Review of Magnusson et. Al.

Overview

Magnusson et. al. present a very detailed and careful study of glacier volume changes and associated geodetic mass balances of the northernmost icecap in Iceland. The authors generate a sequence of DEMs, starting from 1946, using the highly accurate LIDAR as reference and control. The authors then undertake a particularly strenuous task of modelling the errors using geostatistical kriging methods. From this experiment, the authors conclude that the majority of other geodetic mass balance studies use error estimates that are far too conservative, a conclusion that has been known (suggested) before, but not verified. Moreover, the study also makes seasonal corrections to the geodetic mass balances in order to ensure proper comparisons to in-situ mass balances, and concludes that this is a very important correction, especially in regions of high mass turnover, like Iceland. In summary, this is a very nice, detailed study providing a long history of glacier volume change and a thorough methodological discussion around the errors associated with such measurements. While I certainly can recommend this work for publication in the Cryosphere, there are a number of points that need consideration, and I hope these points will help improve this manuscript into an important reference for the future.

The manuscript is rather long and at some points a little difficult to follow from all the details. I can understand that it was not easy to get all the specific details into a logical framework, but I think this may need some re-consideration to determine which information is important to keep, and which may be thrown away. In addition, the application of the entire mass balance model through the entire time series (Fig 10) is highly interesting, but may rather be suitable for a separate publication. There is not enough space in this manuscript to discuss these results, and why the mass balance model may or may not fit to the geodetic mass balance measurements. I think it is fine to use the mass balance model to estimate your seasonal corrections, but consider leaving it at that.

In terms of the error analysis and description, there are a number of points that need clarification and consideration. First, I think there is an important reference missing: Zemp et. al. 2013 http://www.the-cryosphere.net/7/1227/2013/. There needs to be a clear separation between the types of errors one is discussing, in particular the difference between bias and random errors (accuracy and precision). As far as I know, Rolstad et. al. (2009) quantified the influence of the random error component over the glacier area at question. I am not sure whether your simulations are for the total error, including both systematic and random components but I believe you have modelled a potential systematic error over the glacier, and then correct for it? If this is the case, then the two studies may not be as easily comparable, but certainly should be discussed. Furthermore, more detailed description of “Nscore”ing needs to be made. It is ok to reference to others, but it will allow the reader to understand more directly what you are doing. I am particularly curious about the influence of transforming the data into a normal distribution for calculating the errors, and then whether this assumption then allows for more realistic, representative error estimates for DEM differencing, or if alternatively, provides unnecessarily small error estimates since the inherent error distributions are not normally distributed. I think discussion around this assumption should become a more significant part of this manuscript, and certainly may still be open for discussion.

Minor remarks:

Pg4738 L13: “The Lidar DEM…”
L19: “manual editing of the”
Pg4739 L7- Which year was the monoscopic coverage?
L8 – I don’t understand how you can use a modern DEM to orthorectify a historic imagery and actually reveal accurately front positions? Since the front of the glacier is where the most change is happening, here is where one would expect the worst results from orthoprojection of a historic image on a modern DEM…
Pg4740 L26-27 – So here, you are extracting a 2x2m elevation at a sampling distance of 20m. This inherently causes resolution problems for your comparisons related to roughness, curvature and ultimately resolution. To be completely precise, one should average the LIDAR with a 10x10 pixel window around each 20m grid sampling…
Pg4741 L21: I agree that it may be highly unlikely, but it is very easy to test by making a scatterplot of [elevation difference divided by slope tangent] by [aspect]…
P4742 L7: This step needs to be explained in more detail. How does the transform affect the data, what kind of transform is it… It would also be beneficial to defend this choice or provide a background on why you chose this direction.
L11: “…nscored input data, in which the semivariogram…”
P4743 L15-20: I don’t follow the argumentation about the 1946 DEM. According to Fig 4c, I see that the range for low contrast surfaces is closer to 1600 m, but this for changes between 1946 and the LIDAR. In either way, this argumentation needs to be spelled out a bit clearer for the reader.
L24-28: Fine to filter by slope but I imagine there is more a problem about resolution between your reference and slave DEMs, and thus curvature and roughness will have a large impact, since you compare a 2m point elevation to a 20m pixel elevation.
P4744 L6: a 500 by 500 m window is rather large and a bit excessive. This creates unnaturally smoothed values. Why not use 100 by 100 m since you end up sampling at this distance in the end (L26)??
P4745 L16-20: Could the spatially varying mean error really be used for correction? It is only produced from a simulation, and not really based on any truth over the glacier. What strikes me is the trend (in Fig 4d) from lower glacier to upper glacier, this type of bias adjustment, if not certain, would change the data significantly, and probably also the interpretation. Also, in the end, what is the value of the mean of the derived probability function? And where is this derived, only over the glacier, or over the entre sample?
L24-27: It is still not clear what the difference to Rolstad et. al. (2009) is? As far as I see it, is the transform to normal distribution is different, but it doesn’t seem like the range changes so much (between fig 4a and b). Can you clarify more specifically why there is a difference, and what that difference is.
P4748 L19: why “or/and” ? Shouldn’t it always be integrated over the largest area?
P4752: L13: What percent volume change is the correction for each time period. This will just give the reader a better idea on the significance of these corrections.
P4753 L7: Actually, the variables are not independent at all, since to calculate the volume change, you need to have the area…. But, maybe rather the results are not so sensitive to failure of the assumption….

Table 3: place abreviations of the parameters (table column headers) into the caption at the proper description. This will make it easier for the reader to interpret the table.

Figure4a: Why not plot the spherical variogram model here? Seems the range will be the same as when N-scoring the data….
Figure 8: Very interesting plot! Would be worth to describe this plot more specifically in the text, and suggest reasons for the linear scaling between your and Rolstads approach (i.e. Normal distribution vs. Original Distribution).
Fig. 9: Hard to see the different lines that represent the estimates and the errors… Consider changing colors
Fig 10. Consider removing this figure… Doesn't add so much to the entire article, and can probably be an article in itself…