Detailed author reply to review comments on "Uncertainties and re-analysis of glacier mass balance measurements"

M. Zemp et al.

Correspondence to : M. Zemp (michael.zemp@geo.uzh.ch)

This is a supplement to the general author reply to review comments.

(A) Feedback by M. Pelto (Referee)

1)In several locations within the paper the ELA and AAR are treated as solely derived elements from the mass balance record that can then be adjusted based upon mass balance recalibration. This is true in some cases, but in many cases the ELA and AAR are directly determined from ground, aerial or satellite images. As such they are input to not output from the glaciologic method. If this is the case then they cannot be adjusted as part of the recalibration. Of the 12 glaciers that I submit Ba data on to the WGMS all six that have a reported ELA and all 12 with a reported AAR are based solely on direct observations.

Zemp et al.: Clarified in the paper. See general comments section.

2)The glaciologic method relies on extrapolation from index measurement sites across the glacier and requires an understanding of distribution of mass balance across the glacier determined from an initial detailed survey. With significant changes in glacier area and elevation this distribution may change. Ideally every decade or so, the distribution of mass balance would be reassessed using a detailed survey. This was the concept behind the USGS Benchmark glacier series that is discussed by Fountain et al (1997 and Van Beausekom et al (2010). This would help both identify and prevent bias. An example of this is on Lemon Creek Glacier in Alaska in 1961, 1984 and 1998 (Miller and Pelto, 1998).

Zemp et al.: A corresponding recommendation was added to the paper. See general comments section.

3)The authors sketch the ideal situation where a DEM can be generated for the geodetic assessment of mass balance, which is appropriate. There is a second common data type used in geodetic assessment of elevation change that can be compared to mas balance aerial laser altimetry or ICESat imagery along specific profiles often the center line (Gardner et al, 2012; Sapiano et al, 1998). Given the comprehensive nature of this paper, it would be beneficial to have at least a comment on how such a data set should be utilized.

Zemp et al.: We extended the existing statement about the extrapolation issue of geodetic survey profiles in Paragraph 1 of Section 2.3 (Geodetic observation method). See general comments section.

Specific Comments are below

794-10: Besides the measurement points often the observed ELA as an input to the mass balance and this is then not a single point.

Zemp et al.: clarified in text.

795-4: ELA is often observed and not calculated. WGMS now asks for this observation in Ba data submission.

Zemp et al.: Snowline is observed, ELA is calculated. See general comments section.

795-21: Pelto (2000) examined the impact of extrapolation on Columbia Glacier and Lemon Creek Glacier and found that reducing the number of data points from 300 to 40 on these glaciers did not lead to much higher error, but reduction from 10 to 40 did.

Zemp et al.: Reference to Fountain and Vecchia (1999) and to Pelto (2000) included.

797-10: Better wording needed. Note that these are summer accumulation type glaciers where the main accumulation and ablation season coincide.

Zemp et al.: Reworded.

797-21: May not be the ideal location to address this. However, it must be stated, that an underlying principle of the glaciological method is that a few measurements can be used to extrapolate the mass balance distribution across a glacier once that distribution is determined with a detailed survey. A key to reducing errors in the long run for field mass balance programs on glaciers with changing surface areas is the periodic detailed reassessment of the mass balance distribution. This has been done on Lemon Creek Glacier where else?

Zemp et al.: The „underlying principle of the glaciological method" is stated in the first paragraph of Section 3 (P802, L20-22). The point of periodically reassessing the mass balance distribution was included (see general comments section).

798-3: This is a very comprehensive paper, and as such would be ideal to serve as the template for geodetic-glaciological mass balance comparison. A key method that has been used for geodetic mass balance assessment in Alaska and the Canadian Arctic is airborne laser altimetry of a center line profile (Gardner et al, 2012; Sapiano et al, 1998). It would be worth a brief note on how such a comparison should be utilized and tied into a DEM from a different time. Lemon Creek Glacier, Alaska (Miller and Pelto, 1999) is an example where laser profiling and a separate DEM provide verification for surface mass balance, albeit not in the rigorous and best practice scheme outlined here by Zemp et al (2013). In this case what would the ideal approach be for the best comparison?

Zemp et al.: We extended the existing statement about the extrapolation issue of geodetic survey profiles in Paragraph 1 of Section 2.3 (Geodetic observation method). See general comments section.

799-25: The level of detail here for 3-D co-registration etc far exceeds that in other portions of the paper, should this section through page 800 be moved to Appendix B?

Zemp et al.: we moved two sections to Appendices A and B. See also general comments section.

802-21: This statement is true only if the distribution of mass balance is known from detailed surveying so that the few measurement points can used appropriately.

Zemp et al.: We agree with regard to capturing the spatial distribution (and added a corresponding statement, see general comments section) but disagree for temporal variability (see Lliboutry, 1974; Vallon).

812-10: This is not a correct assumption since ELA and AAR are often determined from aerial, ground or satellite imagery and are not output from the glaciological method. If they are not, they should be since they are both reliable indicators that can help in the homogenization process.

812-19: Nor does it change ELA or AAR measurements that were map based field or image observations.

Zemp et al.: We clarified the difference between observed snowline and equilibrium line. See general comments section.

817-16: separate- above described

Zemp et al.: done

818-10: What is the typical density per km2 of measurement points on the glacier.

Zemp et al.: we do not give a typical density of the observational network since this varies from glacier to glacier and over time. Moreover, the density of the sampling sites is not the only parameter to be considere; others are the spatial variability of the mass balance, its stability over time, and the stake-to-stake correlation of the mass balance. If the spatial variability were to be known perfectly and to be stable over time, a single point balance would be enough to record the year-to-year variation, regardless of the glacier size (e.g. Rasmussen, GRL 2004).

820-12: A detailed mass balance distribution survey every decade would be just as crucial in an era of changing glacier area, elevation profile and climate, as is a geodetic survey. The geodetic survey checks the overall cumulative mass balance accuracy the former whether a bias is being introduced from assumptions that may no longer be valid. On Wolverine and Gulkana Glacier Van Beusekom et al (2010) noted that due to glacier retreat and ELA rise the need to re-define appropriate index sites through an expansion and upward migration of the site networks to readjust the original partitioning of the area represented by each stake. They also identified that the sensitivity of estimated balances to sparse input data further motivated deployment of expanded stake networks to better define the shape and stability of the balance gradient. This is an important and detailed reference examining recalibrating mass balance work that should be referenced.

Zemp et al.: A corresponding recommendation (and the reference to Van Beusekom et al., 2010) was added to the paper. See general comments section.

Figure 2: In the caption separate elevation from distribution.

Zemp et al.: done.

Table Appendix C: A better caption for this table is needed. I was not able to follow what each column represented.

Zemp et al.: we added a corresponding list of abbreviations.

(B) Feedback by G. Cogely (Referee)

Substantive Comments

P789

Title It is not necessary to hyphenate "reanalysis". I suggest following the meteorologists (and the *Glossary of Glacier Mass Balance*) and treating it as one word. To avoid possible ambiguity, however, its meaning in glaciology should be defined in the Introduction.

Zemp et al.: We deleted the hyphen. For further clarification, we add a brief definition in Section 3 and use 'reanalysing' rather than 'reanalysis' since the latter is strongly linked to atmospheric reanalysis products. See also the general comments section.

P791

L4 "Until the present". But "Until recently" would be more accurate (see for example the cited papers by Thibert et al. 2008 and Husset al. 2009).

Zemp et al.: done.

L18 "measurements by the glaciological method". A long-standing problem with the WGMS archive is the failure of observers to contribute their geodetic measurements.

Zemp et al.: we agree, see general comments section. No changes made in paper.

L25 "decadal": I feel discomfort at the consistent use in the paper of "decadal", which has an exact meaning, in the sense of "multi-annual" or even "long-term". Either of these alternatives would be preferable.

Zemp et al.: revised in entire text; 'decadal' was replaced in most cases by 'multi-annual'.

P794

L1 "specific (i.e. glacier-wide) mass balance": "specific" does not mean "glacier-wide". It means "per unit area". All that is needed here is "To obtain the glacier-wide mass balance", but the rest of the text needs checking for this.

Zemp et al.: done.

L4-6 This very awkward sentence ends up hinting that one ought not to measure the mass balance of floating ice. Say something like "Floating glacier tongues and ice shelves are not considered here, because their mass balance is often dominated by frontal and basal terms that are not addressed by the glaciological method."

Zemp et al.: done.

P795

L2 "conventional balances": do not introduce this technical term without defining it. More simply, just delete "conventional" and leave the definition where it is now on P796.

Zemp et al.: done.

P796

L1-8 This discussion of conventional and reference-surface balances needs more clarity. Firstly, most measurements are actually mixtures of the two kinds, unless the area adopted in any one year is extrapolated somehow from the shrinkage rate determined from the two most recent geodetic surveys (or, less commonly, measured every year). Secondly and probably more importantly, the reference surface has not only a particular area but also a particular distribution of elevations; with modern GNSS technology, most point measurements nowadays are probably "conventional" already, so that the actual measurements are more like mixtures than ever.
I suggest clarifying explicitly that the analysis in the paper is focussed on conventional balances. A reference-surface balance can only be obtained by correcting both to the reference area and to the reference surface elevation. This introduces further elements of uncertainty, and the uncertainty in the conventional balance is difficult enough to manage anyway. (This is not to dispute the importance of the later analysis aimed at ensuring that comparisons of glaciological and geodetic balances are comparisons of like with like.)

Zemp et al.: Done by introducing the following statement: "Note that the analysis in this paper is focussed on conventional balances. Obtaining reference-surface balances would require correcting both to the reference area and to the reference elevation. This can only be solved with a distributed mass balance model (e.g., Paul, 2010; Huss et al., 2012) and would introduce further elements of uncertainty."

P797

L5-11 One could add to this list glaciers on which most accumulation is by avalanching (e.g. in the Himalaya) and most ablation is by sublimation (e.g. the highest glaciers in Bolivia or Tibet). But see my comment below on section 2.2 as a whole.

Zemp et al.: entire paragraph deleted in order to shorten Section 2. See also general comments section.

L12-15 I would delete this intrusive sentence. The point it makes is entirely valid, but the following sentence does not apply to it and it is not of fundamental importance.

Zemp et al.: done

P794-797

Section 2.2 This comment applies in fact to the whole of section 2.
This paper will be a long and rather tough read even for those with an immediate technical interest in the subject, so I wonder whether it could be made more accessible by reducing the amount of technical material which is simply taken from earlier sources (principally the *Glossary*, although that source was mainly a codification of still earlier work). Some of the material is essential for the setting out of notation and basic ideas, but all the same I would urge the authors to try to reduce section 2 by about a third to a half.

Zemp et al.: We shortened Sections 2, 3, and 4. See also general comments section.

P800

L12 Standard error: this should be defined carefully here, as the standard deviation divided by the square root of the number of independent items of information in the sample. Problems arise later from assuming that this number is equal to the number of items in the sample.

Zemp et al.: done and following sentences clarified.

P802

L16-17 This, and some of the material from the previous paragraph, ought to be in section 2.2.

Zemp et al.: We assume that this comment refers to issues related to interal and basal mass balance components. Section 2.3 cover with the generic differences between the glaciological and the geodetic mass balance whereas Sections 2.1 and 2.2. cover the glaciological (=surface components only) and geodetic (=surface, internal & basal) methods, respectively. For reasons of clarity, we prefer keeping these three sections strictly seperated.

P803
L2 "homogenization" should be defined here (or perhaps earlier at P792 L19).

Zemp et al.: We added explaining text to Section 2.1 (Terminology and components..) and to Section 3 (Conceptual framework for…). See also general comments section.

P805
General Review eqs. 11 to 15 to determine whether all are needed, given that the later analysis is all in terms of cumulative balances.
L16 This is not correct. The number of degrees of freedom in the geodetic measurement is exactly one for present purposes, and the division should be by the number of years (not the root). The quantity $1/N$ is simply a unit conversion factor. The error in a 16-year geodetic balance, when expressed as a rate per year, is very different when divided by 4 instead of 16.

Zemp et al.: We agree and correct Eq. 15. Consequently, we delete Eq. 22 and adjust Eqs. 26 and 27. Changing Eq. 15 does only impact annual geodetic uncertainties calculated from their estimation/model over the period of records (in Section 5.3, Figs. 2, 3, 5, Appendix C) but not does not impact results over the entire period of records. And it does not affect any of the conclusions as the comparions test are provided on cumulated balances. As a consequence of this correction, we decide keeping Eqs. 11 to 15 and add a comment to Eq. 15. This will help readers to reproduce our calculations. See also general comments section.

P807
L2-3 The common habit of comparing confidence regions to see if they "overlap" is not strictly correct. If the uncertainties of the two balances are normally distributed then the probability of their being "different" (drawn from different populations) depends on the difference between them and on their variances. A Student's-t test is needed, and ideally also an F test for equality of the variances.

Zemp et al.: We change "overlap" to "are large enough" in this general sentence, the calculation framework and all details being given in the following paragraph. Stundent's-t and F tests are suited for comparisons of mean and variance, respectively. A student test would be suited for a comparison on empirical estimates of mean (measurement values in our case), the real variance being unknow (just an empirical estimate being available). Here, our estimates of measurement uncertainties result from a physical approach and we consider sigma as known (or estimated with a degree of freedom higher than 40). Therefore our test is like a Studend test of infinite degree of freedom in which the student function tends towards a reduced normal law. Moreover, we assume the results to be normally distributed (which is not a mandatory condition to perform this test). The measurement difference delta.PoR (Eq. 20) is therefore expected to follow a normal law with a variance sigma.common.Por (Eq. 21). As we test the null hypotheses, the reduced variable discrepancy delta (Eq. 23) follows a centred gaussian of unit variance. Acceptance of H0 is regarded as a comparison with this unit variance.

L23 Explain briefly where 1.96, and 1.64 below, come from.

Zemp et al.: The values +/-1.96 and +/-1.64 indicate a confidence level of 95% and 90%, respectively, in a standard (zero-mean, unit-variance) normal distribution and corresponds to the often used 2xsigma and 1xsigma uncertainties) as explained in Section 3.4, lines 807-808, and illustrated in Figure 4.

P810
L3 What is an "interim" calibration? This adjective should probably be deleted.

Zemp et al.: done.

L5 "a significant difference". Make things easier for the reader by using "bias" only as a "within-method" descriptor and using "difference" only for between-method differences.

Zemp et al.: done.

P816
L2-3 The non-additivity of the reanalysis procedure explains the very odd result that it might advise you to correct some parts of a long glaciological record but to refrain from correcting the entire record. It is essential that the paper offer some clear guidance about what to do when this happens. As it stands, the paper implies that there are two decades of the Storglaciären record for which there are two "equal best" approximations to the truth.

Zemp et al.: In reality, geodetic surveys do not completely cover the full time period of the glaciological observations and are subject to systematic and random errors themselves. A per default calibration to geodetic balances does, hence, not make sense. See also general comments section. We added an example on how to deal with this ambiguity in Sections 5.1 and 5.3.

P818
L5 The Cogley and Adams estimate was –200, not –300 (plus an unknown additional amount to allow for neglect of internal accumulation).

Zemp et al.: In the manuscript, the random uncertainty of measurements at point location of 300 refers to Lliboutry (1974); for Cogley and Adams (1998) it is: 200.

L9-10 Why do these numbers not add up, even when 325 and 140 are added "in quadrature"?

Zemp et al.: This is because the values do not refer to exactly the same sample. We adjusted this paragraph to avoid this kind of confusion.

Stylistic Comments

Zemp et al.: all comments considered and the majority of corresponding paragraphs revised.

(C) Feedback by R. Hock (Referee)

Major concerns

1.) My main comment concerns the recommended procedure that in case of discrepancies between glaciological and geodetic results, the glaciological results per default are the ones to be corrected to match the geodetic results (e.g. line 9, page 791; page 812, line 19). I think this general statement/recommendation is not justified. It includes the implicit assumption that the geodetic balance is the 'truth' to which the glaciological balance needs to be adjusted. This may be the case in many of the examples in the literature but in general it is conceivable that a geodetic balance is more off than a glaciological balance. Whether or not the glaciological balance should be corrected to the geodetic balance (or vice versa) has to depend on the uncertainties of each of the balances.

Zemp et al.: We agree that the validation of the two balances must consider the uncertainties of both the glaciological and the geodetic balance. This is exactly why we are using the common variance (Eq. 21) to compare the difference between the two methods. The outcome of the test is just to detect a significant difference between the both mesurements within their scattering. Not to assign the biais to one of the two balances. A fraction of the biais might be shared out by both balances, and the "true" value a third unknown value.

To go further from the above statistics, we need therefore metrology and physical considerations. We aggree with R. Hock's comment that the occurrence and the extent of systematic error in the glaciological balance are not larger than in the geodetic. But the glaciological measurements are repeated every year to cover a given period of record of N years, so that, if present, they might accumulate linearly from year to year, while the geodetic is performed by one differenciation (two measurments) over the same period of N years. Considering physics, the annual glaciological measurement is defined from the previous year surface, while the geodetic is defined relative to a fix (ground) surface. The surface integration of point balances is also a potential source of systematic error in the glaciological method. In the geodetic, the volume to water equivalent conversion is the main source of bias. The geodetic balance seams therefore reasonably less affected by systematic error accumulation than the glaciological and, among both measurements, provides the best unbiased estimation of the balance over a long period.

Adjusting the cumulative glaciological to the geodetic balance allows maintaining the relative annual variability of the glaciological balance and – at the same time – adjusting to the absolute multi-annual value of the geodetic method. See also general comments section. We do not see the benefit or reasoning for adjusting the geodetic to the glaciological series.

Despite all physical considerations and in line with R. Hock comment, we admit that adjusting the cumulative glaciological to the geodetic is a convention and remains somewhat arbitrary as it implies that it is the best estimate of the balance.

2.) Second, the thresholds given on page 822, lines 25ff for the decision whether or not a reanalysis is necessary are problematic; and I don't think they should be included here. An absolute number cannot be given because the threshold bias must depend on the uncertainties in both balances. The absolute difference has to be compared to the uncertainties in both balances and also to the absolute balances. For example, a re-analysis of a glacier is not necessarily justified if the uncertainties in the geodetic balance are larger than the differences between methods no matter whether or not they exceed the given thresholds.

Zemp et al.: The values given in recommendation III of Section 5.4 are only a first indication for a reanalysing need. Note that the need for reanalysing a balance series does not necessarily result in calibration. We change these three thresholds into a more general statement and one value (derived from our dataset). Also, we add comments in abstract and conclusions to avoid the potential misunderstanding between "reanalyse" and "calibrate". As explained above, we do consider the uncertainty of both methods.

General comments

3.) The paper is somewhat difficult to read mostly because it is too long; the authors should try to shorten it in order to make it more readable to a broader audience. The paper combines a review-type overview with a framework/recommendations. Since the former is incomplete and not fully developed anyway, and the latter is the more innovative and main point of the paper, I would suggest that the review-type information is cut to a minimum and repetitions avoided. To shorten, I also suggest that
a) trivial equations (like for example, how an arithmetic mean is computed or a deviation from the mean) are deleted
b) the small summaries after each subchapter are deleted (they are not very informative and may rather be integrated into the conclusions).
c) page 813: add the half page of references to the Table and remove here.

Zemp et al.: (a) we deleted Eq. 22 but keep the others since even the simple equations allow explaining something very clearly and do not use a lot of space. (b) done. (c) done. In addition, we shortened Sections 2 and 3. See also general comments section.

4.) Due to the extensive review-type parts of the paper, the essential part, i.e. the framework and recommendations get a bit lost. Chapter 3 could benefit from some re-writing. Instead of starting with review-type information which seems somewhat off-topic I would suggest that the authors give a short overview of their framework how to reanalyze rather than more text-book like descriptions about stakes and pits etc. This would also help to understand the following subchapters. Chapter 3.3. starts with 'The aim of this

third step is …' but in fact I missed what the first 2 steps are. Why are 'Observations' part of the steps in the re-analysis? They are the input, but what are the actual steps you do? Maybe a flow diagram of the steps of re-analysis would help. Figure 1 is unfortunately not very helpful in that respect. Basically the essence of the 'new' framework gets lost in excessive review-type information and extreme detail (e.g. page 17). What I miss is an overview (text and figure) of the main steps of the re-analysis.

Zemp et al.: Chapter 3 does in fact start with an overview text and figure (Fig. 1) explaining the six reanalysing steps. We added short definitons for the main reanalysing steps, delete the section summaries in Chapters 2 and 3, and add a paragraph on the essence of the new reanalysing framework to the conclusions.

5.) The paper deals exclusively with usually smaller, land-terminating glaciers that have a long-term glaciological program. Hence it is rather limited in scope, considering that a huge fraction of the world's mountain glaciers are marine- or lake-terminating glaciers and many other methods have emerged to quantify mountain glacier mass changes on various scales. This is no problem for the paper but the authors should clearly state this limitation and make clear that this is about land-terminating glaciers, generally small in size, with a glaciological mass balance program.

Zemp et al.: The corresponding statement in Section 2.2 was extended based on the comment by Graham Cogley: "Note that floating glacier tongues and ice shelves are not considered here, because their mass balance is often dominated by frontal and basal terms that are not addressed by the glaciological method (cf., Kaser et al., 2003)."

6.) Terminology: The authors claim to follow Cogley et al., 2011, however, they do not in a few of instances:
a) The term 'specific' is used incorrectly. It is not the glacier-wide balance but the balance in units per m2 (e.g. page 794 and 795, line 16).
b) conventional/reference balances (see below)

Zemp et al.: a) done. b) done.

7.) Those familiar with calibrating and validating models may be somewhat confused by the use of the terms calibration and validation in this context (although their usage here is technically correct). It may be useful to define what you mean; and / or use re-analyze instead of calibrate. Currently the paper uses both terms for the same process, and it may be better to stick consistently to 're-analyze' as also indicated in the title.

Zemp et al.: We are not using 'reanalysing' and 'calibrating' for the same process. See specific and general comments section. Text clarified based on recommendations by Graham Cogley.

8.) Acronyms should be avoided as much as possible. PoR is not necessary and only makes the paper (even harder than it is anyway) to read. In many cases it can just be omitted or replaced by balance period or just period.

Zemp et al.: We replaced PoR by 'balance period' in most cases but introduced and use the acronym when related to equations or tables.

Detailed comments

1.) page 793, line 11: Delete 'processes' → 'sum of all components of accumulation and ablation …"
line 13 is only true for a land-terminating glacier. This needs to be included.

Zemp et al.: done.

2.) Page 794: non-inclusion of floating glacier tongues or ice shelfs are mentioned but not lake- or marine terminating glaciers (which often have a grounded tongue).

Zemp et al.: done.

3.) Page 794, line 10. Replace 'inaccessible' by 'unmeasured'. Often, regions are accessible but still not measured and one needs to extrapolate.

Zemp et al.: done.

4.) Page 795, line 1: add reference(s)

Zemp et al.: done.

5.) Page 795, line 3: conventional balance: add a definition or at least a reference since not everybody is familiar with this concept

Zemp et al.: term deleted here and explained later at Eq. 4

6.) Page 795, line 14: What about internal accumulation?

Zemp et al.: The components of the internal and basal balance are discussed in Section 2.4 (Generic differences between glaciological and geodetic balance).

7.) Page 796: the discussion about conventional and reference surface balance is incomplete and not correct: the reference-surface balance does not only use a constant area but also a constant hypsometry, i.e. measured point balances first need to be extrapolated to the previous elevation before integrating over the previous ice extent.
For the sake of shortening this paper, the entire discussion about these 2 types of balances

may be deleted (or shortened), since it does not really relate to the re-analysis framework discussed here.

Zemp et al.: The issue of conventional and reference-surface balance is common to all mass balance programs and an essential step of the homogenization. Hence, we do not delete this paragraph but clarified it based on the comments by Regine Hock and Graham Cogley.

8.) Page 797: page 3: add 'land-terminating glaciers'

Zemp et al.: done.

9.) Page 797 last line: something is wrong in the English

Zemp et al.: corrected based on comments by Graham Cogley.

10.)Page 798: 'the volume change is converted into a specific mass change by:

Zemp et al.: we changed to ‚specific mass balance' but keep the information about the unit.

11.)Page 798 line 22: why is the assumption of linear change necessary. Don't think that is true; it rather implies an assumption about the geometry of the bed topography.

Zemp et al.: We agree that the assumption of a linear area change through time might be wrong. However, this is the most common approach for geodetic volume change assessments and based on the observations available (i.e., S0, S1, dV). This two points together is our reasoning for explicitely formulating the basic assumption behind Eq. 9.

12.)Page 799: There is excessive detail and emphasis on sighting and plotting, (and too much detail also on page 800), however, I lack a more detailed mentioning and discussion about density assumptions in converting volume changes into mass changes, since this is an aspect with hugely diverging assumptions in the literature.

Zemp et al.: The issue about density conversion is discussed in Section 2.4 (Generic differences between glaciological and geodetic balance).

13.)Page 801, line 23: replace 'surface' by 'climatic-basal' mass balance

Zemp et al.: We replace ‚surface' by ‚surface, internal, and basal balance'. The term 'climatic-basal balance' is misleading since it indicates that surface and internal balances are climatic components and that the basal balance is completely independent from climatic influences.

14.)Page 802-803: lots of repetition of problems that were mentioned earlier. Can be shortened a lot and better be replaced by a brief overview of the main steps of the framework.

Zemp et al.: see comments above about shortining the paper.

15.)Page 806: what do you mean by validation?

Zemp et al.: We added short definitions for key terms in Chapter 3.

16.)Page 807, line 14: delete 'the PoR covering' (can be done in most cases where PoR pops up).

Zemp et al.: done.

17.)Page 811, line 9: replace 'calibrated' by 'reanalyzed' (also elsewhere).

Zemp et al.: Not replaced because 'calibrated' and 'reanalyzed' have two different meanings. We clarified this in the Chapter 3.

18.)Page 811, line 15: This seems arbitrary. What is the physical rationale that any bias would exclusively come from errors in the summer balance rather than the winter balance. Ablation tends to be far more homogenous (and simply elevation dependent) due to the large air temperature dependency, and therefore can relatively well be captured by a few measurements. However, snow accumulation varies greatly within tens of meters and rarely are sufficient probings available to capture the spatial distribution. Hence, one would expect the opposite.

Zemp et al.: Yes but summer balance is usually not measured. See specific and general comments sections.

19.)Page 812, line 11, 16. Here both are used: re-analyzed and calibration. This is confusing.

Zemp et al.: Both terms are used in their specific meaning (as defined and explained in Chapter 3). The context is given in line 15.

20.)Chapter 5.4 is weak. It mixes text-book like statements about basic theory (like VII) with recommendations; hence the (more important) recommendations are diluted. Make this chapter more concise and focus on the recommendations and their rationale.

Zemp et al.: We agree that some statements are basic theory but – unfortunately – still the cause of misconceptions in many studies. Hence, we believe it is important to (re-)emphasize these basic issues.

21.)Conclusions are too general and should be more concise. Maybe reiterate your main points about the framework.

Zemp et al.: done.

22.)Same symbol is used for density and reduced discrepancy

Zemp et al.: No, we use 'rho' for density and 'delta' for reduced discrepancy.