV $\rho_{new\ snow}$*) and using the selected snow pit as a single-item test ($\rho_{new\ snow}$ from snow pit j). In this way, we avoid dependence between the calibration and validation datasets in assessing the new snow density.

The results give evidence that the standard deviation of the differences between the *LOOCV $\rho_{new\ snow}$* values and the corresponding single-item test values ($\rho_{new\ snow}$ from snow pit j) is 18 kg m^{-3}. The error of the average value of $\rho_{new\ snow}$ can therefore be estimated dividing this standard deviation for the square root of the number of the considered snow pits. It turns out to be 6 kg m^{-3}. The new and the old estimates of $\rho_{new\ snow}$ (149 and 140 kg m^{-3}, respectively) therefore do not have a statistically significant difference. The individual snow accumulation periods instead have a naturally higher error and the single snow pit estimates for $\rho_{new\ snow}$ range from 128 to 178 kg m^{-3}. In addition, we attempted to extend this analysis considering each single snow layer (h_{ij}) instead of each snow pit j. In particular, we tried to associate to each snow pit layer the corresponding new snow measured by the sonic ranger (Citterio et al., 2007). However, this approach turned out to be too subjective to contribute more quantitative information about the real representativeness of the ρnew snow value we found.

All these analyses are shown in the "Discussion" section of the new version of the manuscript.

Finally, before uploading the reviewed manuscript, as suggested by the second Referee, the standard of English spelling and grammar has been improved by a professional, mother-tongue consultant.

Point-by-point answers to your comments follow:

Discuss the potential errors in individual events and how this impacts your peak estimation. A large event (big increase in snow depth) that has a large deviation from the mean density will result in larger errors (e.g. a heavy wet snowfall). What is the potential for this to occur at this site? Add a discussion about missing data as this is the greatest threat to failure of the technique. Can you do gap filling with photographed snow stakes? Would you recommend redundant sensors? More specific comments are listed below.

We have addressed these issues in the "Discussion" section of the new version of the manuscript.
As far as large events are concerned, they are rather rare at the studied site: only 3 days in the 11-year period covered by the data recorded more than 40 cm of new snow (the number of days decreases to 1 if the threshold increases to 50 cm). More in detail, we have the following distribution of new snow: 382 days with values between 1 and 10 cm, 95 days with values between 10 and 20 cm, 33 days with values between 20 and 30 cm, 11 days with values between 30 and 40 cm. Beside investigating the distribution of new snow values, we have checked also if the days in the different new snow intervals have significantly different average temperatures. We have not found any signal.
We agree that missing data is a relevant issue. The introduction of the second sonic ranger (Sommer USH8) at the end of the snow season 2013-2014 was an attempt to reduce the impact of this problem. The second sonic ranger, however, was still in process of testing in the last years of the period investigated within this paper (we e.g. changed the sensor model). We are confident that in the years to come it can help reducing the problem of missing data. Multiple sensors for fail-safe data collection are indeed highly recommended.
The four stakes installed at the corners of the snow pillow at the beginning of the 2014-2015 snow season were another idea for collecting more data. Unfortunately, they were broken almost immediately after the beginning of the snow accumulation period. They can be another way to deal with the problem of missing data, provided we figure out how to avoid breakage during the winter season
Finally, we have attempted to extend the analysis considering each single snow layer (h_{ij}) instead of each snow pit j. In particular, we have tried to associate to each snow pit layer the corresponding new snow measured by

the sonic ranger (Citterio et al., 2007). However, this approach turned out to be too subjective to contribute more quantitative information about the real representativeness of the ρnew snow value we found.

Note that the units for SWE should be reported in mm water equivalent (w.e.) or kg m-2 and not m w.e. . Snow depths should also be reported in cm and not m. It would also be useful if you used an abbreviation for "sonic ranger-derived SWE" such as SWE_{SR} and use this throughout the paper.

We have modified throughout the manuscript accordingly.

Abstract

Page 1, Line 15: "…on the Forni Glacier in Italy."

We have modified the sentence accordingly.

Lines 19-20: This sentence misses the mark. From what I have read, you are not really assessing the mean value of new snow density…this was done elsewhere. I think you miss explicitly stating the aim of the analysis and the value of this paper. You should state here that you are using mean new snowfall density and automated depth measurements to estimate the SWE of new snowfall and accumulating to estimate peak SWE and evaluating against other methods.

We have modified the sentence accordingly: "The aim of the analyses is to estimate the SWE of new snowfall and the annual peak of SWE based on the average density of the new snow at the site (corresponding to the snowfall during the standard observation period of 24 hours) and automated depth measurements, as well as to find the most appropriate method for evaluating SWE at this measuring site."

Line 21: "rather good" is a vague and subjective description of the estimation. Avoid this and/or quantify the estimation.

We have calculated the RMSE between SWE derived from SR50 sonic ranger and measured by means of snow pillow and we have found a RMSE of 45 mm w.e. We have added this information in the abstract accordingly.

Introduction

Page 2, Line 37: "…often only snowfall measurements are available…"

We have modified the sentence accordingly.

Line 38: "assess" should be "calculate" and "depending" should be "depends". Fresh snow density also depends on surface conditions, correct?

We have modified the sentences accordingly. We have added "surface conditions" in addition to "atmospheric conditions".

Line 50: CryoNet is more of a network attached to the GCW initiative, rather than a "project"

We have modified the sentence accordingly.

Line 71: "detail" not "details"

We have modified the sentence from "the accuracy should be verified in details for a large variety of events" to "the accuracy should be carefully verified for a large variety of events".

Page 3, Lines 89-93: For item ii, I'm not sure that you are "defining the reliability of…" because you don't really have a solid reference (more on that later) to be able to do that. I would rather you said that you were "assessing the capability of…". You also use the term "obtained SWE" here…now I assume that you are referring to the derived SWE from the depth measurements so you need to be more specific here.

We have modified the sentence accordingly: "ii) assess the capability to obtain SWE values from the depth measurements".

Data and Methods

Page 3, Line 100: You refer to "These sensors" but you should rather say "These measurements were made at the two stations…" since not all of the measurements were made with sensors.

We have modified the sentence accordingly.

Line 106: Was the T/RH sensor shielded? I assume yes.

The thermo-hygrometer is with shield. We have modified the sentence accordingly adding "shielded".

Page 4, Line 110: You should cite Beaumont (1965) here when introducing the snow pillow methods. Also some grammar issues with the first sentence on this page.

We have modified the sentence accordingly citing Beaumont (1965) and checking the grammar issues. The new version is: "AWS Forni SPICE is equipped with a snow pillow (Park Mechanical steel snow pillow, 150 x 120 x 1.5 cm) and a barometer (STS ATM.1ST) for measuring the snow water equivalent (Table 1, Beaumont, 1965). "

Line 116: I think "constrictions" should be "challenges".

We have modified the sentence accordingly.

Line 121: "represented" should be "provided" or "supplied". "deep" should be "thorough".

We have modified the sentences accordingly.

Line 122: "working" should perhaps be "operation".

We have modified the sentence accordingly.

Line 123: "…due to ice flow, etc.) is required before installation…"

We have modified the sentence accordingly.

Lines 125-128: Not sure if this paragraph adds anything to the paper

We have deleted this paragraph accordingly.

Line 130: I think that I know what you mean by "adjust" but you should clarify this.

We have modified the sentence accordingly in order to better explain what we mean. The new version is : "In this way, the AWSs stand freely on the ice, and move together with the melting surface during summer (with a mean ice thickness variation of about 4 m per year)."

Line 132: What are ring faults?

The ring faults are a series of circular or semicircular fractures with stepwise subsidence, caused by englacial or subglacial meltwater creating voids at the ice-bedrock interface and eventually the collapse of cavity roofs.
We have modified the paragraph accordingly. The new version is: "Due to the formation of ring faults, in November 2015 both AWSs were moved to the Forni glacier central tongue (46°23'42.40"N and 10°35'24.20"E at an elevation of 2675 m a.s.l., the red star in Fig. 1a). Ring faults are a series of circular or semicircular fractures with stepwise subsidence (caused by englacial or subglacial meltwater) that could compromise the stability of the stations because they could create voids at the ice-bedrock interface and eventually the collapse of cavity roofs (Azzoni et al., 2017; Fugazza et al., submitted)."

Line 146-147: "…from snow depth acquired by sonic rangers and estimated new snowfall density." Clarify what you mean by the last sentence "In particular…". This is a crucial piece of the methodology and you should describe this better. What time did you use for start/end? Did you do any noise filtering? How do you account for redistribution, settling, and sublimation during the day? (should also be discussed in the Discussion section).

In order to estimate the "daily positive differences in depth (Δh)", we considered the hourly snow depth values recorded by the sonic ranger in a day and we calculated the difference between the last and the first reading. Whenever this difference is positive (at least 1 cm), it corresponds to a new snowfall:

$$\Delta h = \begin{cases} reading\ at\ h23 - reading\ at\ h00\ (if\ positive\ and\ > 1\ cm) \\ 0\ (if\ negative\ or\ \leq 1\ cm) \end{cases}$$

This issue is better explained in the revised version of the manuscript (see answers to the main comments).

As far as settling is concerned, $\Delta h_{snow-pit-j}$ from Eq. 2 (see answers to the main comments) would indeed be higher if Δh_{tj} values were calculated considering an interval shorter than 24 hours. However, this would not be possible because on the one hand, the sonic ranger data's margin of error is too high to consider hourly resolution, and on the other hand, new snow is defined by the WMO within the context of a 24-hour period. Therefore, settling could not be considered in our analyses since new snow as defined by the WMO already includes the settling that occurs in the 24-hour period.

Obviously, our sonic ranger data may be affected by snow transported by wind. The effect that is potentially more relevant is new snow that is recorded by the sonic ranger but then blown away in the following days. It is therefore considered in $\Delta h_{snow-pit-j}$ but not in $SWE_{snow-pit-j}$, thus causing an underestimation of $\rho_{new\ snow}$ (see Eq. 3 – answers to the main comments). The snow transported to the measuring site can also influence $\rho_{new\ snow}$ even if in this case the effect is less important as it measured both by the sonic ranger and by the snow pit. Here, the problem may be an overestimation of $\rho_{new\ snow}$ as snow transported by wind usually has a higher density than new snow. We considered the problem of the effect of wind on snow cover when we selected the

station site on the glacier. Even though sites not affected by wind transport simply do not exist, we are confident that the site we selected has a position that can reasonably minimize this issue.

Moreover, sublimation processes would have an effect that is similar to those produced by new snow that is recorded by the sonic ranger but then blown away in the following days. In any case, the value we found for the site average new snow density (i.e. 149 kg m^{-3}) does not seem to suggest an underestimated value.

All these eventual effects are described in the "Discussion" section.

As regards the quality-check activities, we submitted the dataset to a strict control to avoid under- or over-measurements, to remove outliers and nonsense values, and to filter possible noises. We will modify the relative paragraph in the "Method" section accordingly in order to better explain our approach (see answer to the main comments).

Page 5, Lines 152-153: What do you mean by "unique date"?

We have modified the sentence accordingly in order to better clarify the approach (see answer to the main comments).

Lines 150-153: This is the methodology that isn't clearly described (see note above).

In the new version of the manuscript, we have modified this part of the "Method" section (following also the suggestions of the second Referee) in order to better clarify our procedure in estimating the density of the new snow ($\rho_{new\,snow}$) used for deriving SWE from snow depth data (see answers to the main comments).

Results

Beside expressing the errors or biases in absolute units, it would be helpful if the relative bias was stated. Same can be said for subsequent sections.

We have modified all the showed values adding the error accordingly.

Lines 159-161: Can you comment here or in the discussion on any potential impacts on the analysis due to the site move.

The relocation of the AWSs could influence snow conditions, as there could be a different snow accumulation due to a different radiation input and diverse wind regimes. However, as the distance between the two sites is about 500 m, the difference in elevation is only 44 m and the aspect is very similar, we do not expect a noticeable impact of the site change on snow depth.

Lines 166-169: The intercomparison between the SR50 and the USH8 is interesting but largely irrelevant for this paper. I would omit this. What you could mention is that the two instruments had a correlation of ?? and that from 2015 onwards, redundancy in the snow depth measurement could mean better data for the SWE estimate.

We have omitted the intercomparison and the relative figure accordingly. We have moved this part into the "Method" section, specifying the correlation between the datasets from Campbell and Sommer sensors.


Line 174: What is "a very good agreement"?

We have added the root mean square error, that is equal to 58 mm w.e.

Line 180: "elaborate" should be "accumulate"

We have modified the sentence accordingly.

Figure 4: it would be very useful if the missing SR50 data could be indicated (e.g. different colour line, etc). This would certainly help with the interpretation of the graph. One might argue that all of the estimated SWE should be set to missing after large gaps in the snow depth data since it is an accumulated result.

We have modified the figure showing the period without snow depth data. The new version of the figure is:
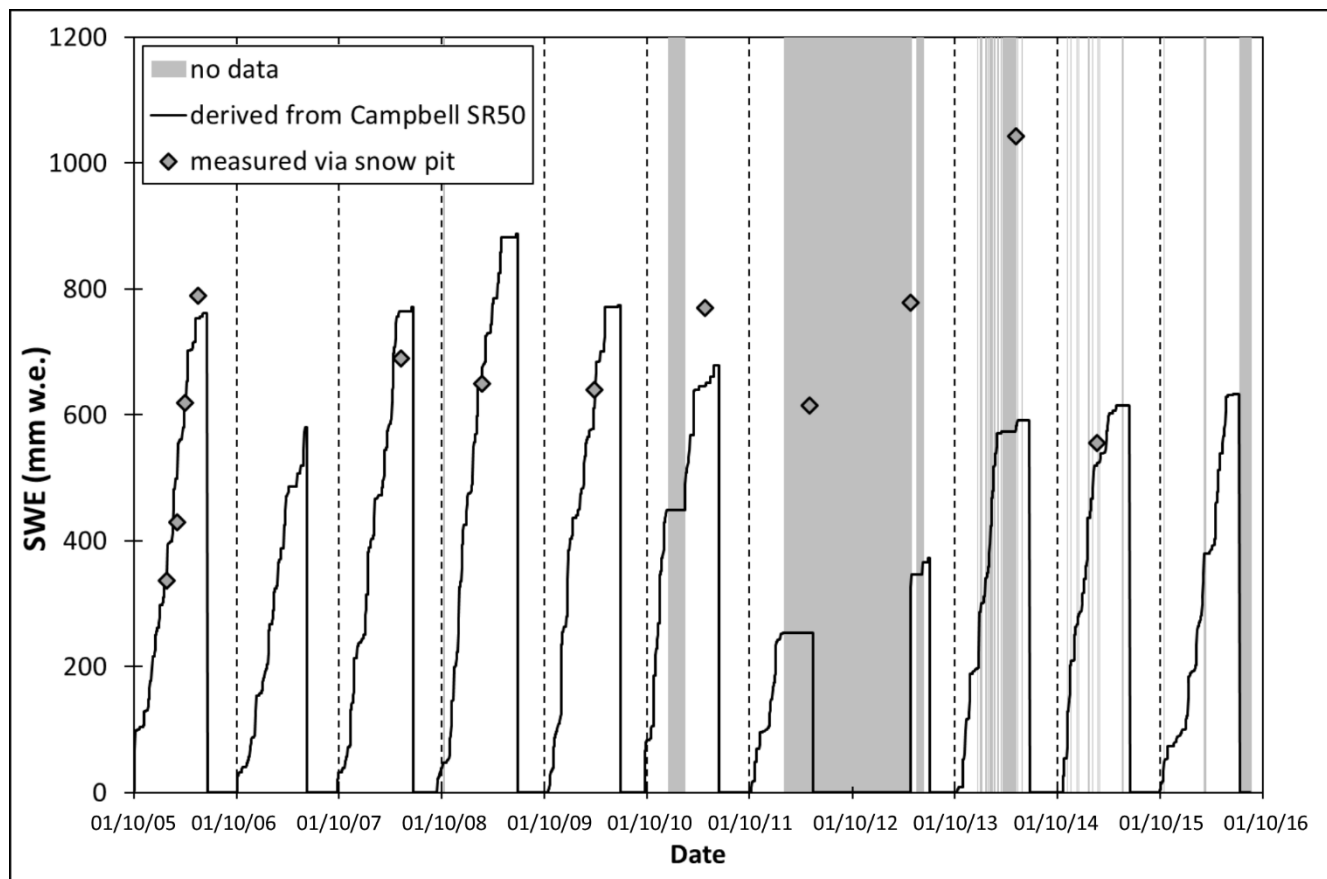
Page 6, Line 188: "…thus suggesting a correct working of the sensor." This is more than a sensor, but rather a technique or process. This sentence is awkward and should be reworded.

We have modified the sentence accordingly. The new version is: "thus suggesting that in spite of the problems at the beginning of the snow season, the snow pillow seems to give reasonable results"

Line 192: typo "derided".

We have modified in "SWE_{SR}" according to your previous comment.

Line 194: Change "raises" to "increases".

We have modified the sentence. The new version is: "There is a general underestimation of SWE_{SR} compared to the snow pillow values, considering the 2014-2015 data, though the agreement strengthens in the 2015-2016 dataset (Fig. 5): 54 mm w.e. and 29 mm w.e. of RMSE regarding 2014-2015 and 2015-2016, respectively."

Lines 196-197: You mention the snow course data using the snow tube and suggest a "large spatial variability". Perhaps you could report the average and standard deviation of this data and in the discussion, relate this variability to the differences you see between the SR50 SWE estimation and the snow pillow and snow pit.

We have modified accordingly adding the mean value of 165 cm and the standard deviation of 29 cm. In addition, as reported in the following comment, we have discussed this variability in the "Discussion" section.

Lines 200-201: Oversampling of the tube is a potential error and should be included in the discussion and not in the results. You need to estimate this error and put it into context with the differences between the methods.

As stated in Johnson et al. (2015), numerous studies have been conducted to verify snow tube accuracy in determining SWE (Freeman, 1965; Work et al., 1965; Beaumont, 1967; Peterson and Brown, 1975; Goodison, 1978; Farnes et al., 1982), the most recent being by Sturm et al. (2010) and Dixon and Boon (2012). The most recent comparison of snow tubes by Dixon and Boon (2012) examined the Standard Federal, Meteorological Service of Canada, and SnowHydro snow tubes. The earlier studies had found that snow tubes tended to over-measure SWE from about 4–11%, whereas the recent studies by Sturm et al. (2010) and Dixon and Boon (2012) found that snow tubes could under-measure or over-measure SWE from -9% to 11%. Even if we allow for ±10% margin of error in our snow tube measurements, the high SWE variability is confirmed.

This part is discussed in the "Discussion" section of the revised manuscript.

Discussion

The start of the Discussion paragraph needs a couple of sentences of introduction to "frame" the problem and set up the following paragraphs.

We have added a paragraph accordingly in order to better introduce the "Discussion" section: "Defining a correct algorithm for modeling SWE data is very important for evaluating the water resources deriving from snow melt. The approach applied to derive SWE_{SR} is highly sensitive to the value used for the new snow density, which can vary substantially depending on both atmospheric and surface conditions. In this way, the error in individual snowfall events could be quite large. Moreover, the technique depends on determining snowfall events, which are estimated from changes in snow depth, and the subsequent calculation and accumulation of SWE_{SR} from those events. Therefore, missed events due to gaps in snow depth data could invalidate the calculation of peak SWE_{SR}. For these reasons, we focused our analyses on understanding how an incorrect assessment of ρnew snow or a gap in snow depth data may affect the estimation of the SWE."

Lines 205-208: The whole success of technique hinges on the approximation of the average new snowfall density but yet this only gets a few sentences in the discussion. This needs to be expanded. When would you expect errors to be greatest? At this site, can you potentially get large snowfalls that have much greater (or lower) densities

than the average, which could then bias your accumulated SWE? Any suggestions on how to better estimate new snowfall SWE, perhaps at sites where you don't have snow pits to back-calculate the density.

We have modified this paragraph accordingly. The new version is:

"First, we evaluated the $\rho_{new\,snow}$ estimate (applying Eq. 3, found to be equal to 149 kg m^{-3} considering the 2005-2015 dataset), by means of the leave-one-out cross-validation technique (LOOCV, a particular case of leave-p-out cross-validation with p = 1), to ensure independence between the data we use to estimate $\rho_{new\,snow}$ and the data we use to assess the corresponding estimation error. In this kind of cross-validation, the number of "folds" (repetitions of the cross-validation process) equals the number of observations in the dataset. Specifically, we applied Eq. 3 once for each snow pit (j), using all other snow pits in the calculation (*LOOCV $\rho_{new\,snow}$*) and using the selected snow pit as a single-item test ($\rho_{new\,snow}$ from snow pit j). In this way, we avoid dependence between the calibration and validation datasets in assessing the new snow density. The results are shown in Table 2. They give evidence that the standard deviation of the differences between the *LOOCV $\rho_{new\,snow}$* values and the corresponding single-item test values ($\rho_{new\,snow}$ from snow pit j) is 18 kg m^{-3}. The error of the average value of $\rho_{new\,snow}$ can therefore be estimated dividing this standard deviation for the square root of the number of the considered snow pits. It turns out to be 6 kg m^{-3}. The new and the old estimates (149 and 140 kg m^{-3}, respectively) therefore do not have a statistically significant difference. The individual snow accumulation periods instead have a naturally higher error and the single snow pit estimates for $\rho_{new\,snow}$ range from 128 to 178 kg m^{-3}. In addition, we attempted to extend this analysis considering each single snow layer (h_{ij}) instead of each snow pit j. In particular, we tried to associate to each snow pit layer the corresponding new snow measured by the sonic ranger (Citterio et al., 2007). However, this approach turned out to be too subjective to contribute more quantitative information about the real representativeness of the $\rho_{new\,snow}$ value we found.

Moreover, we investigated the *SWE* sensitivity to changes in $\rho_{new\,snow}$. In particular, we calculated *SWE$_{SR}$* using different values of new snow density ranging from 100 to 200 kg m^{-3} at 25 kg m^{-3} intervals (Fig. 6). An increase/decrease of the density by 25 kg m^{-3} causes a mean variation in *SWE$_{SR}$* of ±106 mm w.e. for each hydrological year (corresponding to about 17% of the mean total cumulative *SWE* considering all hydrological years), ranging from ±43 mm w.e. to ±144 mm w.e. A reliable estimation of $\rho_{new\,snow}$ is therefore a key issue.

In addition to an accurate definition of new snow density, an uninterrupted dataset of snow depth is also necessary in order to derive correct *SWE$_{SR}$* values. It is therefore necessary to put in place all the available information to reduce the occurrence of data gaps to a minimum. It is also important to stress that potential errors in individual snowfall events could affect peak *SWE$_{SR}$* estimation. A large snowfall event with a considerable deviation from the mean new snow density will result in large errors (e.g. a heavy wet snowfall). These events are rather rare at the Forni site: only 3 days in the 11-year period covered by the data recorded more than 40 cm of new snow (the number of days decreases to 1 if the threshold increases to 50 cm). More in detail, we found the following distribution of new snow: 382 days with values between 1 and 10 cm, 95 days with values between 10 and 20 cm, 33 days with values between 20 and 30 cm, 11 days with values between 30 and 40 cm. Beside investigating the distribution of new snow values, we also checked if the days in the different new snow intervals have significantly different average temperatures. The calculated average temperature values are -5.7 ± 4.5°C, -5.2 ± 4.2°C, and -4.8 ± 3.2°C (for days with 1-10 cm, 10-20 cm, and >20 cm of new snow depth, respectively), suggesting that there is no significant change of air temperature in these three classes. As far as data gaps are concerned, the introduction of the second sonic ranger (Sommer USH8) at the end of the 2013-2014 snow season was an attempt to reduce the impact of this problem. The second sonic ranger, however, was still in process of testing in the last years of the period investigated within this paper. We are confident that in the years to come it can help reduce the problem of missing data. Multiple sensors for fail-safe data collection are indeed highly recommended. In addition, the four stakes installed at the corners of the snow pillow at the beginning of the 2014-2015 snow season were another idea for collecting more data. Unfortunately, they were broken almost immediately after the beginning of the snow accumulation period. They can be another way to deal with the problem of missing data, provided we figure out how to avoid breakage during the winter season."

Line 205: "Once our procedure was verified, we performed…"

We have modified this paragraph accordingly to the previous comment.

Line 209: What is a "general good agreement". Quantify this.

We have modified the sentence accordingly from "Beside to a general good agreement between the measures performed with the different sensors, there are also some problems." to "As regards the instrumentation, we found some issues related to the derived snow information."

Line 214: "not constant" should be "inconsistent"

We have modified the sentence accordingly.

Page 7, Lines 223-231: You reference some issues with the snow pillow and then state that your snow pillow only seems to be working with depths greater that 50cm. Can you relate any of the referenced potential issues to the errors that you are seeing at the site? I'm not convinced that your issues are related to those referenced so this discussion needs to be stronger.

We have modified this part accordingly. The new version is: "The results from the snow pillow are difficult to explain as this sensor has been working for only two winter seasons and we are still in the process of testing. Analyzing data from the years to come will allow a more robust interpretation. However, we have searched for a possible explanation of this problem and this error could be due to the configuration of the snow pillow."

Lines 232-246: This paragraph is more of a justification for using this technique and is more appropriate for the introduction than the discussion. The sentences (Lines 246-250) about the snow pillow is relevant, however. The subsequent discussion about the snow pit should have its own paragraph.

We have modified this paragraph accordingly, moving part of it into the "Introduction" section. In addition, we have discussed about the snow pit measurements in a separate paragraph.

Line 248-249: You should refer to Fig 5 and discuss this error in context with what you are seeing at the site. This should be combined with the snow pillow discussion. Also, is there a reference for this statement about minimum snow pillow measurements?

We have modified this section as described in the previous comments.

Line 252: What do you mean by "whole glacier accumulation amount"?

We have modified the sentence accordingly from "whole glacier accumulation amount" to "total snow accumulation amount".

Line 253: "…the peak cumulative SWE…" Same applies for the next sentence.

We have modified the sentences accordingly.

Line 257: Typo "closed" should be "close"

We have modified the sentence accordingly.

Lines 259-261: The two sentences beginning with "Finally…" seem out of place or awkward. Please re-organize these.

We have modified the sentences accordingly from "Finally, with data acquired by the SR50 sonic ranger a correct curve of SWE was derived. The unique issue is represented by the definition of the beginning of the accumulation period, but this can be overcome using albedo data." to "Finally, the SR50 sonic ranger features the unique problem of the definition of the start of the accumulation period, but this can be overcome using albedo data."

Page 8, Lines 261-263: The discussion about the SR50/USH8 is largely unsubstantiated and should be omitted.

We have deleted the discussion regarding the USH8 sonic ranger accordingly.

I would like to see me recommendations in the discussion about how to make this technique better and how to adapt it for other remote sites. Perhaps make some recommendations on further testing, possibly to take advantage of other data collected during SPICE.

We have modified the "Discussion" section accordingly, following all your previous suggestions and comments.

Conclusions

Line 267: Was the acronym SPICE not defined earlier in the paper? If it was, it doesn't need to be defined here.

We have deleted the definition of the acronym accordingly.

Line 269: The sentence beginning with "This has allowed…" needs some commas to read correctly.

We have modified the sentence accordingly. The new version is: "This has allowed an accurate comparison and evaluation of the pros and cons of using the snow pillow, sonic ranger, snow pit, or snow weighing tube, and of estimating SWE from snow depth data."

Line 270: "…SWE values using a fresh snow density…"

We have modified the sentence accordingly.

Lines 270-273: I have some problems with the statement "The results achieved…". You only tested two sensors for measuring snow depth so I wouldn't go as far as saying that the SR50 is the "most suitable device"…perhaps it is but you don't have enough information to substantiate this. You present a technique for estimating peak SWE on a glacier and you compare this technique against another technique (e.g. the snow pillow) but you don't really have a true reference so I'm not sure you can say that it is "most suitable". Rather, I would reference your relative error and despite the issues, suggest that the technique is "suitable" for estimating SWE.

We have modified the sentences accordingly from "The results achieved during the SPICE experiment support our procedure for deriving SWE values and the applied fresh snow density of 140 kg m^{-3} (Senese et al., 2014),

and suggest that, once ρ_{fresh} snow is known, the SR50 sonic ranger can be considered the most suitable device on a glacier to record snowfall events and to measure snow depth values in order to derive the point SWE."

To "We found that the mean new snow density changes based on the considered period was: 140 kg m^{-3} in 2005-2009 (Senese et al., 2014) and 149 kg m^{-3} in 2005-2015. The difference is however not statistically significant. We first evaluated the new snow density estimation by means of LOOCV and we found an error of 6 kg m^{-3}. Then, we benchmarked the derived SWE_{SR} data against the information from the snow pillow (data which was not used as input in our density estimation), finding a RMSE of 45 mm w.e. These analyses permitted a correct definition of the reliability of our method in deriving SWE from snow depth data. Moreover, in order to define the effects and impacts of an incorrect $\rho_{new\,snow}$ value in the derived SWE amount, we found that a change in density of ±25 kg m^{-3} causes a mean variation of 17% of the mean total cumulative SWE considering all hydrological years. Finally, once $\rho_{new\,snow}$ is known, the sonic ranger can be considered a suitable device on a glacier, or in a remote area in general, for recording snowfall events and for measuring snow depth values in order to derive SWE values. In fact, the methodology we presented can be interesting for other sites as it allows estimating total SWE using a relatively inexpensive, low power, low maintenance, and reliable instrument such as the sonic ranger, and it is a good solution for estimating SWE at remote locations such as glacier or high alpine regions."

Line 274: What are the relative variations as a result of the density change?

We have modified the sentence accordingly adding the relative variation: "Moreover, in order to define the effects and impacts of an incorrect $\rho_{new\,snow}$ value in the derived SWE amount, we found that a change in density of ±25 kg m^{-3} causes a mean variation of 17% of the mean total cumulative SWE considering all hydrological years."

Instead of ending the conclusions with a comment about the error, perhaps end with a comment on the applicability to other remote sites.

We have added accordingly: "The sensors generally used (e.g. heated tipping bucket rain gauges, heated weighing gauges, or disdrometers) can provide more accurate measurements compared to the ones installed at the Forni Glacier. The problem is that in remote areas like a glacier at a high alpine site, it is very difficult to install and maintain them. The main constrictions concern i) the power supply to the instruments, which consists in solar panels and lead-gel batteries, and ii) the glacier dynamics, snow flux and differential snow/ice ablation that can compromise the stability of the instrument structure. Therefore, for our limited experience in such remote areas, a sonic ranger could represent a useful approach for estimating SWE, since it does not require expert personnel, nor does it depend on the date of the survey (as do such manual techniques as snow pits and snow weighing tubes); it is not subject to glacier dynamics, snow flux or differential ablation (as do graduated rods installed close to an automated camera and snow pillows), and the required power is not so high (unlike heated tipping bucket rain gauges). The average new snow density must, however, be known either by means of snow pit measurements or by the availability of information from similar sites in the same geographic area."

We very much appreciate the time and effort you put into the comments.

Sincerely,

Antonella Senese and Co-author

Referee #2

Dear Referee

Thanks for helping us in improving the manuscript.

We have considered all your comments in the revised version of the manuscript. The title, as you suggest and in agreement also with the second Referee, is: Estimating the snow water equivalent on remote and glacierized high elevation areas (Forni Glacier, Italy).

We have therefore modified the "Method" section in order to better explain the approach applied for estimating the site-average-new-snow density ($\rho_{new\,snow}$) and the sonic-ranger-depth-derived *SWE*. In the previous version of the manuscript, we performed the analyses using the mean new snow density (140 kg m^{-3}) that was obtained by Senese et al. (2012) considering the 2005-2009 dataset, and then we discussed to what degree this value is able to describe the data from the years to come. In the revised version of the manuscript, we have started with updating the site average new snow density estimation exploiting all the available datasets and we have performed all the subsequent analyses using this value.

The text concerning this issue in "Method" section is:

> In addition to the measurements recorded by the AWSs, since winter 2005-2006, personnel from the Centro Nivo-Meteorologico (namely CNM Bormio-ARPA Lombardia) of the Lombardy Regional Agency for the Environment have periodically used snow pits (performed according to the AINEVA protocol, see also Senese et al., 2014) in order to estimate snow depth and *SWE*. In particular, for each snow pit *j*, the thickness (h_{ij}) and the density (ρ_{ij}) of each snow layer (*i*) are measured for determining its snow water equivalent, and then the total $SWE_{snow-pit-j}$ of the whole snow cover (*n* layers) is obtained:
>
> $$SWE_{snow-pit-j} = \sum_{i=1}^{n} h_{ij} \cdot \frac{\rho_{ij}}{\rho_{water}} \tag{1}$$
>
> where ρ_{water} is water density. As noted in a previous study (Senese et al., 2014), the date when the snow pit is dug is very important for not underestimating the actual accumulation. For this reason, we considered only the snow pits excavated before the beginning of snow ablation. In fact, whenever ablation occurs, successive *SWE* values derived from snow pits show a decreasing trend (i.e., they are affected by mass losses).
>
> The snow pit *SWE* data were then used, together with the corresponding total new snow derived from sonic ranger readings, to estimate the site average $\rho_{new\,snow}$, in order to update the value of 140 kg m^{-3} that was found in a previous study of data of the same site covering the 2005-2009 period (Senese et al., 2012a). We need to update our figures for $\rho_{new\,snow}$ as it is the key variable for estimating *SWE* from the sonic ranger's new snow data. Specifically, for each snow pit *j*, the corresponding total new snow was first determined by:
>
> $$\Delta h_{snow-pit-j} = \sum_{t=1}^{m} (\Delta h_{tj}) \tag{2}$$
>
> where *m* is the total number of days with snowfall in the period corresponding to snow pit *j* and Δh_{tj} corresponds to the depth of new snow on day *t*. In particular, we considered the hourly snow depth values recorded by the sonic ranger in a day and we calculated the difference between the last and the first reading. Whenever this difference is positive (at least 1 cm), it corresponds to a new snowfall. All data are subject to a strict control to avoid under- or over-measurements, to remove outliers and nonsense values, and to filter possible noises. $\sum_{t=1}^{m} (\Delta h_{tj})$ is therefore the total new snow measured by the Campbell SR50 from the beginning of the accumulation period to the date of the snow pit survey. The average site $\rho_{new\,snow}$ was then determined as:
>
> $$\rho_{new\,snow} = \frac{\sum_{j=1}^{k} SWE_{snow-pit-j}}{\sum_{j=1}^{k} (\Delta h_{snow-pit-j})} \tag{3}$$
>
> where *j* identifies a given snow pit and the corresponding total new snow and the sum extends over all *k* available snow pits. Instead of a mere average of $\rho_{new\,snow}$ values obtained from individual snow pit surveys, this relation gives more weight to snow pits with a higher $SWE_{snow-pit}$ amount.
>
> The *SWE* of each day (*t*) was then estimated by:
>
> $$SWE_{SR-t} = \begin{cases} \Delta h_t \frac{\rho_{new\,snow}}{\rho_{water}} & if\ \Delta h_t \geq 1\ cm \\ 0 & if\ \Delta h_t < 1\ cm \end{cases} \tag{4}$$

We also applied the leave-one-out cross-validation technique (LOOCV, a particular case of leave-p-out cross-validation with p = 1) to ensure independence between the data we use to estimate $\rho_{new\,snow}$ and the data we use to assess the corresponding estimation error. Specifically, we applied Eq. (3) (see answer above) once for each

snow pit (j), using all other snow pits in the calculation (*LOOCV $\rho_{new\,snow}$*) and using the selected snow pit as a single-item test ($\rho_{new\,snow}$ from snow pit j). In this way, we avoid dependence between the calibration and validation datasets in assessing the new snow density.

The results give evidence that the standard deviation of the differences between the *LOOCV $\rho_{new\,snow}$* values and the corresponding single-item test values ($\rho_{new\,snow}$ from snow pit j) is 18 kg m^{-3}. The error of the average value of $\rho_{new\,snow}$ can therefore be estimated dividing this standard deviation for the square root of the number of the considered snow pits. It turns out to be 6 kg m^{-3}. The new and the old estimates of $\rho_{new\,snow}$ (149 and 140 kg m^{-3}, respectively) therefore do not have a statistically significant difference. The individual snow accumulation periods instead have a naturally higher error and the single snow pit estimates for ρnew snow range from 128 to 178 kg m^{-3}. In addition, we attempted to extend this analysis considering each single snow layer (h_{ij}) instead of each snow pit j. In particular, we tried to associate to each snow pit layer the corresponding new snow measured by the sonic ranger (Citterio et al., 2007). However, this approach turned out to be too subjective to contribute more quantitative information about the real representativeness of the ρnew snow value we found.

All these analyses are shown in the "Discussion" section of the new version of the manuscript.

Finally, before uploading the reviewed manuscript, as suggested by the second Referee, the standard of English spelling and grammar has been improved by a professional, mother-tongue consultant.

Point-by-point answers to your comments follow.

- the exact determination of the fixed density

In the new version of the manuscript, we have modified this part of the "Method" section (following also the suggestions of the first Referee) in order to better clarify our procedure in estimating the density of the average new snow ($\rho_{new\,snow}$) used for deriving SWE from snow depth data (see answers to the main comments).

- the calculation of the daily positive snow depth differences

In order to estimate the "daily positive differences in depth (Δh)", we considered the hourly snow depth values recorded by the sonic ranger in a day and we calculated the difference between the last and the first reading. Whenever this difference is positive (at least 1 cm), it corresponds to a new snowfall:

$$\Delta h = \begin{cases} reading\ at\ h23 - reading\ at\ h00\ (if\ positive\ and\ >1\ cm) \\ 0\ (if\ negative\ or\ \leq 1\ cm) \end{cases}$$

This issue is better explained in the revised version of the manuscript (see answers to the main comments).

- the limitations and uncertainty of the method

For validating our procedure in estimating the new snow density, we have applied the leave-one-out cross-validation technique (LOOCV, a particular case of leave-p-out cross-validation with p = 1) in order to assess both the error of the estimation of the average site $\rho_{new\,snow}$ and the error we perform if we estimate $\rho_{new\,snow}$ of each single snow pit by means of our approach (see answers to the main comments).

In addition, we have investigated the *SWE* sensitivity to changes in $\rho_{new\,snow}$: we have calculated *SWE$_{SR}$* using different values of new snow density ranging from 100 to 200 kg m^{-3} at 25 kg m^{-3} intervals.

We have also added to the "Discussion" section some new information on the occurrence of outliers in our data set (the highest new snow value is slightly higher than 40 cm) and on the distribution of new snow. We have then underlined that we do not have a significant link between new snow and corresponding daily average temperature that could e.g. indicate that higher new snow values are more frequent at the beginning and at the end of the snow season.

Moreover, we have investigated if potential errors in individual snowfall events could affect peak SWE_{SR} estimation. In fact, a large snowfall event with a considerable deviation from the mean new snow density will result in large errors (e.g. a heavy wet snowfall). These events are rather rare at the Forni site: only 3 days in the 11-year period covered by the data recorded more than 40 cm of new snow (the number of days decreases to 1 if the threshold increases to 50 cm). More in detail, we found the following distribution of new snow: 382 days with values between 1 and 10 cm, 95 days with values between 10 and 20 cm, 33 days with values between 20 and 30 cm, 11 days with values between 30 and 40 cm. Beside investigating the distribution of new snow values, we also checked if the days in the different new snow intervals have significantly different average temperatures. The calculated average temperature values are -5.7 ± 4.5°C, -5.2 ± 4.2°C, and -4.8 ± 3.2°C (for days with 1-10 cm, 10-20 cm, and >20 cm of new snow depth, respectively), suggesting that there is no significant change of air temperature in these three classes.

Another possible source of error in estimating new snow density and in deriving daily SWE is represented by rainfall events. In fact, one of the effects is an enhanced snow melt and then a decrease in snow depth, as rain water has a higher temperature than the snow. Therefore, especially at the beginning of the snow accumulation season, we could detect a snowfall (analyzing snow depth data) but, whenever it was followed by a rainfall, the new fallen snow could partially or completely melt, thus remaining undetected when measured at the end of the accumulation season using snow pit techniques. This is therefore another potential error that, besides the ones previously considered, could lead to underestimation of the $\rho_{new\ snow}$ value, even if, as already mentioned, the found value of 149 kg m^{-3} does not seem to suggest this. On the other hand, rain can also increase the SWE measured by the snow pit without giving a corresponding sign in the snow depth measured by the sonic ranger whenever limited amounts of rain fall over cold snow. Anyway, rain events are extremely rare during the snow accumulation period.

In addition to the validation of new snow density estimation, we have benchmarked the derived SWE data against the ones measured by the snow pillow (data not used as input in our density estimation). This validation permitted to correctly define the reliability of our method in deriving SWE from snow depth data.

All these data elaborations and analyses are included in the "Discussion" section.

- transferability to other sites and climates

In our opinion, the methodology we presented can be interesting for other sites as it allows estimating total SWE using a relatively inexpensive, low power, low maintenance, and reliable instrument such as the sonic ranger, and it is a good solution for estimating SWE at remote locations such as glacier or high alpine regions. We have highlighted this point in the "Conclusions" of the revised manuscript.

The fixed density of 140 kg/m3 has already been used in earlier papers (2012, 2104) from you. Indeed the comparison between the estimated SWE from snow depth measurements and SWE from snow pits (the current Fig. 4) has already been published in The Cryosphere in 2014 (Fig. 2 in https://www.the-cryosphere.net/8/1921/2014/tc-8-1921-2014.pdf). However, there the estimated SWE data seem to be different from the one shown in the current paper? Can you explain?

The values shown in the previous paper are the same but cumulating daily SWE data the yearly end date is different. In this study, we showed the SWE values only during the snow accumulation period, neglecting the ice ablation period. In fact, in this period the fallen snow is completely melted within a few days and always before the beginning of the accumulation period. Instead, in Senese et al. (2014) we cumulated daily SWE values until 30th September of each year. Only for this reason, the two figures seem different, but they show the same datasets.

However, in the revised manuscript, we have shown SWE data derived using the update new snow density (149 kg m[-3], instead of 140 kg m[-3])

- Your fixed density is not a fresh snow density, since you totally neglect settling and are not able to determine small snow falls due to the measurement uncertainty of the snow depth sensor. The found density of 140 kg/m3 can therefore not be compared with published fresh snow densities found in literature and is relatively large since it has to compensate the missing snowfall amounts mentioned above.

We have modified our terminology accordingly from "fresh snow density" to "new snow density". In addition, we agree that missing data is a relevant issue.
The introduction of the second sonic ranger (Sommer USH8) at the end of the 2013-2014 snow season was an attempt to reduce the impact of this problem. The second sonic ranger, however, was still in process of testing in the last years of the period investigated within this paper. We are confident that in the years to come it can help reduce the problem of missing data. Multiple sensors for fail-safe data collection are indeed highly recommended. In addition, the four stakes installed at the corners of the snow pillow at the beginning of the 2014-2015 snow season were another idea for collecting more data. Unfortunately, they were broken almost immediately after the beginning of the snow accumulation period. They can be another way to deal with the problem of missing data, provided we figure out how to avoid breakage during the winter season.

- The impact of rain events at the beginning and end of the snow season has not been discussed so far.

We have added some discussion about the impact of rain events: another possible source of error in estimating new snow density and in deriving the daily *SWE* is represented by rainfall events. In fact, one of the effects is an enhanced snow melt and then a decrease in snow depth, as rain water has a higher temperature than the snow. Therefore, especially at the beginning of the snow accumulation season, we could detect a snowfall (analyzing snow depth data) but, whenever it was followed by a rainfall, the new fallen snow could partially or completely melt, thus remaining undetected when measured at the end of the accumulation season using snow pit techniques. This is therefore another potential error that, besides the ones previously considered, could lead to underestimation of the $\rho_{new\,snow}$ value, even if, as already mentioned, the found value of 149 kg m[-3] does not seem to suggest this. On the other hand, rain can also increase the *SWE* measured by the snow pit without giving a corresponding sign in the snow depth measured by the sonic ranger whenever limited amounts of rain fall over cold snow. Anyway, rain events are extremely rare during the snow accumulation period."

- Since your focus is the determination SWE from snow depth your title needs to be changed to something like: Estimation of SWE from automatic snow depth measurements during accumulation on a high alpine glacier.

Following also the suggestions of the first Referee, we have modified the title accordingly from "Snow data intercomparison on remote and glacierized high elevation areas (Forni Glacier, Italy)" to "Estimating the snow water equivalent on remote and glacierized high elevation areas (Forni Glacier, Italy)".

- The content of the abstract is odd. The abstract needs to be rewritten.

Following also the suggestions of the first Referee, we have modified the abstract accordingly. The new version is:

"We present and compare 11 years of snow data (snow depth and snow water equivalent, *SWE*) measured by an Automatic Weather Station corroborated by data resulting from field campaigns on the Forni Glacier in Italy. The aim of the analyses is to estimate the *SWE* of new snowfall and the annual peak of *SWE* based on the average density of the new snow at the site (corresponding to the snowfall during the standard observation period of 24 hours) and automated depth measurements, as well as to find the most appropriate method for evaluating *SWE* at this measuring site. The results indicate that the daily SR50 sonic ranger measures allow a rather good estimation of the *SWE* (RMSE of 45 mm w.e. if compared with snow pillow measurements), and the available snow pit data can be used to define the mean new snow density value at the site. For the Forni Glacier measuring site, this value was found to be 149 ± 6 kg m^{-3}. The *SWE* derived from sonic ranger data is quite sensitive to this value: a change in new snow density of ±25 kg m^{-3} causes a mean variation in *SWE* of ±106 mm w.e. for each hydrological year (corresponding to about 17% of the mean total cumulative *SWE* considering all hydrological years), ranging from ±43 mm w.e. to ±144 mm w.e.."

- The possible impact of the dislocation of the station needs to be discussed.

The relocation of the AWSs could influence snow conditions, as there could be a different snow accumulation due to a different radiation input and diverse wind regimes. However, as the distance between the two sites is about 500 m, the difference in elevation is only 44 m and the aspect is very similar, we do not expect a noticeable impact of the site change on snow depth.

- You need to present some numbers about the uncertainty of the involved sensors, manual measurements and the uncertainty of the presented results. That means the papers definitely needs more quantitative information about the performance of your method.

Following also the suggestions of the first Referee, we have added and discussed the uncertainty of the method for deriving the new snow density and of all the measurements and techniques applied in this study for quantifying *SWE*:
- The average site new snow density is found to be affected by an error of ±6 kg m^{-3} (by means of the leave-one-out cross-validation).
- The comparison between SWE derived by snow depth data and SWE measured by means of the snow pillow showed a RMSE of 45 mm w.e.
- Snow tubes could under-measure or over-measure SWE of about ±10% (as found by Sturm et al., 2010, and Dixon and Boon, 2012).

- Information about the measurement frequency and aggregation of the data shown in the figures need to included.

Data points are sampled at 60-second intervals and averaged over a 60-minute time period for the SR50 sonic ranger, wind sensor and barometer, over a 30-minute time period for the sensors recording air temperature, relative humidity, solar and infrared radiation, and liquid precipitation, and over a 10-minute time period for the USH8 sonic ranger and snow pillow. All data are recorded in a flash memory card, including the basic distribution parameters (minimum, mean, maximum, and standard deviation values). We have added all these information in the revised manuscript accordingly.

In addition, we have added in all the figure captions the aggregation of data.

- A comparison with data from other sites is needed in order to be able to judge the usefulness of the method.

Comparing our results with snow depth data from 10 stations spread over the Italian Alps, the snow depth peaks are in agreement with findings over the Italian Alps. In fact, analyzing data in the period 1960–2009 from 10 stations above 1500 m a.s.l. Valt and Cianfarra (2010) reported a mean snow depth of 233 cm (from 199 to 280 cm). At the Forni Glacier, we observed a mean snow depth peak of 222 cm, ranging from 134 to 280 cm. We have added this comparison in the "Results" section.

- The English language of the paper is often odd and needs to be improved in revised version.

Before uploading the reviewed manuscript, as suggested, the standard of English spelling and grammar has been improved by a professional, mother-tongue consultant.

The following comments and corrections are the ones reported by the Referee in PDF of the paper.

Title: We have modified it accordingly.

Line 14, Abstract: We have deleted the word "snowfall" accordingly.

Lines 15-18, Abstract: We have deleted the sentences accordingly.

Line 19, Abstract: "fresh snow" Please use the terms defined in the international classification for seasonal snow on the ground.

We used the terminology reported in the 2008 WMO Guide, where the term "fresh snow" is used instead of the term "new snow". In fact, the snowfall is defined as "the depth of freshly fallen snow deposited over a specified period (generally 24 h)". However, following the international classification for seasonal snow on the ground (Fierz et al., 2009), the height of new snow is "the depth in centimetres of freshly fallen snow that accumulated on a snow board during a standard observing period of 24 hours". Therefore, it seems that both "fresh" and "new" snow refer to the snow fallen in a period of 24 hours. We have changed "fresh snow" with "new snow" in accordance with what the Referee asked for.

Lines 21-22, Abstract: "The results indicate that the daily SR50 sonic ranger measures allow a rather good estimation of the SWE, and the provided snow pit data are available for defining the site mean value of fresh snow density." unclear.

We have reworded this sentence of the Abstract.

Line 24, Abstract: We have modified all the SWE values from "m" to "mm" accordingly.

Lines 32-33: "The study of spatial and temporal variability of the water resource deriving from snow melt (i.e. Snow Water Equivalent, SWE) is very important for the estimation of the hydrological balance at catchment scale.." strange english.

We have reworded this sentence of the Abstract.

Line 42: We have added "(the largest valley glacier in Italy)" accordingly.

Line 87: The *impact of the fresh snow age is nowhere mentioned!*

As previously mentioned, we have added some discussion about the processes affecting snow during the time window of a day, corresponding to the period in which we quantify the daily differences in snow depth data used for deriving SWE values.

Line 108: How do you measure only liquid? I guess you mean unheated precipitation gauge.

We have added "(by means of an unheated precipitation gauge)" accordingly.

Lines 111-112: Since there is no information on the web available about this pillow, please provide the information about the size.

We have added some information regarding the snow pillow and the pressure sensor: "AWS Forni SPICE is equipped with a snow pillow (Park Mechanical steel snow pillow, 150 x 120 x 1.5 cm) and a barometer (STS ATM.1ST) for measuring the snow water equivalent (Table 1, Beaumont, 1965)."

Line 114: We have added "the" and "in order to observe the snow depth" accordingly.

Line 121: Following also the suggestions of the first Referee, we have modified with "provided".

Lines 129-131: I guess you have to adjust it manually? How often was it moved upward again? How did the surface height (or ice thickness) change during this 12 years.

We do not mean that we manually move the AWS. It leans on the surface of the glacier, it is not fixed into the ice. Therefore, during the ice ablation period, the AWS follows the lowering of the surface. Following also the suggestions of the first Referee, we have modified the sentence from "adjust" to "move together with". The mean annual ice thickness change is about 4 m per year.

Lines 132-134: The radiation and wind conditions must have changed due to this dislocation, which may also impact the snow conditions?

As the distance between the two sites is about 500 m, the difference in elevation is only 44 m and the aspect is very similar, so we do not expect a noticeable impact of the site change on snow depth.

Line 146: Which time? You did not provide any information about the measurement frequency of the different sensors?

We have added in the "Method" section: "Data points are sampled at 60-second intervals and averaged over a 60-minute time period for the SR50 sonic ranger, wind sensor and barometer, over a 30-minute time period for the sensors recording air temperature, relative humidity, solar and infrared radiation, and liquid precipitation, and over a 10-minute time period for the USH8 sonic ranger and snow pillow. All data are recorded in a flash memory card, including the basic distribution parameters (minimum, mean, maximum, and standard deviation values)."

Line 146: What about the settling of the snow cover during this time?

As far as settling is concerned, $\Delta h_{snow-pit-j}$ from eq. 2 (see answers to the main comments) would indeed be higher if Δh_{tj} values were calculated considering an interval shorter than 24 hours. However, this would not be possible because on the one hand, the sonic ranger data's margin of error is too high to consider hourly resolution, and on the other hand, new snow is defined by the WMO within the context of a 24-hour period. Therefore, settling could not be considered in our analyses since new snow as defined by the WMO already includes the settling that occurs in the 24-hour period.

Line 149: We have modified the sentence from "snow days" to "days with snowfalls".

Lines 150-153: This procedure is not clear. Could you provide more details. It seems to be crucial for the paper!

In the new version of the manuscript, we have modified this part of the "Method" section (following also the suggestions of the first Referee) in order to better clarify our procedure in estimating the density of the new snow ($\rho_{new\ snow}$) used for deriving SWE from snow depth data (see answers to the main comments).

Line 149: Why is this value presented as a new result in the abstract, if it has already been found in previous studies. But none of these studies really demonstrated how exactly the value was determined.

Following also the suggestions of the first Referee, we have modified this paragraph and the Abstract. In the revised manuscript, we have explained better how the new snow density is estimated (see answers to the main comments).

Lines 162-163: How does this compare with neighboring stations?

We have added a comparison with data from 10 stations spread over the Italian Alps: "These values are in agreement with findings over the Italian Alps in the period 1960–2009. In fact, Valt and Cianfarra (2010) reported a mean snow depth of 233 cm (from 199 to 280 cm) for the stations above 1500 m a.s.l."

Line 162: We have modified the sentence accordingly adding "second".

Line 169: You don't know without independent validation! Small snowfalls are impossible to detect due to uncertainty of the snow depth sensors and since there is often concurrent settling even snowfalls larger the measurement uncertainty can not be detected. impossible to detect and

Following also the suggestions of the first Referee, we have deleted this paragraph regarding the comparison between Campbell and Sommer sensors.

Lines 182-184: We have deleted this sentence accordingly.

Lines 186-187: This is strange! Could you provide any explanation? Why should the measurement be correct afterwards?

The results from the snow pillow are difficult to explain as this sensor has been working for only two winter seasons and we are still in the process of testing. Analyzing data from the years to come will allow a more robust interpretation. However, we have searched for a possible explanation of this problem and this error could be due to the configuration of the snow pillow.

Lines 191-192: Why do you think the snow pillow working correctly before January 2015, but not before January 2016?

We found that the valid readings from the snow pillow occur whenever there is a snow cover thicker than about 50 cm. Because of the observed inter-annual variability in snow depth values, the beginning of the valid readings from the snow pillow differs from one year to the next. In fact, in the accumulation period 2015/2016 this threshold is reached after January 2016.

Line 199: I suggest to add these measurements in Fig. 6.

We have added in the Figure the values measured by means of the snow weighing tube, accordingly. The new version of the figure is:
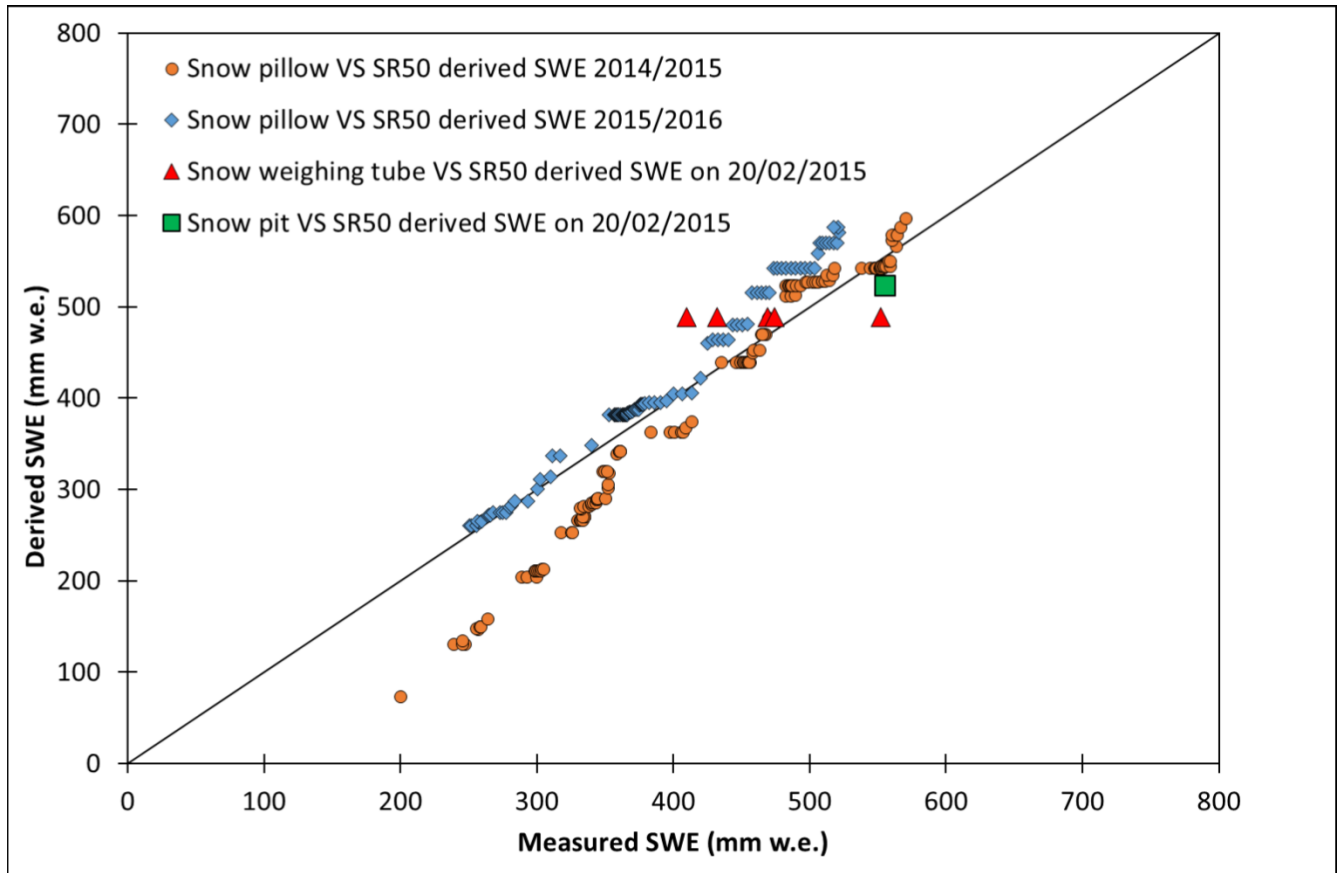
Line 205-206: Improve english!

Following also the suggestion of the first Refeee, we have modified this paragraph.

Line 206: We have added "of the density by" accordingly.

Lines 207-208: Percentage difference instead of the absolute values would provide much more information.

Following also the suggestion of the first Referee, we have added "corresponding to about 17% of the mean total cumulative SWE considering all hydrological years".

Line 212: really tens and not ten?

The ice glacier surface features a lot of different conditions: bare ice, ponds of different size and depth, presence of dust and fine or coarse debris that can be scattered over the surface or aggregated. This surface heterogeneity translates into a differential ablation. In fact, each material has a different value of albedo and heat transfer and then the below ice will receive a different amount of energy for melting. The result is a surface roughness of tens of centimeters.

Lines 223-224: Not clear what you mean?

Following also the suggestions of the first Referee, we have completely modified this paragraph.

Lines 227-229: Which would not be true in your case since you measure over ice!

Following also the suggestions of the first Referee, we have completely modified this paragraph.

Line 232-246: What is this? Does not belong to Discussion and has not direct connection to your results.

Following also the suggestions of the first Referee, we have moved this part to the "Introduction" section.

Lines 254-256: What is this temperature threshold for the AWS1?

We have added accordingly "From the Forni Glacier, the application of the +0.5°C daily temperature threshold allows for a consistent quantification of snow ablation while, instead, for detecting the beginning of the snow melting processes, a suitable threshold has proven to be at least −4.6°C."

Line 260: Following also the suggestions of the first Referee, we have modified this part from "The unique issue is represented by the definition of…" to "Finally, the SR50 sonic ranger features the unique problem of the definition of the start of the accumulation period, but this can be overcome using albedo data."

Figures: We have modified all the figures and captions accordingly.

We very much appreciate the time and effort you put into the comments.

Sincerely,

Antonella Senese and Co-author