Reply to 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.

> Chapter 4 (accuracy assessment and validation) will be changed and extended as we will reassess the uncertainties of our methods and results according to all reviewers comments on these issues, including a re-evaluation of the question whether errors of the DEMs used are spatially correlated or not. We will try to apply the approach by Rolstad et al. 2009 and applications thereof (e.g. Motyka et al. 2010) to our study.

*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:

"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 an average mass balance of −0.60 m w.e. yr^{-1}. Both the quality of the different source data used and methodologies applied can to some extent explain the differences in derived average mass balance. 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 o the SRTM DEM is not comparable to the recently compiled 2m 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 C-band of SRTM (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 3 m under the September 1999 surface, which would be in good agreement with values reported for the Karakoram (Gardelle et al., 2012b) or southeastern Alaska (Melkonian et al., 2014). Apart from radar penetration, the DEM processing prior to differencing may cause uncertainty, for instance if…"

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.

"… 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 (P 4583, Line 4ff).

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 25; P 4585, Line 7; P 4590, Line 15; P 4594, Line 23; P 4595, Line 1; P 4595, Line 17; P 4595, Line 27; P 4596, Line 5).

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.

"Abermann et al. (2010) and Fischer et al. (2014) show that use of as accurate as possible and high-resolution source data is…"

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

Changed accordingly.

"…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.

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.

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.

"The initial glacier surface topography at the beginning of the observation period (hereafter referred to as t_1) is given…"

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 (P 4584 Lns 25ff). 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.

"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 for rugged high-mountain topography depending on individual map sheets."

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 above).

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

Now reworded in order to be clearer.

"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…"

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:

"The deviation of the glacier-individual average mass balance $\dot{B}_g$ (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 glacier-wide mass balance variability (Kuhn et al., 1985). The mean mountain-range mass balance from Huss (2012) for an individual year *i*, $B_{i,\mathrm{mr}}$, accounts for temporal mass balance variability. The annual mass balance $B_{i,g}$ for year *i* and any 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 Bi;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 provide 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:

"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).

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:

"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.

As the uncertainty in surface elevation ($\sigma_{\Delta z}$) is the first of the enumerated variables we rewrote this text passage (also implementing comments of reviewer #3 on P4589 Eq. 4) as follows:

"The uncertainty in surface elevation, volume and mass changes presented in this study is mainly given by the uncertainty related to the two DEMs used. Following Etzelmüller (2000), the former, $\sigma_{\Delta z}$, is defined as:"

From section 2.2 it should now be more clear 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 just what we get from the product informations of the DHM25 Level 1 and the swissALTI^{3D} DEMs (swisstopo, 2000; swisstopo, 2013).

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).

We did not correct for systematic error between the DEMs, which we estimated to -1.7 m as computed by the mean difference of the DEMs over stable terrain. Also

following comments of the other reviewers, the further issues reviewer #2 addresses here will be implemented in the re-evaluation of the uncertainty and errors of our approach.

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 -1.7 m as computed by the mean difference of the DEMs over stable terrain. Also following comments of the other reviewers, the further issues reviewer #2 addresses here will be implemented in the re-evaluation of the uncertainty and errors of our approach.

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. We extended the corresponding text passage as follows:

"… with *n* the number of pixels for which DEM comparison over stable terrain is carried out, is ±0.7 m averaged over the total area considered around individual glaciers."

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

Rewritten as recommended.

"This points to a slight horizontal shift in NW–SE direction of the elevation information included in the DHM25 Level 1 DEMs."

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

Exactly. Should now be clearer.

"Because 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 $\pm 10^{-2}$ 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.

P 4593, Line 18: Still existing in 2010?

Now reworded in order to be clearer.

"For the entire Swiss Alps, the area-weighted average mass balance of all 1420 glaciers included in the SGI2010 was –0.62±0.03 m w.e. yr^{-1} during our reference period 1980–2010."

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

Done.

"A weak correlation (r=0.22) was found for median elevation (Fig. 10b), and a stronger one (r=0.42) for mean slope over the lowermost 25% of the glacier (Fig. 10c).

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:

"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 mean area 1973–2010, median elevation, surface slope of the glacier terminus, and dominant aspect, hereafter referred to as mean aspect, was performed. Huss (2012) showed that these four geometrical indices can explain some of the variability of observed long-term mass balances. For the surface slope of the glacier terminus, the testing of different values indicated that taking the average surface slope over the lowermost 25% of the glacier at t_1 resulted in the highest correlation."

On P 4595, Line 11, 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.

"Because part of the significant scatter in Figure 10a-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. 10a-c) by computing the mass balance average 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:

"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.

"For the entire Swiss Alps, the area-weighted average mass balance of all 1420 glaciers included in the SGI2010 was –0.62±0.03 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. 7, 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.

Figure 4 is now extended accordingly.

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. We did not perform a temporal homogenization before this comparison. Figure 5 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. 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 8 and 9 to more easily allow the reader to see 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.