Dear Referee

Thanks for helping us in improving the manuscript.

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

We will therefore modify 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 of the following years. In the revised version of the manuscript, we will start with updating the site average new snow density estimation exploiting all the available dataset and we will perform all the subsequent analyses using this value.

The text concerning this issue in "Method" section will therefore be:

In addition to the measures performed by means of 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 been carrying out periodic 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 then the total $SWE_{snow\text{-}pit\text{-}j}$ of the whole snow cover (*n* layers) is obtained:

$$SWE_{snow-pit-j} = \sum_{i=1}^{n} h_{ij} \cdot \frac{\rho_{ij}}{\rho_{water}} \quad (1)$$

where ρ_{water} is density of water. As stated 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 carried out 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 measures, 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 research on data of the same site covering the 2005-2009 period (Senese et al., 2012a). We need updating $\rho_{new\ snow}$ as it is the key variable for estimating *SWE* from sonic ranger new snow data. Specifically, for each snow pit *j*, the corresponding total new snow was first determined by:

$$\Delta h_{snow-pit-j} = \sum_{t=1}^{m} (\Delta h_{tj}) \quad (2)$$

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 of 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 non-sense 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:

$$\rho_{new\ snow} = \frac{\sum_{j=1}^{k} SWE_{snow-pit-j}}{\sum_{j=1}^{k} (\Delta h_{snow-pit-j})} \quad (3)$$

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:

$$SWE_{SR-t} = \begin{cases} \Delta h_t \frac{\rho_{new\ snow}}{\rho_{water}} & if\ \Delta h_t \geq 1\ cm \\ 0 \quad if\ \Delta h_t < 1\ cm \end{cases} \quad (4)$$

We will also apply the leave-one-out cross-validation (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 will apply Eq. (3) (see answer above) once for each snow pit (*j*), using all other snow pits in the relation ($\rho_{new\ snow}$ LOOCV) and using the selected snow pit as a single-item test ($\rho_{new\ snow}$ from snow pit *j*). In this way, we will avoid dependence between calibration and validation dataset in assessing the new snow density.

The results will give evidence that the standard deviation of the differences between the $\rho_{new\ snow}$ LOOCV 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 $\rho_{new\ snow}$ will then be estimated dividing this standard deviation for the square root of the number of the considered snow pits. We will show that it is 6 kg m^{-3}. We will finally show that the new and the old estimates of $\rho_{new\ snow}$ (149 and 140 kg m^{-3}, respectively) do not have a statistical significant difference.

We will also discuss this issue focusing on the snow pit layers (see Eq. (1) – answer above).

Finally, before uploading the reviewed manuscript, as suggested by the second Referee, the standard of English spelling and grammar will be 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 will modify 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 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, it corresponds to a possible new snowfall:

$$\Delta h = \begin{cases} reading\ at\ h23 - reading\ at\ h00\ (if\ positive) \\ 0\ (if\ negative) \end{cases}$$

This issue will be 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 will apply the leave-one-out cross-validation (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 will investigate the *SWE* sensitivity to changes in $\rho_{new\ snow}$: we will calculate SWE_{SR} using different values of new snow density ranging from 100 to 200 kg m^{-3} with a step of 25 kg m^{-3}.

We will also add 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 will then underline 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. Finally, we will present a first analysis of the snow pit.

Moreover, we will investigate if potential errors in individual snowfall events could affect peak SWE_{SR} estimation. In fact, a large snowfall event with a large deviation from the mean new snow density will result in larger errors (e.g. a heavy wet snowfall). Large events 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 checked also if the days in the different new snow intervals have significantly different average temperatures. We did not find any signal.

Another possible source of error in estimating new snow density and in deriving daily SWE is represented by rainfall event. 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 is followed by a rainfall, the fallen new snow could partially or completely melt and then it could not be measured by snow pit carried out at the end of the accumulation season.

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

All these data elaborations and analyses will be 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 will highlight 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 will show 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 will modify 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 snow season 2013-2014 was an attempt to reduce the impact of this problem. The second sonic

ranger, however, was still in its 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 next years it can help reducing the problem of missing data. Redundant sensors 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 to have more data. Unfortunately, they were broken almost immediately after the beginning of the accumulation period. They can be another way to face the problem of missing data, provided we will find out how to avoid they will break 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 will add some discussion about the impact of rain events. In fact, another possible source of error in estimating new snow density and in deriving daily SWE is represented by rainfall event. 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 is followed by a rainfall, the fallen new snow could partially or completely melt and then it could not be measured by snow pit carried out at the end of the accumulation season. This is therefore another potential error that, besides to the ones considered in the answers to the first Referee, could potentially give an underestimated value of $\rho_{new\ snow}$. We again highlight however that the value we find (149 kg m^{-3}) does not seem to suggest an underestimation.

- *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 will modify the title accordingly from “Snow data intercomparison on remote and glacierized high elevation areas (Forni Glacier, Italy)” to “Estimating 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 will modify the abstract accordingly.

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

The dislocation 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. However, we will add some discussions about the site move in the revised manuscript following also the suggestions of the first Referee.

- 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 will add and discuss 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 from about -10% to 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 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 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 will add all these information in the revised manuscript accordingly.
In addition, we will add 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.

- 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 will be 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 will modify it accordingly.

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

Lines 15-18, Abstract: We will delete 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 will change "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 will reword this sentence of the Abstract.

Line 24, Abstract: We will modify 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 will reword this sentence of the Abstract.

Line 42: We will add "(the largest valley glacier in Italy)" accordingly.

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

As previously mentioned, we will add 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 will add "(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 will add some information regarding the snow pillow and the pressure sensor: "The 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 will add "the" and "to be able to determine the snow depth" accordingly.

Line 121: Following also the suggestions of the first Referee, we will modify 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 will modify 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, we do not expect a noticeable impact of the site change on snow depth. However, we will add some discussions about the site move in the revised manuscript following also the suggestions of the first Referee.

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

We will add in the "Method" section: "Data points are sampled at 60-second intervals and averaged over a 60-minute time period for 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 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 be higher if Δh_{tj} were calculated considering a shorter interval than 24 hours. However, on one side this would not be possible because the sonic ranger data have too high error to consider hourly resolution, and on the other side, new snow is defined by WMO considering a 24-hour period. In our opinion, therefore, we do not need considering settling as new snow defined by WMO does already include the settling that occurs in the period used to measure new snow.

Line 149: We will modify 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 will modify 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 will modified this paragraph and the Abstract. In the revised manuscript, we will explain 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 will add 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 for the stations above 1500 m a.s.l."

Line 162: We will modify 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 will delete this paragraph regarding the comparison between Campbell and Sommer sensors.

Lines 182-184: We will delete this sentence accordingly.

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

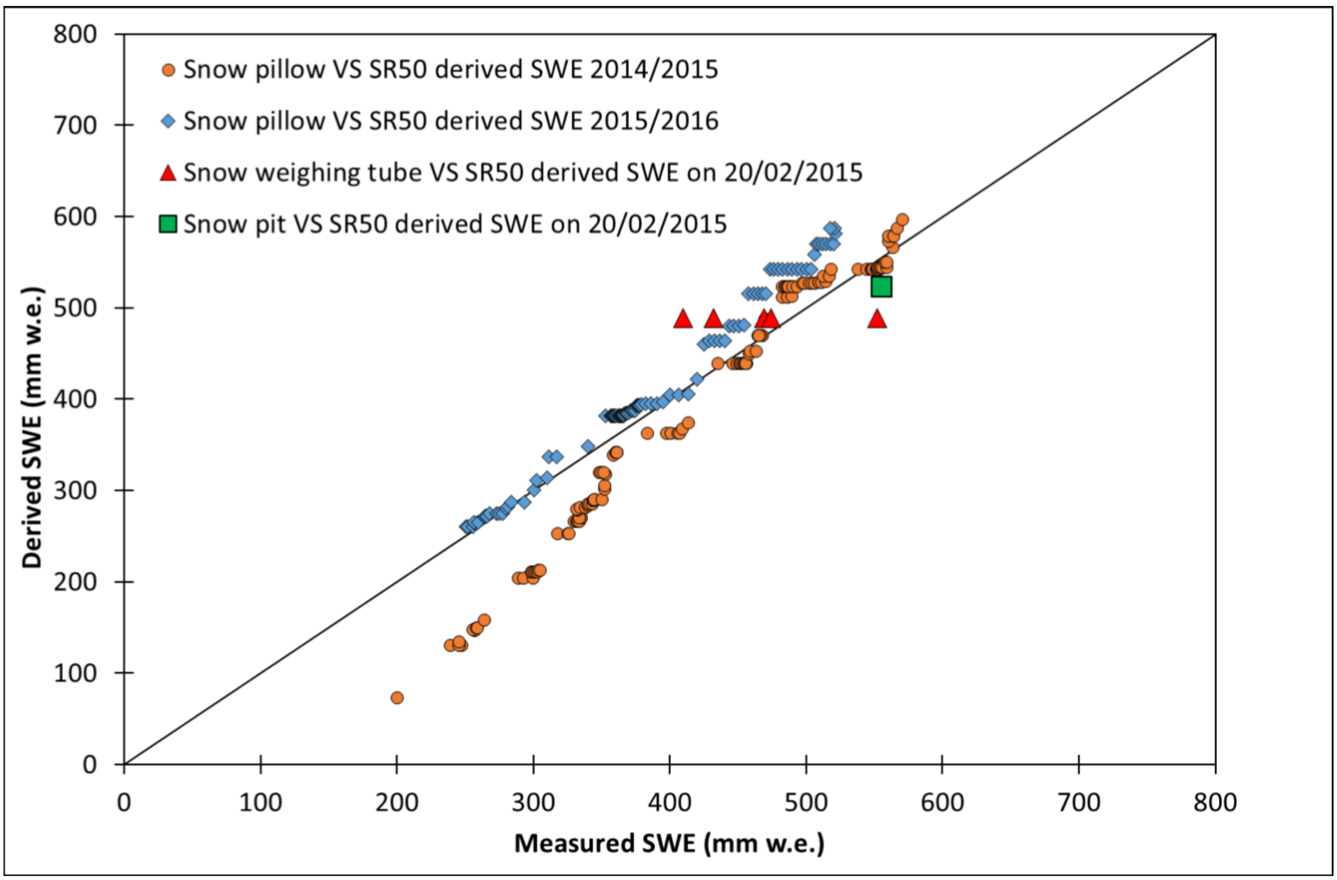
The results from the snow pillow are difficult to explain as this sensor has been working for only two winter seasons and then we are still in the process of testing. Analyzing data of the next years will allow a more robust interpretation. However, we have searched for a possible explanation of this over-weighing accordingly and perhaps this error may 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 will add 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 will modify this paragraph.

Line 206: We will add “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 will add “corresponding to about 14% 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 will completely modify 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 will completely modify 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 will move this part to the “Introduction” section.

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

We will add 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 will modify this part from “The unique issue is represented by the definition of…” to “In spite of all these issues, the SR50 sonic ranger features the unique problem represented by the definition of…”

Figures: We will modify 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