Reviews of the paper tcd-8-4581-2014 submitted to The Cryosphere (Mauro Fischer, corresponding author):

Dear Martin,

We want to thank for your work as the scientific editor of our paper "Surface elevation and mass changes of all Swiss glaciers 1980–2010."
We answered and commented on every single issue addressed by the three anonymous reviewers below. – *Reviewer comments* are formatted in Times 12 italic, our response in Times 12 normal, and the corresponding revised text including information about the **corresponding line numbers in the new TC manuscript version** in Times 10 normal/bold.

Comments by Anonymous Referee #1:

1) DEM differencing / Error Assessment: The authors assume errors in each of the DEM products are uncorrelated, as implied by Eq. 4 where errors are added in quadrature, put provide no evidence that this is in fact true. Uncorrelated errors will be smaller than those showing a correlation structure. A more robust approach is to quantify the spatial correlation lengths through an assessment of the difference map over stable terrain, following Rolstad et al, 2009 (J. Glac. 55/192), and applied for example in Motyka et al., 2010 (J. Glac. 56/198). The more sophisticated error assessment is needed because the author's current approach of taking the mean differences over stable terrain may mask spatial autocorrelation and its impact on the error distribution. In fact, Figure 4 seems to indicate the elevation differences over stable terrain do have strong autocorrelation structure.

> We reassessed the uncertainty in the surface elevation, volume and mass changes presented in this study with a simple implementation of the approach by Rolstad et al. (2009) to the entire Swiss Alps. – We carefully rewrote and extended the entire chapter 4 and adapted and corrected all information concerning uncertainty (text passages and graphs). Please refer to the concerning text passages, figures and numbers in the revised manuscript.

2) Source DEMs and Validation data: The authors choose the dataset of Huss (2010a,b) for validation purposes. This creates some confusion because the Huss 2010 data relies on at least one of the same DEM products (the DHM25 Level 1 data) used in the present analysis. Therefore, the uniqueness of this validation product relative to the new analysis here comes into question. Furthermore, the Huss 2010 data utilizes the SRTM product, which the authors identify as problematic in their introduction. The authors should provide more justification for their use of the Huss 2010 dataset for validation purposes.
Another product used by Huss 2010 is a series of DEMs from aerial photographs, presumably the same ones mentioned here on P4585, and implied to have higher quality than the DEMs used for the present volume change assessment. If better quality DEMs exist than the Swiss ALTI product, it is not clear why they were not used here. The authors should justify their choice of DEM products relative to other work done in this area.

> The validation data relies on ice volume changes derived from a series of high-accuracy photogrammetrical DEMs for sub-decadal to multi-decadal time intervals (see section 2.3). Hence, the DHM25 Level 1 DEMs which we used for the glacier

surface topography at the beginning of our observation period are not used for validation purposes. Also, neither for/in Huss et al. 2010a (GRL) nor for/in Huss et al. 2010b (Erdkunde), which we refer to for validation, the SRTM DEM was used. We have clarified this issue in section 2.3.

Lns 174ff:
"Time series of surface mass balance for glaciers of different type and size class covering the entire Swiss Alps over the last decades (Huss et al., 2010a, b) are used to validate the geodetic mass balances presented here. These series rely on ice volume changes derived from high-accuracy photogrammetrical DEMs for sub-decadal to multi-decadal time intervals (Bauder et al., 2007). By using a distributed mass balance modelling approach including comprehensive field data sets (winter accumulation, summer ablation and discharge measurements), annual mass balance series were calculated that agree with the observed geodetic mass changes."

Regarding the photogrammetrically derived DEMs used for our study, the clarifications made in section 2.3 (Validation data) as well as the new section 4.1 (Quality of input DEMs over selected glaciers) should now allow the reader to understand why and when photogrammetrically derived DEMs were used.

P 4582, Line 20: Replace "The currently observed atmospheric warming caused striking. . ." with "Recent atmospheric warming has caused increased mass loss. . ."

Reformulated accordingly.

Lns 26ff:
"Recent Recent atmospheric warming has caused increased mass loss of glaciers all over the world (e.g. Zemp et al., 2009; Radić and Hock, 2014), which significantly contributes to …"

P 4583, Lines 1-2: In addition to the mass losses reported since the mid-1980s, make a statement about the longer-term losses observed in this region.

Now extended and rewritten accordingly.

Lns 34ff:
"Glaciers of the European Alps showed general mass loss and shrinkage since the Little Ice Age (LIA) maximum around 1850 (Zemp et al., 2008), despite intermittent phases of positive mass and area changes around 1890, during the 1910s and from the late 1970s to the mid-1980s. Since then, pronounced glacier retreat has been reported again (Paul et al., 2011; Huss, 2012)."

P 4583, Line 4: Change "…data is…" to "…data are…"

Implemented accordingly.

Lns 43ff:
"Mass balance data are available either from annual field measurements on individual glaciers …"

P 4583, Line 9: Delete "also"

Deleted.

Lns 50ff:
"…(DEMs) available worldwide and the fact that inaccessible areas and entire glacier systems…"

P 4583, Line 24f: "might cause problems" is quite vague. Specify that Berthier et al found systematic biases at high elevations.

Now clearer.

Lns 72ff:
"Furthermore, applying the medium-resolution SRTM DEM to high-mountain areas is problematic due to the systematic biases reported for high elevations (Berthier et al., 2006)."

P 4584, Line 16-17: "In number, small, thin. . ." is unclear. Instead provide some specific information from the inventory (e.g. ranges of glacier sizes and thickness found in this region). Also, "glacierets" is a standard term (according to the Cogley et al glossary), but "ice patches" is not. Unless you define a difference between these, just use one.

Following these comments, the whole paragraph is now rewritten.

Lns 102ff:
"The study area covers the entire Swiss Alps, where glaciers generally showed rapid mass loss until today after a short period of mass gain between the late 1970s and the mid-1980s (Huss et al., 2010a). Overall, small, thin and rather steep glaciers dominate. In number, almost 90% of all glaciers were smaller than 1.0 km^2 in 2010. At that time, the total glacierized area was 944.3±24.1 km^2, corresponding to an area change of –362.6 km^2 (–27.7%, or –0.75% yr^{-1}) since 1973 (Fischer et al., 2014). If we apply the approach by Huss and Farinotti (2012) to all Swiss glaciers, the average estimated mean ice thickness was 63 m in 2010, and 92% of the estimated total ice volume of 59.9 km^3 was stored within only 10% of the 1420 glaciers comprised in the SGI2010."

P4585, Line 3: Do these vertical accuracies vary with elevation due to poor contrast of snow covered areas in the accumulation area? Also, is there any information on horizontal accuracy of the DEMs?

In order to answer these questions and give more precise information on the accuracy of the DHM25 Level 1 DEMs, we rewrote and extended the corresponding text passage as follows:

Lns 121ff:
"The positional accuracy is reported to range between 2.5 and 7.5 m. The vertical accuracy was estimated by comparison of known spot heights with corresponding cell values of the DHM25 Level 1 DEMs and ranges between 3.7 and 8.2 m on average for rugged highmountain topography depending on individual map sheets (Rickenbacher, 1999; swisstopo, 2000). Because spot heights are typically located at topographical extreme points like hilltops or depressions, the actual vertical accuracy over "average terrain", as, for instance, glacier surface topography, is probably considerably higher."

P4585, Lines 16-23: Can these DEMs from airborne photogrammetry be used as a formal independent check on the DEMs you actually use, rather than the rough quality assessment (< +/- 1m) given here? If you do bring the independent DEMs into your analysis then state clearly what their accuracy is and how geodetic controls were applied.

The DEMs reviewer #1 is referring to here are now used as a formal independent check of the DEMs we actually use (see new subchapter 4.1 and new Figure 4).

Lns 298ff:
"**4.1 Quality of input DEMs over selected glaciers**
Photogrammetrically derived DEMs of the same acquisition dates as the DHM25 Level 1 DEMs for t_1 and the swissALTI^{3D} DEMs for t_2 are available for 13 glaciers covering 207.3 km^2 (15.9% of the total glacierized surface area) in 1973 and 12 glaciers covering 88.2 km^2 (9.3%) in 2010 (Bauder et al., 2007, updated). Statistical comparisons of these DEMs show that for glacierized surface topography, the actual vertical accuracies of both the DHM25 Level 1 and the swissALTI^{3D} DEMs is likely much better

in the glacier-wide mean than assumed by the average error values taken from literature. Both mean (μ = –1.19 m) and median ($\tilde{x}$ = –1.36 m) elevation differences are slightly negative when subtracting the DHM25 Level 1 DEMs from the photogrammetrical DEMs at t_1. For the swissALTI^{3D} DEMs at t_2, mean (μ = 0.50 m) and median ($\tilde{x}$ = 0.21 m) elevation differences are slightly positive and show even closer match with the photogrammetrical DEMs. The standard deviations (1σ level) and interquartile ranges (iqr) of the elevation differences are comparable for both the DHM25 Level 1 DEMs and the swissALTI^{3D} DEMs. Hence, the uncertainties in glacier surface elevation are probably very similar for individual grid cells both at t_1 and t_2 (Fig. 4)."

P 4587, Line 18: The value of 850 +/-60 kg m-3 is from Huss (2013) directly, and so that reference needs to follow this sentence. It is true this compares well with data from these other references, but the primary reference should be Huss (2013).

Implemented accordingly.

Lns 215ff:
"Due to the fairly long observation periods, $f_{\Delta V}$ is set as a constant of 0.85, corresponding to a density of volume change of 850±60 kg m^{-3} (Huss, 2013), which is consistent with other studies (Sapiano et al., 1998; Fischer, 2011; Zemp et al., 2013)."

P4591: A histogram showing the elevation change distribution and standard deviation bounds would help the reader to visualize your results. Also, are the DEM offsets normally distributed? If not, the IQR is a more suitable statistic (see Larsen et al., 2007).

Now implemented as suggested by reviewer #1 (see new Figure 5b).

P 4592, Lines 8-14: DEM co-registration is an important step prior to DEM differencing, especially in mountainous terrain where small planimetric offsets can result in large errors. If you have done the co-registration following Nuth and Kaab, I suggest including this formally as one of your processing steps, instead of at the end of your error assessment.

We calculated the influence of co-registration according to Nuth and Kääb (2011) for the 45 largest glaciers. Because the co-registration of the source DEMs prior to the DEM differencing had only a negligible influence on resulting mass changes (changes inferior to uncertainty of the mass changes), we did not co-register the source DEMs prior to DEM differencing. We reformulated the corresponding text passage in order to be clearer.

Lns 353ff:
"We assume this shift to originate from the creation of the DHM25 Level 1 source data and therefore calculate the influence of its correction via co-registration according to Nuth and Kääb (2011) for the 45 largest glaciers spread over the entire Swiss Alps with an ice surface area of 650 km^2 at t_1. The effect of this correction on the average mass balance of individual glaciers turns out to be in the order of ±10^{-4} to 10^{-2} m w.e. yr^{-1}. Therefore, we consider the impact of the detected DEM shifts on calculated surface elevation, volume and mass changes as negligible for the entire Swiss Alps and do not co-register the source DEMs prior to DEM differencing."

P 4592, Lines 19-20: "Partly differs significantly…" seems contradictory.

Considering Figure 5 we do not think so. However, following reviewer #3s comments on P 4592, Lns 19f, we reworded this text passage.

Lns 452ff:
"For individual glaciers, mean mass balance from Huss et al. (2010a,b) partly differs considerably from our results over the same period."

P 4593, Section heading: The following sections include both results and discussion. Rename the section, or separate out discussion elements into a different section.

Section 5 (Results) indeed includes both results and discussion. – We do not think that it hampers the understandability of our manuscript if the results and a description thereof are not clearly separated from a discussion/first findings of the results. We therefore argue that it's ok to keep the structure of section 5 as it is.

P 4593, Lines 1-2: this is a run-on sentence.

Text passage rewritten accordingly.

Lns 466ff:
"The recent (t_2) area-altitude distribution was derived from the combination of the SGI2010 with the swissALTI3D DEMs. For t_1 the SGI1973 and the DHM25 Level 1 DEMs were used."

P 4593, Lines 6-7: Reword this to state that the area distribution changed such that the maximum area is now at a higher elevation. In the current formulation, the sentence implies the ice moved upstream or that there was a thickening at that elevation.

Reworded accordingly.

Lns 471f:
"The most heavily glacierized areas were located almost 200 m higher in 2010 (Fig. 7a)."

P 4593, Lines 8-9: The section heading is "changes with altitude" but this sentence presents your total delta V for this first time, which is not specifically a result to do with hypsometric changes.

The section heading is now "Changes in surface elevation and area-altitude distribution"

P4595, Line 8: Report the significance level of all of your correlations (p value). A low or high r value is not an indicator of statistical significance.

p values of all correlations are $<1.0*10^{-6}$. Therefore, the given correlation coefficients are significant. For clarification, we added p values in Figure 10 (Figure 11 in the new manuscript).

P4595, Lines 11-15: Can the authors provide some statistics on the strength of the trends that emerge when examining mean values for 5% quantiles?

Correlation coefficients for 5% quantile mean values and corresponding p values of are now included in Figure 10 (Figure 11 in the new manuscript).

P4595, Lines 15-16: If the area/mass balance relation is statistically robust, it has important implications for regional mass balance assessments based on conventional mass balance data. See the recent findings of Gardner et al. (Science, 2013), who propose that the offset between modern geodetic and older mass balance assessments based on conventional mass balance programs could be due to the bias of small glaciers toward more negative mass *balances.*

Following our response to reviewer #1s comments on P 4595, Lns 11-15, we added information about the robustness of the trends for mean values of 5% quantiles. We rewrote the corresponding text passage as follows:

Lns 556ff:
"For average area 1973-2010, the correlation is negative. However, if the smallest glaciers (<0.1 km^2) are neglected, mean average mass balances did not vary considerably for different size classes (Fig. 11a)."

P4595, Lines 16-17: This sentence is not formulated quite correctly. The work of Johannesson et al (1989) relates the glacier response time to a characteristic ice thickness and the rate of elevation change at the terminus. Clarify how this theory predicts a glacier's response to a shift in climate as a function of the appropriate parameters. Also, see the later work of Harrison and others (macroscopic theory of glacier response to climate).

Also due to our implementations of reviewer #1s comments on P 4595, Lns 15-16, we deleted the sentence reviewer #1 is referring to here.

P4596, Section 6: Some but not all of these comparisons have error bounds reported. Can the authors include errors on all of their reported values? This will aid in assessing if the differences are significant or not.

If we did not report error bounds for the reported values it was because they were not directly derived from the DEM differencing but from our time series of annual mass balance (see section 3.1, temporal homogenization via mountain-range mass balance data) in order to compare to reported values of other studies over the same time intervals. These values have an additional uncertainty component resulting from the temporal homogenization which we can not exactly determine.

P 4597: Lines 26-27: ". . .implies that only ice would have melted." This is an incorrect statement. Assuming a density of ice is equivalent to the application of Sorge's Law, which assumes the rate of accumulation and firn densification are time invariant (see Bader, 1954). Snow still melts during the measurement period, but the net loss of mass is in the form of glacier ice.

We rewrote the corresponding text passage.

Lns 666ff:
"Assuming a density of the volumetric change of 900 kg m^{-3} implies that both the mean firn density and firn thickness and area would not have changed within the observation period. For glaciers in the European Alps, however, significant changes in both firn coverage and density are reported (e.g. Carturan et al., 2013; Helfricht et al., 2014)."

P 4597, Lines 27-28: The authors should assess the impact of the density assumption on their mass balance calculations by performing calculations over a range of density values (see for example Johnson et al., 2013, J. Glac.). Then you can say definitively whether the density assumption accounts for the discrepancy with other studies you cite.

Here we compare our results to the study of Paul and Haeberli (2008). As they did assume a constant density of the volume change of 900 kg m^{-3}, it does not make sense to further perform calculations over a range of density values. We assumed the same density of volume change of 900 kg m^{-3} here in order to find out how much of the disagreement in calculated average mass balance between Paul and Haeberli (2008) and our study is related to the difference in density of the volumetric change. We can therefore exclude effects as described in Johnson et al. (2013).

Fig 1: I assume the tick marks are in units of decimal degrees? If so place a degree symbol on the labels, here and throughout.

Implemented accordingly.

Fig 6b: Change caption to read "volume change" not "volume loss". Specify in the caption which x-axis relates to which plotted element. Shouldn't the zero location of the lower and upper x-axes be at the same horizontal location on the plot?

We replaced "volume loss" with "volume change". We argue that there is no real need to further specify in the caption which x-axis relates to which plotted element (stated in the legend). The zero location of the lower and upper x-axes are now at the same horizontal location.

Comments by Anonymous Referee #2:

1) The authors could indicate that their error estimate (@ the 1-sigma confidence level, right?) for individual glaciers is probably conservative because they do not account for the reduction of the error that results from averaging over a large number of pixels on each glacier. But I believe their choice to stay on the conservative side of thing is best. At the scale of the Swiss Alps (Eq 5) the authors assumed that the error for each individual glacier is independent from the neighboring glacier (and thus errors are summed in quadrature) resulting in a very small uncertainty of 0.03 m w.e./yr. This is not so conservative. Can the authors justify that the errors for individual glaciers are uncorrelated? Also the authors should add an error for the temporal correction factor, apparently not included yet.

> We reassessed the uncertainty in the surface elevation, volume and mass changes presented in this study with a simple implementation of the approach by Rolstad et al. (2009) to the entire Swiss Alps. – We carefully rewrote and extended the entire chapter 4 and adapted and corrected all information concerning uncertainty (text passages and graphs). Please refer to the concerning text passages, figures and numbers in the revised manuscript.

*2) Difference with Paul & Haeberli, GRL,2008 (PH2008) 1980s-1999 mass balance estimate. The authors state P4597 L4 that those differences are "considerable" (-0.78 m w.e./yr in PH2008 versus -0.60 m w.e./yr in the present study for the same time period, error bar to be added on the latter value by the way). I would not describe those difference as "considerable", no need to "hammer" an earlier study this way. Unfortunately PH2008 did not provide some error bars in their study but their uncertainties would have been probably relatively high (compared to the errors bars from the present study) and thus the differences would lie within the uncertainty. In fact, I find those differences quite reasonable and one can even speculate more quantitatively about the origin of the systematic differences (as Fischer et al. already do). If all of the difference (PH2008 have a Swiss-wide mass balance more negative by 0.18 m w.e./yr for this 14 year time period) is attributed to the penetration of the SRTM C-Band radar signal into snow and ice (under the sept-1999 surface), one can inferred a mean penetration (0.18*14/rho) of about 2.8 to 3 m (depending on the density used 0.85 or 0.9). A value that makes sense if compared to recent papers on the topic (e.g., Gardelle et al., JoG, 2012; Melkonian et al., JoG, 2014) although unfortunately, to my knowledge, no specific estimate of the SRTM radar penetration depth has been made in the Alps. I leave it to the authors to decide if such a discussion would fit in their revised MS. My opinion is yes, it is worth providing those back of the envelope estimates of the SRTM penetration (or improved ones taking into account the 24% difference due to the fact that the density used was different in the two studies + 17% due to the errors in time span).*
A proposed addition to the paper would be to compare at the basin scale (Table 1) or at the individual glacier level, the relationship (or the lack of it) between area changes and mass balance. The authors have all the data to so.

> We do not want to "hammer" the study by Paul and Haeberli (2008) at all. – We explicitly payed attention to the wording used when comparing our results to those of Paul and Haeberli (2008). In favor of a more neutral phrasing we deleted "considerable". We now extended the discussion of the SRTM C-band wave penetration into snow and ice and the corresponding uncertainty in DEM differencing this may cause. In our opinion, a comparison between area changes and mass balance at the individual glacier or basin scale level is beyond the scope of this paper. We rewrote the paragraph reviewer #2 is referring to here as follows:

Lns 615ff:
"To derive surface elevation and mass changes for the entire Swiss Alps, Paul and Haeberli (2008) compared the DHM25 Level 1 DEMs to the SRTM DEM from February 2000 and combined the former with the SGI1973 and the latter with the SGI2000 created from medium-resolution (30 m) satellite imagery. They assumed t_1=1985 as constant and report an average mass balance of –0.78 m w.e. yr^{-1} between 1985 and 1999. Over the same reference period, we find a cumulative average mass balance for all Swiss glaciers of –0.60 m w.e. yr^{-1} by temporally homogenizing mass changes derived from DEM differencing (cf. Eq. 3). Both the quality of the different source data used and methodologies applied can explain the differences in derived average mass balance to some extent. According to Jarvis et al. (2008), the vertical accuracy of the 90 m resolution SRTM DEM is ±30 m. Over glacierized areas in Switzerland, however, it is probably considerably higher (Paul, 2008). Nevertheless, the quality of the SRTM DEM is not comparable to the recently compiled 2 m swissALTI^{3D} DEMs. If the SRTM DEM is used, the impacts of the penetration of radar waves into snow and ice should be considered (Berthier et al., 2006; Gardelle et al., 2012b), as they can reach up to 10 m for the SRTM C-band (Dall et al., 2001). This could explain the more negative surface elevation changes over accumulation areas observed by Paul and Haeberli (2008) who compared the SRTM DEM to photogrammetrically derived DEMs by Bauder et al. (2007). If all of the difference between the average mass balance 1985–1999 from Paul and Haeberli (2008) and our approach would be attributed to the penetration of the radar signal into snow and ice, one could infer a mean penetration of the SRTM C-band of about 8 m beneath the February 2000 surface (0.18 m w.e. $yr^{-1} \cdot 14$ yr and assuming a mean firn density of 0.65 g cm^{-3} and an accumulation area ratio of 0.5). This would be somewhat higher compared to values reported for the Karakoram (Gardelle et al., 2012b) or southeastern Alaska (Melkonian et al., 2014). Apart from radar penetration, the DEM preprocessing prior to differencing may cause uncertainty, for instance if source DEMs have to be downscaled to the same gird cell resolution (Carturan et al., 2013a). For our study, the 2 m resolution swissALTI3D DEMs were upscaled to 25 m, corresponding to the cell size of the DHM25 Level 1 DEMs. Not only the quality of the source DEMs but also of the glacier outlines used is important. This especially applies if changes of small and very small glaciers have to be assessed. Fischer et al. (2014) show that the accuracy of glacier outlines derived from semi-automatic satellite remote sensing approaches using medium-resolution satellite imagery does not satisfy change assessments of glaciers <1.0 km^2.
Related to the methodological approaches used, the following aspects likely also cause part of the disagreement in calculated average mass balances between Paul and Haeberli (2008) and our study: Assuming a density of the volumetric change of 900 kg m^{-3} implies that both the mean firn density and firn thickness and area would not have changed within the observation period. For glaciers in the European Alps, however, significant changes in both firn coverage and density are reported (e.g. Carturan et al., 2013a; Helfricht et al., 2014). Following Huss (2013), we apply a mean conversion factor of 850 kg m^{-3} here. Furthermore, the average mass balance for individual glaciers (Eq. 2) can be significantly biased if differences in the reference years of the source data are not considered. For Grosser Aletschgletscher, assuming t_1 as 1985 instead of 1980 changes the result by 0.14 m w.e. yr^{-1} (–0.87 instead of –0.73 m w.e. yr^{-1}). Defining t_1 as 1985 for all glaciers of Switzerland would change the mean geodetic mass balance by 0.03 m w.e. yr^{-1}. Hence, differences in the reference years of the source data account for 17%, the different assumptions of the density of volume change for 24%, and the quality of the source data used for 59% of the disagreement in calculated average mass balances between the present study and Paul and Haeberli (2008)."

Title: I suggest adding "during" before "1980-2010"

We would rather keep the title as it is – short and clear.

P 4582, Line 9: I do not think "resulting" is needed.

Now, "resulting" is omitted.

Lns 11f:
"…of the source data used, mass changes are temporally homogenized…"

P 4583, Line 2: Authors could reference here some non-Swiss studies (from colleagues like

Abermann, Carturan or Vincent).

Paul et al. (2011) and Huss (2012) – by chance two Swiss studies – were chosen here because both studies investigated the entire European Alps (Paul et al. (2011) for area changes, Huss (2012) for mass changes). References to studies investigating selected regions of the European Alps mentioned by reviewer #2 – which we do know – are given right below (now Lns 40ff).

P 4583, Line 4: Maybe add here one reference for the Swiss Alps (Paul, GRL, 2004) and one for the French Alps (Gardent, GPC, 2014)?

Implemented. However, we prefer to directly refer to the latest and most updated references here (Fischer et al. (2014) AAAR for the Swiss Alps, Gardent et al. (2014) GPC for the French Alps).

P 4583, Line 25: "cf." not needed before a reference (true in general not only here)

Now, "cf." is omitted everywhere before references.

P 4583, Line 27: "most accurate" is unclear. Do the authors mean more accurate than the SRTM DEM? It is difficult to define the accuracy needed for the DEM to be useful because it depends a lot on the time interval between the compared DEMs.

Meant is: as accurate as possible. Rewritten accordingly.

Lns 77ff:
"Abermann et al. (2010) and Fischer et al. (2014) show that use of as accurate as possible and high-resolution source data is of particular importance for change assessments of these smallest glacier size classes."

P 4584, Line 8: "a consistent period, 1980-2010"

Changed accordingly.

Lns 91f:
"…we temporally homogenize resulting mass changes to a consistent period, 1980–2010."

P 4584, Line 9: "used" not needed I think

Now, "used" is omitted.

Lns 92f:
"This is necessary due to significant differences in acquisition date of the source data."

P 4584, Line 10: "accompanying studies of this type" not needed (seems obvious that you will analyze the source of errors from your method)

Now, "accompanying studies of this type" is omitted.

Lns 93ff:
"We discuss various sources of possible error, perform an in-depth accuracy assessment of our results and ..."

P 4584, Line 12: Is "comment" the right word? Maybe "analyze"?

We prefer to leave this as it is because we did not perform a "complete" analysis of the factors controlling the spatial variability of observed long-term geodetic mass balance. We only touched on this subject.

P 4584, Line 24: " at the time of the beginning of the observation period" does not read well. What about "The initial glacier surface topography"

Rewritten accordingly.

Lns 116f:
“The initial glacier surface topography at the beginning of the observation period (hereafter referred to as t_1) is given by …”

P 4585, Line 2: Clarify if you did or not the interpolation. Are those DEM available freely (the old and recent ones)? If yes provide the URL.

The interpolation to the regular DHM25 Level 1 grids was performed by swisstopo. We argue that this is obvious from the text (now Lns 115ff). Both the old and new DEMs are not freely available (see Acknowledgements).

P 4585, Line 3: "estimated" is not needed. If it is reported, it has necessarily been estimated.

According to the comments of reviewer #1 on P 4585, Ln 3, we rewrote the text passage. Now, it should be clear why the "estimated" is still there.

Lns 122ff:
“The vertical accuracy was estimated by comparison of known spot heights with corresponding cell values of the DHM25 Level 1 DEMs and ranges between 3.7 and 8.2 m on average for rugged highmountain topography depending on individual map sheets (Rickenbacher, 1999; swisstopo, 2000).”

P 4585, Line 5: Why only contour lines? Spot heights also I guess.

For obvious reasons, there are no spot heights over glacierized surfaces. Therefore, the DHM25 Level 1 DEMs over glacierized areas were only interpolated from digitized contour lines.

P 4585, Line 7: cf. not needed.

Implemented accordingly (see comment above).

P 4586, Line 20: "identical coding scheme" is unclear to me.

Now reworded in order to be clearer.

Lns 191ff:
“Because glacier polygons of the SGI2010 were coded and named according to the 1973 outlines they fell into or overlapped with (Fischer et al., 2014), elevation changes could be calculated for individual glacier entities as a next step by ...”

P 4587, Line 14: "balcance" → "balcance".

Done.

P 4588, Line 5: Why providing the equation for a year i and not the full equation for the 1980-2010 period? I think it would make the understanding of the equation easier and in closer agreement with Figure 3. In fact, it was not entirely clear to me how the adjustment was performed. Did the authors take into account the fact that glacier-wide mass balances can vary by a factor of 4-5 (and more from your figure 10...) from one glacier to another while temporal variations are known to be rather homogeneous at the scale of the mountain range? For the adjustment one could imagine a scaling factor for each glacier that would be the ratio of the individual glacier mass balance for the observation period and the mountain range mass balance for the same period. This factor would then be applied to the correcting term. Maybe this is already what the authors did (according to Figure 3) but if so, it could be better described in the text.

Reviewer #2 is right, this is exactly what we did. If we want the reader to understand our temporal homogenization of average mass balances of individual glaciers *g*, we have to provide the equation for a year *i*. From Figure 1 it becomes obvious that the observation period Δt (t_2–t_1) over which the area-averaged specific geodetic mass balance rate ($\dot{B}_g$) is calculated strongly varies for individual glaciers. The deviation of $\dot{B}_g$ from the mountain-range mean over the same respective observation period (overline_B_{t2}–B_{t1}) is, as reviewer #2 correctly mentiones, the "scaling factor" for glacier *g*. $B_{i,\mathrm{mr}}$ is the mean mountain range-mass balance for an individual year *i*. Hence, $B_{i,\mathrm{mr}}$ accounts for the temporal variations. So $B_{i,g}$ is calculated from the mean annual mountain-range mass balance (as derived by Huss (2012) from measured data) of year *i* and the scaling factor for glacier *g* calculated over the observation period of glacier *g*. In this way we were able to calculate annual mass balance data for all 1420 glaciers recorded in the SGI2010 for the time span of *i*=1960 to *i*=2010. It is only then that cumulative mass balances over our defined reference period 1980/81-2009/10 could be calculated for individual glaciers and it is only then that average mass balances of individual glaciers can be compared between individual glaciers. We rewrote the corresponding text passage and hope to be clearer now:

Lns 229ff:

"The deviation of average mass balance $\dot{B}_g$ for glacier *g* (dashed grey line in Fig. 3) from the mountain range mean (black line in Fig. 3) over the respective observation period overline_B_{t2}–B_{t1} is used as a scaling factor to account for the variability in glacier-wide mass balance (Kuhn et al., 1985). The mean mountain-range mass balance from Huss (2012) for an individual year *i*, $B_{i,\mathrm{mr}}$, accounts for the temporal mass balance variability. The annual mass balance $B_{i,g}$ for year *i* and every glacier *g* is thus calculated with

Equation (3).

Because 2010 is the reference year t_2 for most of the investigated glacier entities and the mean observation period is ≈30 years (Fig. 1), the hydrological years 1980/81–2009/10 are defined as the reference observation period over the entire Swiss Alps over which annual mass balances for individual glaciers $B_{i,g}$ are cumulated (grey line in Fig. 3). Using this approach, mass changes are temporally homogenized, can be compared and further analyzed."

P 4588, Line 12: "Analysis of control". I wonder whether this is really "Method". The paragraph rather provides the background behind the study of the factor control the variability of glacier mass balance. This paragraph could probably be split between the introduction (background) and section 5.3.

We fully agree with reviewer #2 that the first paragraph of section 3.2 is providing background information rather than methodological procedures. Nevertheless, we would rather not include the first paragraph of section 3.2 in chapter 1 because this would hamper a clear golden thread through the introduction and extend the latter too much. Also, the analysis of controls of resulting changes is not the main focus of this paper (cf. our response to the comment of reviewer #1 on P 4585 Ln 12). However, the information given in the first paragraph of section 3.2 is necessary to understand what we did and why regarding the analysis of controls. We therefore argue that it is justifiable and best to leave sections 3.2 and 5.3 as they are.

P 4588, Line 13: What do the authors mean by "representative"?

By representative samples we point to a quantity of glaciers which has to be large enough in order not to be influenced by glacier-wide mass balance variability but to reflect average changes at the regional scale (e.g. within a glacierized catchment). By representative observation periods we point to a period of time which has to be long enough in order not to reflect only weather conditions but climatic trends. We extended the corresponding sentence as follows:

Lns 248ff:
"Averaged over representative samples (number of glaciers) and observation periods (number of years), glacier area and elevation changes are usually in agreement with changes in air temperature and precipitation recorded over the investigated areas and time intervals (e.g. Abermann et al., 2009; VanLooy and Forster, 2011; Carturan et al., 2013)."

P 4588, Line 19: Another relevant reference is (Vincent et al., 2005) showing a factor of more than 4 for the cumulative mass balances in the French Alps.

Now, we also refer to Vincent (2002) JGR, which is probably better suited to refer to here compared to Vincent et al. (2005).

Lns 253ff:
"Within a mountain-range and despite similar climatic changes, the differences in long-term mass balance can however be significant between individual – and even adjacent – glaciers (e.g. Kuhn et al., 1985; Vincent, 2002; Larsen et al., 2007; Abermann et al., 2011)."

P 4589, Line 14: "to be able in explaining" sounds a bit weird. To be checked. Maybe "to be efficient in explaining"?.

The sentence is now rewritten as follows:

Lns 287ff:
"Huss (2012) showed that these four geometrical indices can explain some of the variability in observed long-term mass balances."

P 4589, Line 19: The "latter" is unclear because the previous sentence enumerated different variables. Make it clear what sigma_dz is (although obvious). The authors should also make it clear whether they discuss (as I believe) uncertainty at the 1-sigma confidence level.

According to various comments of all three reviewers, we reanalyzed the uncertainty in surface elevation, volume and mass changes presented in our study. The subchapter 4.3 (Uncertainty) should now give enough information about how the different uncertainty estimates were derived and what they mean, and reviewer #2s comment here is regarded as answered.

P 4591, Line 11: Was this mean difference of -1.7 m between the DEM corrected or not? Is *this mean difference varying spatially? (it is varying with altitude and is stronger at altitudes where the glaciers are located by the way...). This value (or better its spatial and altitudinal variation throughout the study area) could be used as an estimate of the systematic error for the elevation difference, a source of errors not accounted for yet. It would result in a more reasonable estimate of the mass balance error (in particular at the scale of the Swiss Alps).*

Pleas refer to our comments right below here.

P 4591, Line 17: Was this mean difference of -1.7 m between the DEM corrected or not? Is this mean difference varying spatially? (it is varying with altitude and is stronger at altitudes where Although I find it interesting, it is not very explicit why the authors calculated this stochastic errors (not used elsewhere I think). Also, in this equation "n" should be the number of independent measurements. It is known that there is some spatial auto-correlation in the errors on the elevation difference so that the number of effective sample is lower than the total number of pixel (Rolstad, JoG, 2009).

We did not correct for systematic error between the DEMs, which we estimated to –1.47±6.82 (1σ) m as calculated from the mean elevation differences of both input DEMs over stable terrain. Also following comments of the other reviewers, the further issues reviewer #2 addresses here are now implemented according to our re-evaluation of the uncertainties in surface elevation, volume, and mass changes.

P 4591, Line 20: "on average": was the error calculated for the stable terrain around each individual glacier and then the average computed? Not entirely clear to me.

This is exactly what we did here. However, we omitted the corresponding text passage in the new manuscript as we do not refer to the stochastic uncertainty anymore (because the latter would stand for one of the 'extreme approaches' to estimate the uncertainty in DEM differencing).

P 4592, Line 5: "a slight horizontal shift".

Rewritten as recommended.

Lns 351:
"This points to a slight horizontal shift in NW–SE direction of the elevation information included in both input DEMs."

P 4592, Lines 9-10: Magnitude of the correction for individual glaciers?

Exactly. Should now be clearer.

Lns 358ff:
"The effect of this correction on the average mass balance of individual glaciers turns out to be in the order of $\pm10^{-4}$ to $\pm10^{-2}$ m w.e. yr^{-1}."

P 4592, Line 22: In Figure 5, I suggest that the authors add the magnitude of their formal error estimates calculated for each individual glaciers so that it can easily be compared to the mass balance differences and would confirm that this formal error estimate is sound.

Now error bars are added to Figure 5 (Figure 6 in the new manuscript).

P 4593, Line 18: Still existing in 2010?

Now reworded in order to be clearer.

Lns 490ff:
"For the entire Swiss Alps, the area-weighted average mass balance of all 1420 glaciers included in the SGI2010 was -0.62 ± 0.07 m w.e. yr^{-1} during our reference period 1980–2010."

P 4595, Line 10: I suggest replacing "good" by "stronger"

Done.

Lns 546ff:
"A weak correlation (r=0.22) was found for median elevation at t_1 (Fig. 11b), and a stronger one (r=0.42) for mean slope over the lowermost 25% of the glacier surface at t_1 (Fig. 11c)."

P 4595, Line 11: The reader wonder why 25% was chosen and not, for example, 10% or 50%? Can the authors justify their choice? Did they test different values and chose the one that led to the higher correlation?

We actually did test different values for which we found the highest correlation for slope over the lowermost 25% of the surface at t_1. 50% would correspond to the whole ablation area at t_1, i.e. to far more than the glacier terminus, if we consider the median elevation of a glacier to be a proxy for the climatic equilibrium line altitude (ELA). In regard to the fact that small glaciers with rather small elevation ranges dominate the sample of glaciers in the Swiss Alps, it is reasonable to assume that the elevation range over the glacier terminus corresponds to more than only the lowermost 10% of the surface at t_1. These rather qualitative arguments provide further support for 25% as a reasonable value. In consequence, we rewrote and extended the 2nd paragraph of section 3.2. as follows:

Lns 281ff:
"In order to identify the controlling factors and to better understand the spatial variability of the observed surface elevation ans mass changes, a correlation analysis between the average mass balance over the reference period 1980–2010 and classes of (i) mean area, (ii) median elevation, (iii) surface slope of the glacier terminus, and (iv) dominant aspect, hereafter referred to as mean aspect, was performed. Huss (2012) showed that these four geometrical indices can explain some of the variability in observed long-term mass balances. For the surface slope of the glacier terminus, the testing of different values indicated that the average surface slope over the lowermost 25% of the glacier at t_1 resulted in the highest correlation."

In section 5.3, it should then be clear why we chose to take the mean slope over the lowermost 25% of the glacier surface at t_1 to account for the geometrical factor slope over the glacier terminus.

P 4595, Line 13: "5%-quantiles of the data" could be explained a bit more (Did the authors separate the whole sample into 20 bins with an equal number of samples in each bin and then compute the MB average in each bin, right?).

Text passage now rewritten accordingly.

Lns 549ff:
"Because part of the significant scatter in Figure 11a-c is likely caused by glacier-individual uncertainties and local effects, we also calculated the respective mean values for 5%-quantiles of the data (triangles in Fig. 11a-c) by computing the mean average mass balance for 20 classes of equal

sample size."

P 4595, Line 17: " because of their longer response time" is not an explanation by itself. Clarify.

Also due to our implementations of reviewer #1s comments on P 4595, Lns 15-16, we deleted the sentence reviewer #2 is referring to here.

P 4596, Line 9: Again not clear why a shorter response time means a less negative mass balance. Is it because those glaciers already adjusted rapidly to warming since 1980s (or since LIA?) so that they already reached a state (at higher elevations) where they are closer to equilibrium to climate and thus do not respond so strongly?

In order to be clearer, we extended the corresponding text passage as follows:

Lns 589ff:
"Because of the stronger influence of the shortwave radiation component and the fact that they are located at higher elevations, south-exposed glaciers generally react less sensitively to air temperature changes than north-exposed glaciers. Also, south-exposed glaciers are often smaller and thinner, and therefore generally have a shorter response time and thus less negative mass balance."

P 4598, Line 21: Did a few glaciers exhibited significantly positive mass balances?

No.

Conclusion: are this dataset available to others? e.g., modelers?

We plan to make the dataset accessible to anyone interested via the WGMS webpage after final publication.

Table 1 sigma_B_ref is defined in the main text but I do not think B_ref is.

Overline_B_{ref} is now defined.

Lns 490ff:
"For the entire Swiss Alps, the area-weighted average mass balance of all 1420 glaciers included in the SGI2010 was –0.62±0.07 m w.e. yr^{-1} during our reference period 1980-2010. For the main hydrological catchments, it ranged between –0.52 m w.e. yr^{-1} and –1.07 m w.e. yr^{-1} (Fig. 8, overline_B_{ref} in Tab. 1)."

Figure 4. Author could add an inset with the distribution (histogram) of the elevation differences on the stable terrain with basic statistics such as mean, media, standard deviation.

Implemented accordingly (new Figure 5b).

Figure 5. The bold dashed line could be made more different (color? Bolder? Dotted?) than the 0.1 intervals. Did a regional pattern emerged in this differences (that would be related to regional biases in the DEMs)? It was not entirely clear to me if the mass balance that are compared here cover the exact same time period (the text say "we choose 31 glaciers from the datasets of Huss et al. (2010a, b) for which volume changes based on the independent, photogrammetrically derived DEMs show closest temporal accord with our respective measured period". Did the authors performed a temporal homogenization before this comparison. If not, Figure 5 should show the time period covered by the photogrammetric DEMs.

Now the dashed line is bolder and error bars are included in Fig. 5 (Figure 6 in the new manuscript). We did not perform a temporal homogenization before this comparison. The new Figure 6 now integrates the time period covered by the photogrammetric DEMs.

Figure 7. Are three digits needed? The fonts could be larger to improve readability.

Implemented accordingly (Figure 8 in the new manuscript).

Figure 8. Why adding the average mass balance for the entire Swiss Alps here?

We added the average mass balance for the entire Swiss Alps to both Figures 9 and 10 in the new manuscript to allow the reader to see more easily which glaciers showed mass changes below and which above the Swiss mean.

Figure 10d. Define the whisker plot (because not all authors use the same representation)

Implemented accordingly (see captions of Figure 11 in the new manuscript).

Comments by Anonymous Referee #3:

1) The study is solid overall. Most comments are minor and given below in the "Specific comments" section. The accuracy assessment has the most room for improvement, although it is extensive already. At this point, it calculates the DEM uncertainties in two different ways, by using the DEM accuracies provided by swisstopo, and by conducting an independent DEM comparison over unglacierized terrain. While this is good, the study applies two extreme approaches for determining uncertainties (resulting in 'nominal' and a 'stochastic' uncertainty, assuming either fully correlated or completely uncorrelated errors). The study does not apply a third approach, which quantifies spatial correlation in the difference grid through variograms. This approach has been used in recent work and should be implemented here as well. See Truessel and others (2013, J.Glac, 59 , p.153) and references therein (Motyka and others, 2010, Rolstad and others, 2009) for more information.

> We reassessed the uncertainty in the surface elevation, volume and mass changes presented in this study with a simple implementation of the approach by Rolstad et al. (2009) to the entire Swiss Alps. – We carefully rewrote and extended the entire chapter 4 and adapted and corrected all information concerning uncertainty (text passages and graphs). Please refer to the concerning text passages, figures and numbers in the revised manuscript.

P 4584: 'within only some few' –> in only a few

> According to the comments of reviewer #1 on P 4584, Lns 16-17, we rewrote the corresponding text passage. Now, the sentence containing "within only some few" is omitted.
>
> **Lns 102ff:**
> "The study area covers the entire Swiss Alps, where glaciers generally showed rapid mass loss until today after a short period of mass gain between the late 1970s and the mid-1980s (Huss et al., 2010a). Overall, small, thin and rather steep glaciers dominate. In number, almost 90% of all glaciers were smaller than 1.0 km^2 in 2010. At that time, the total glacierized area was 944.3±24.1 km^2, corresponding to an area change of –362.6 km^2 (–27.7%, or –0.75% yr^{-1}) since 1973 (Fischer et al., 2014). If we apply the approach by Huss and Farinotti (2012) to all Swiss glaciers, the average estimated mean ice thickness was 63 m in 2010, and 92% of the estimated total ice volume of 59.9 km^3 was stored within only 10% of the 1420 glaciers comprised in the SGI2010."

P 4585, Line 15: Figure 6 indicates that there is some glacierized area below 2000 m asl. What DEM is used there, which technique, and which date?

> Below 2000 m a.s.l., it is also a 2 m DEM but of even higher accuracy because it was created with airborne laserscanning (ALS). We complemented this accordingly but do not give too much detail/answer all the questions of reviewer #3 here because it only concerns a minor fraction of the glacierized surfaces analyzed in our study. We further refer to the product information of the swissALTI^{3D}, where detailed descriptions of the whole data set can be read.
>
> **Lns 150ff:**
> "For areas below 2000 m a.s.l., the swissALTI^{3D} DEMs are more accurate (±0.5 m at the 1σ level) since they were created using airborne laser scanning data (swisstopo, 2013)."

P 4585, Line 22: Mention why: Because the errors are not systematic, they get reduced when averaging over an entire glacier. This finding should be considered in your error assessment.

Now, there is both a new subchapter (4.1) and a new Figure (Fig. 4) addressing these comments of reviewer #3.

Lns 298ff:
"**4.1 Quality of input DEMs over selected glaciers**
Photogrammetrically derived DEMs of the same acquisition dates as the DHM25 Level 1 DEMs for t_1 and the swissALTI^{3D} DEMs for t_2 are available for 13 glaciers covering 207.3 km^2 (15.9% of the total glacierized surface area) in 1973 and 12 glaciers covering 88.2 km^2 (9.3%) in 2010 (Bauder et al., 2007, updated). Statistical comparisons of these DEMs show that for glacierized surface topography, the actual vertical accuracies of both the DHM25 Level 1 and the swissALTI^{3D} DEMs is likely much better in the glacier-wide mean than assumed by the average error values taken from literature. Both mean (μ = –1.19 m) and median (~x = –1.36 m) elevation differences are slightly negative when subtracting the DHM25 Level 1 DEMs from the photogrammetrical DEMs at t_1. For the swissALTI^{3D} DEMs at t_2, mean (μ = 0.50 m) and median (x~ = 0.21 m) elevation differences are slightly positive and show even closer match with the photogrammetrical DEMs. The standard deviations (1σ level) and interquartile ranges (iqr) of the elevation differences are comparable for both the DHM25 Level 1 DEMs and the swissALTI^{3D} DEMs. Hence, the uncertainties in glacier surface elevation are probably very similar for individual grid cells both at t_1 and t_2 (Fig. 4)."

P 4586, Line 1: There is some DEM data from the 60s. A 1960 DEM in conjunction with the smaller 1970 mask will underestimate the volume loss. Or were areas and mass balances stable between the 1960s and 1970s?

Yes, for very few and mostly very small glaciers the combination of a 1960s DHM25 Level 1 DEM with the 1973 outlines might actually underestimate the volume loss. For these glaciers, however, this underestimation is most likely smaller than the uncertainty in the volume change itself. Actually, the mass budget of glaciers in the entire European Alps were close to balanced conditions between 1960 and the mid-1980s, and area changes during this period were only minor, in particular for small glaciers (Huss, 2012, TC). We extended the corresponding text passage accordingly.

Lns 159ff:
"The considerable time difference between the acquisition of the SGI1973 source data and individual DHM25 Level 1 DEMs used for t_1 (Fig. 1) is acceptable as only small area changes and an almost balanced mass budget of glaciers were reported for the European Alps between 1960 and the mid-1980s (Glaciological Reports, 1960-2015; Paul et al., 2004; Huss, 2012)."

P 4586, Line 20: 'Due to' –> Thanks to

According to the comments of reviewer #2 on P 4586, Ln 20, we rewrote the corresponding sentence. Now, "due to" is omitted.

Lns 191ff:
"Because glacier polygons of the SGI2010 were coded and named according to the 1973 outlines they fell into or overlapped with (Fischer et al., 2014), elevation changes could be calculated for individual glacier entities as a next step by ..."

P 4587, Line 18: 850 +-60: add something like "according to Huss (2013)"

Implemented, also according to the comments of reviewer #1 on P 4587, Ln 18.

Lns 215ff:
"Due to the fairly long observation periods, $f_{\Delta V}$ is set as a constant of 0.85, corresponding to a density of volume change of 850±60 kg m^{-3} (Huss, 2013), which is consistent with other studies (Sapiano et al., 1998; Fischer, 2011; Zemp et al., 2013)."

P 4588, Lines 1-11: Add a sentence of justification for this approach. Why is this approach *valid? –> If the mountain range balance has a positive anomaly, then glacier A is also likely to have a positive anomaly. In general, this paragraph reads less well than other parts of the paper. Rephrase/add information so that the reader grasps the idea more quickly. Possibly add another equation: B_norm = sum B_i,g from i = 1980 to 2010 divided by 30.*

Please refer to our answer on reviewer #2s comments to P 4588, Ln 5. We reworded and changed the corresponding text passage in order to be clearer.

Lns 229ff:
"The deviation of average mass balance $\dot{B}_g$ for glacier g (dashed grey line in Fig. 3) from the mountain range mean (black line in Fig. 3) over the respective observation period overline_B_{t2}–B_{t1} is used as a scaling factor to account for the variability in glacier-wide mass balance (Kuhn et al., 1985). The mean mountain-range mass balance from Huss (2012) for an individual year i, $B_{i,\mathrm{mr}}$, accounts for the temporal mass balance variability. The annual mass balance $B_{i,g}$ for year i and every glacier g is thus calculated with

Equation (3).

Because 2010 is the reference year t_2 for most of the investigated glacier entities and the mean observation period is ≈30 years (Fig. 1), the hydrological years 1980/81–2009/10 are defined as the reference observation period over the entire Swiss Alps over which annual mass balances for individual glaciers $B_{i,g}$ are cumulated (grey line in Fig. 3). Using this approach, mass changes are temporally homogenized, can be compared and further analyzed."

P 4588, Line 20: simplify to 'which can explain this variability to a certain extent.'

Done.

Lns 258f:
"Different factors have been identified which can explain this variability to a certain extent."

P 4589: Equation 4: Add reference. For example (Etzelmueller, 2000: "On the quantification of surface changes using grid-based Digital Elevation Models"). What does it stand for? –> Standard propagation of random errors What does it yield? –> The combined per pixel uncertainty. Note that there are other (better) ways to obtain delta sigma z, using variograms (e.g., applied in Motyka and others, 2010).

As we completely re-evaluated the uncertainties in surface elevation, volume and mass changes presented in our manuscript, equation 4 is now omitted. The former Equation 4 included the assumption of spatially totally correlated errors in surface elevation changes (given by the standard deviation of measured point elevation errors of both input DEMs). This would have been one 'extreme approach' of estimating the uncertainty. Please refer to our answers on all reviewer's general comments about the uncertainty assessment.

P 4589, Line 24: What do the vertical accuracies provided by swisstopo (2000) stand for? (one sigma?)

Following also to the comments of reviewer #2 on P4589, L19, we added information on what reviewer #3 is referring to here. From section 2.2 (Lns 214ff in the new manuscript) it should now be clearer that σ_{DEM1} and σ_{DEM2} refer to the average error (except for areas below 2000 m a.s.l. of the swissALTI^{3D} DEMs). This is all we get from the product informations of the DHM25 Level 1 and the swissALTI^{3D} DEMs (swisstopo, 2000; swisstopo, 2013).

P 4590, Line 7: "multiplying with the initial glacier area". A simple multiplication would mean that the per pixel uncertainties are correlated across the glacier area, which is probably not the case given your statement on p. 4585, Line 22 (i.e., your error bounds would be too high). On the other hand, treating the per pixel uncertainties as random would yield errors that are probably too low (as shown in your Eq. 8). The recommended intermediate approach would be that applied by Motyka and others (2010).

We totally agree with reviewer #3 here. Again, we reassessed the uncertainties of our methods and results according to all reviewers comments by implementing a simple form of the approach recommended by Rolstad et al. (2009). Please refer to the new chapter 4 and the various general comments on the re-evaluation of our uncertainty assessment.

P 4590: Equation 5: How do you justify that the sigmas are not just summed up? Assuming that the measurements of individual glaciers are uncorrelated? Add a reference if this equation was used in previous work.

We now give a more conservative estimate of the uncertainty in the total volume change of all Swiss glaciers over the observation period. The glacier-individual uncertainties in volume change are now just summed up (see Equation 6 of the new manuscript).

*P 4590: Equation 6: May be more readable if you combine factors 1 and 2 as well as 3 and 4 into (F1*F2)ˆ2 + (F3*F4) ˆ2*

Implemented as suggested (now Equation 7).

P 4590, Line 16: make clear that the uncertainty comes from Huss (2013).

Implemented accordingly (similar to reviewer #3s comments on P 4587, Ln 18).

Lns 425ff:

"…with a conversion factor $f_{\Delta V}$ of 0.85 and a corresponding uncertainty $\sigma_{f\Delta V}$ of 0.06 derived from the assumed mean density of the total volume change of 850 kg m^{-3} and a corresponding uncertainty of ±60 kg m^{-3} (Huss, 2013)."

P 4591: Equation 7: Again, clearly justify why the numerator is not just summed up.

We now just summed up the numerator, i.e. give a more conservative estimate of the total uncertainty in the average mass balance of all Swiss glaciers during our reference period 1980-2010 (see Equation 8 in new manuscript).

P 4591, Line 9: "over stable terrain" How much terrain did you consider (how many km2, is the area evenly distributed among the aspect categories? Etc.).

We considered about twice the total area glacierized at t_1 for DEM comparison over stable terrain. Because we analyze DEM differences within a mask around glacier entities, we know that different aspect categories are representatively distributed. This should be evident from Fig. 4. We extended the corresponding text passage as follows:

Lns 322ff:

"An analysis of possible elevation-, slope-, and aspect-dependent biases in DEM differencing is carried out by comparing the DHM25 Level 1 DEMs with the swissALTI^{3D} DEMs over stable terrain. The spatial distribution of surface elevation changes outside the glaciers is calculated within a mask around every entity (Fig. 5a) and over about twice the area glacierized at t_1."

P 4591, Line 11: Show the corresponding distribution in addition to Figure 4. Examples are given in Larsen and others (2007) and Truessel and others (2013). Also, state that you did not correct for this systematic shift.

Now implemented accordingly (new Figure 5b).

Lns 326ff:
"The mean offset between both DEMs, which we do not correct for, is -1.47 ± 6.82 (1σ) m (Fig. 5b)."

P 4591, Line 15: ". . .literature-based uncertainty estimates." I assume you mean the values assigned by swisstopo (2000). If so, state this.

Done.

Lns 328ff:
"The mean standard deviation over all stable terrain close to glacierized areas of ±6.82 m agrees with the nominal uncertainty $\sigma\Delta z$ assessed with literature-based uncertainty estimates (Rickenbacher, 1999; swisstopo, 2000)."

P 4591: Equation 8: You calculate the stochastic uncertainty without explaining your motivation for doing so. Also you don't discuss why you refrain from using the stochastic uncertainty for your final error estimates. In fact, the stochastic error is likely too low, because the elevation changes of individual pixels are correlated to some extent. Again, the approach applied in Truessel (2013) and Motyka (2010) would yield error estimates that lie somewhere between the two extreme cases calculated in your work.

We totally agree with reviewer #3 here. We do not include the stochastic uncertainty over stable terrain anymore, because it would again be one 'extreme approach' of assessing the uncertainty in surface elevation, volume and mass changes. Again, please refer to the new chapter 4 and the various general comments on the re-evaluation of our uncertainty assessment.

P 4591, Line 23: Discuss reasons for this increase with elevation. I would have assumed that this is due to the more rugged terrain (i.e., steeper slopes), but you rule that out in the next sentence. Other reasons?

Implemented as suggested.

Lns 340ff:
"Together with the scatter of the bias, which also increases towards higher elevations, this is probably due to the general occurrence of more rugged and steeper terrain with increasing elevation."

P 4592, Line 8: Elaborate on the Nuth and Kaeaeb approach: Did the approach suggest any shift, etc.

We calculated the influence of co-registration according to Nuth and Kääb (2011) for the 45 largest glaciers. Because the co-registration of the source DEMs prior to the

DEM differencing had only a negligible influence on resulting mass changes (changes inferior to uncertainty of the mass changes), we did not co-register the source DEMs prior to DEM differencing. We reformulated the corresponding text passage (also according to the comments of reviewer #1 on P 4592, Lns 8ff) in order to be clearer.

Lns 353ff:
"We assume this shift to originate from the creation of the DHM25 Level 1 source data and therefore calculate the influence of its correction via co-registration according to Nuth and Kääb (2011) for the 45 largest glaciers spread over the entire Swiss Alps with an ice surface area of 650 km^2 at t_1. The effect of this correction on the average mass balance of individual glaciers turns out to be in the order of $\pm 10^{-4}$ to 10^{-2} m w.e. yr^{-1}. Therefore, we consider the impact of the detected DEM shifts on calculated surface elevation, volume and mass changes as negligible for the entire Swiss Alps and do not co-register the source DEMs prior to DEM differencing."

P 4592, Line 20: 'considerably' rather than 'significantly'

Changed accordingly.

Lns 452ff:
"For individual glaciers, mean mass balance from Huss et al. (2010a,b) partly differs considerably from our results over the same period"

P 4592, Line 22: It would help if error bars were integrated into Fig. 5. This would indicate how reasonable your error estimates are.

Now error bars are included in Fig. 5 (Figure 6 in the new manuscript).

P 4592, Line 26: 'same order of magnitude'. Is this something you assume based on Figure 5? If so, you should add a 'likely'. Or do you have additional analyses that would support this statement?

Yes, indeed, we assume this based on Figure 5 (Figure 6 in the new manuscript). Now, a "likely" is added.

Lns 459ff:
"…the accuracy of the average geodetic mass balance is likely in the same order of magnitude as if derived with more precise source DEMs created, for instance, by photogrammetrical techniques."

P 4592, Line 27: Delete 'for instance, by photogrammetric techniques.'

We prefer not to do so because otherwise it makes no sense to write '…is likely in the same order of magnitude…' before. We wrote '…for instance, by photogrammetric techniques…' because we validate our results with data derived from differencing of photogrammetrical DEMs (see also Fig. 6 in the new manuscript).

P 4593, Line 8: 'whereof' –> of which

Done.

Lns 474f:
"…, of which glaciers still present in 2010 account for…"

P 4593, Line 11: 'lowermost elevations'. Maybe mention that you have this typical 'knee' in the curve, with max. elevation changes above the lowermost elevation, due to the glacier *retreat.*

We extended the corresponding text passage as follows:

Lns 477ff:
"Corresponding average elevation changes were in good agreement with theoretical considerations by Schwitter and Raymond (1993) and continuously decreased from largest changes nearly at lowermost elevations (terminus of valley glaciers) towards zero in the accumulation area."

P 4593, Line 16: "state of disequilibrium". Elaborate a little more on this. What are the reasons for the elevation changes above 3500 m asl (Surface mass balance? influence of flow dynamics?)

We rewrote the corresponding sentence as follows:

Lns 484ff:
"The observed thinning at high altitudes and over the accumulation areas of glaciers results from a combination of ice flow dynamics and reduced accumulation and emphasizes the current state of disequilibrium of glaciers in the Swiss Alps."

P 4594, Line 24: 'is a good example to explain'–> 'illustrates the influence of '

Now the corresponding text passage is omitted.

P 4595, Line 8: What does significant mean here? Did you test for significance or does it stand rather for 'considerable'? In general, make sure to calculate significance levels and be careful with interpreting non-significant relationships.

Here, "significant" refers to the significance level of the correlations mentioned (and shown in Fig. 10). To clarify this (also according to reviewer #1), we computed p values for all correlations and added them to Figure 10.

P 4595, Line 9: Elaborate how you obtained the correlations for the aspects. Did you fit a straight line into the points, previously sorted by eight aspect bins? Or did you actually use the sine and cosine components as done in previous work (Evans and Cox, 2005)?

We actually rearraged the initially eight classes of dominant aspect (N, NE, E, SE, S, SW, W, NW) into five new classes of equivalent potential clear sky radiation (N, NW/NE, W/E, SW/SE, S) prior to the correlation analysis. We extended the corresponding text passage in section 3.2 (methods chapter):

Lns 293ff:
"…For mean aspect, the initially eight classes were rearranged into five classes of equivalent potential clear-sky radiation (N, NW/NE, W/E, SW/SE, S) prior to the correlation analysis."

P 4595, Line 10: 'a good one' –> the strongest one. An r = 0.42 indicates that about 18% of the variability can be explained with the slope variable. Also, state why you used the slope of the lowermost 25%? Huss (2012) used the slope of the lowermost 10%.

Implemented accordingly, also reffering to comments of reviewer #2 on P 4595, Ln 10.

Lns 546ff:
"A weak correlation (r=0.22) was found for median elevation at t_1 (Fig. 11b), and a stronger one (r=0.42) for mean slope over the lowermost 25% of the glacier surface at t_1 (Fig. 11c)."

We tested different values for which we found the highest correlation for slope over the lowermost 25% of the surface at t_1. 50% would correspond to the whole ablation area at t_1, i.e. to far more than the glacier terminus, if we consider the midpoint elevation of a glacier to be a proxy for the climatic equilibrium line altitude (ELA). In regard to the fact that small glaciers with rather low elevation ranges dominate the sample of glaciers in the Swiss Alps, it is reasonable to assume that the elevation range over the glacier terminus corresponds to more than only the lowermost 10% of the surface at t_1. These rather qualitative arguments provide further support for 25% as a reasonable value. In consequence, we rewrote and extended the 2nd paragraph of section 3.2. as follows:

Lns 291ff:
"In order to identify the controlling factors and to better understand the spatial variability of the observed surface elevation ans mass changes, a correlation analysis between the average mass balance over the reference period 1980–2010 and classes of (i) mean area, (ii) median elevation, (iii) surface slope of the glacier terminus, and (iv) dominant aspect, hereafter referred to as mean aspect, was performed. Huss (2012) showed that these four geometrical indices can explain some of the variability in observed long-term mass balances. For the surface slope of the glacier terminus, the testing of different values indicated that the average surface slope over the lowermost 25% of the glacier at t_1 resulted in the highest correlation."

P 4595, Line 13: '5-% quantiles'. Explain how you obtained them (I assume sorted by the respective variable and then filled into the 5% bins by number).

Text passage now rewritten accordingly, also reffering to comments of reviewer #2 on P 4595, Ln 13.

Lns 549ff:
"Because part of the significant scatter in Figure 11a-c is likely caused by glacier-individual uncertainties and local effects, we also calculated the respective mean values for 5%-quantiles of the data (triangles in Fig. 11a-c) by computing the mean average mass balance for 20 classes of equal sample size."

P 4595, Line 15: Did you conduct a correlation analysis for those binned values? Are the fits significant? What are the corresponding correlation coefficients?

Correlation coefficients for 5% quantile mean values and corresponding p values of are now included in Figure 10 (also according to reviewer #2s comments on P 4595, Lns 11-15).

P 4595, Line 17: 'longer response times' implying that they are 'more out of equilibrium' or 'lag behind the climatic forcing '. State that here.

Also due to our implementations of reviewer #1s comments on P 4595, Lns 15-16, we deleted the sentence reviewer #3 is referring to here.

P 4596, Line 17: 'the same methods as ' –> our method for

Reworded accordingly.

Lns 605f:
"Applying our method for temporal homogenization of mass changes…"

P 4596, Line 18: '-0.65' add error

Referring also to our answer to a similar comment of reviewer #1, we did not report error bounds for two reported values because they were not directly derived from the DEM differencing but from our time series of annual mass balance (see section 3.1, temporal homogenization via mountain-range mass balance data) in order to compare to reported values of other studies over the same time intervals. These values have an additional uncertainty component resulting from the temporal homogenization which we can not determine.

P 4596, Line 23: '-0.39' add error

See our answer above.

P 4597, Line 5: Combine the two paragraphs.

Done.

P 4597, Line 7: significantly –> considerably

Changed accordingly.

Ln 629ff:
"Over glacierized areas in Switzerland, however, it is probably considerably higher (Paul, 2008)."

P 4598, Line 6: What percentage is due to outline quality and what percentage due to DEM quality?

Unfortunately we do not know this. Neither can we give reasonable estimates here.

Figure 3. add t_1 and t_2 to the plot replace "measured period" with "observation period", "measured" with "observed"

Implemented as suggested.

Figure 4. Discussed above. Add an additional figure with error distribution. Is there a small polygon in Griesgletscher that should not be in there? (the one intersecting the 2700 m contour)?

Figure 4 (Figure 5 in the new manuscript) is now corrected and extended according to the comments of reviewer #3.

Figure 4. The slopes to the NW of Griesgletscher appear to have a systematic shift while the slopes to the SE do not (admitting that the slopes in the SE sections are flatter, implying that the same shift may show up as smaller elevation difference). Nevertheless, this begs the question whether the shifts are systematic over large areas or systematic only on a "local scale". The latter case would not be corrected with the approach of Nuth and Kaeaeb, I think (in general, such errors would be difficult to correct properly). Also, Fig. 4 indicates that an additional buffer around the 1970 outlines would have been appropriate, as the terrain is

particularly unstable in recently deglacierized areas.

These points are now briefly addressed in the text.

Figure 5. Discussed above. Add bars with uncertainties.

Now error bars are added to Figure 5 (Figure 6 in the new manuscript).

Figure 7. No need to show all of Switzerland. Crop the left side and the top and so that you can show the glacierized areas larger.

Changed accordingly (Figure 8 in the new manuscript).

Figures 8. and 9. Is there a way to add uncertainties for each or selected glacier(s), which would allow the reader to better interpret the results?

Adding uncertainties to all glaciers shown in Figures 8 and 9 (Figure 9 and 10 in the new manuscript) was not be possible as it would have hampered the visibility. Uncertainties in average mass balances of all glaciers will be visible and accessible to anyone as we plan to publish the entire dataset via the WGMS webpage after final publication of the paper.

Figure 10. Hard to read as is. Increase thickness of box plots, increase point size. d) Does a linear correlation coefficient make sense for aspect categories? What do the whiskers stand for? 1.5 IQR?

We implemented the suggestions by reviewer #3 accordingly. A linear correlation does indeed not make much sense for aspect categories, but for classes of equivalent potential clear sky radiation as for what we computed it for Figure 10d (Figure 11d in the new manuscript) does.