

Response to Reviewers, TCD Manuscript
“Modelling glacier change in the Everest region, Nepal Himalaya”
Shea et al.

29 January 2015

All authors would like to sincerely thank Ben Marzeion, Ann Rowan, and Franco Salerno for their constructive and insightful reviews. We have attempted to address all their comments below, and we are confident that the revised manuscript will be substantially strengthened.

We would also like to take this opportunity to comment that this modeling exercise was initially intended to be a sensitivity study. How might glaciers in the Everest region respond to increases in temperature, or changes in precipitation? Our approach represented the first integration of distributed models of glacier mass balance and redistribution with field observations. And while the model itself is a simplification of the complex physical processes that determine how glaciers respond to climate change, it is a necessary simplification.

There are too many unquantifiable uncertainties to make true “predictions” of future glacier change in the region. Based on our responses below, we can identify three main sources of uncertainty: 1) parametric; 2) structural; and 3) climate model inputs. A full Monte-Carlo calibration procedure might lead to reduced uncertainty in the model parameters, but is currently not feasible for a distributed model running at such a fine spatial and temporal resolution. The model itself contains significant structural uncertainties, from the lack of true physics governing ice flow to the unknown (unknowable?) initial ice volume conditions. Finally, there is uncertainty in the input climate fields that arises from both our downscaling approach and the CMIP5 simulations. We attempted to be cautious about the results in our original manuscript, and will endeavor to fully explain these uncertainties in our revised version.

Our initial approach was to use the model as a tool to explore possible future glacier change in the Everest region, and we described these as glacier change scenarios. That said, our results should not be considered as projections or predictions of future glacier change. Our model suggests that glaciers in the region will be sensitive to future temperature increases. And given the projected rate of temperature increases we are confident that the glaciers in the Everest region will be melting at a greater rate in the 21st century. Their final state at the end of the century though is still an open question, though our results will hopefully provide some guidance.

Finally, we look forward to future research, field data, and modeling developments that can lead to improved representation of the glacier dynamics, improved climate observations and downscaled climate fields and estimates of high-altitude precipitation, a more physically-based melt modeling approach, and more confidence in the future behavior of the monsoon, which can have great impacts on the mass balance of glaciers in the region. We are unable to deal with these issues in this paper, but we have made a start. At this stage, due to computing time, a lack of knowledge about debris-covered dynamics or data for validation, we cannot address in details some of the reviewers’ comments but we are confident that it will be possible in the future.

Based on the comments of the reviewers, our revised manuscript will contain the following main changes:

- 1) Improved description and discussion of the initialization process and uncertainty

- 2) The use of full T+P anomalies for future glacier change scenarios
- 3) Discussion of structural, parametric, and input uncertainties

We have addressed all general and specific comments below. The original reviewers comment is given in regular font, and our response is given directly below in **bold**. Text added to the manuscript is also given in ***bold italic*** font.

Review #1: Ben Marzeion

Shea and colleagues present a model of glacier mass balance and ice redistribution and apply it to the glaciers of the Everest region. The manuscript is well organized, and the results are generally well presented. The research is clearly relevant, and the authors are following a very promising approach. However, I find three major issues that need to be addressed to ensure that the conclusions are actually supported by the results. Additionally, I have quite a few smaller suggestions that probably can be taken into account with relatively little work.

General Comments:

- I do not understand how the initialization run is working (P5386, L13-19): Assuming that the applied T anomaly and the resulting mass balances are correct, how do you get the ice thicknesses and extent in 1961? Do you apply the mass balances until extents (and ice thicknesses) reach an equilibrium? How long does this take? Or if you do something else, please describe. I think this is not a trivial problem, because the lagged response of glacier extents to mass balance changes will affect the following calibration runs; i.e. how you create the state in 1961 will impact how the glaciers behave in the following decades, additionally to the different parameters you use. Of course, also the parameter set you use for the initialization run will play a role (e.g., you will see changes in area even if you keep the forcing constant, but change R). Or you could imagine a situation where the glacier extent in the calibration run is not yet adjusted to the ice thickness changes, and the resulting changes in the following calibration run are a mixture of adjustment to the new parameter set, and the equilibration between extents and ice thickness.

Response: We agree that the model initialization is not a trivial problem! Unfortunately, there are no datasets available to constrain the initial (1961) ice volume conditions, and comparisons with earlier glacier inventories (e.g. see our response to Franco Salerno) introduce additional errors. Our approach was to estimate “current” ice thicknesses with the 2007 glacier distribution using the SRTM-derived surface slopes and the equilibrium ice thickness (Eq. 8). From this configuration, ice volumes and extents in 1961 were estimated by executing a historical run forced by the 1961 – 2007 time series of precipitation and temperature and an applied temperature trend of -0.025 K/yr to correct for the observed warming. We assume that the model initialization period of 47 years represents climate conditions prior to 1961. However, it is possible that there are significant uncertainties in our estimates of initial (1961) thicknesses and extents, given the forcings and parameter set used, and the lag in glacier geometry responses to climate forcings. These uncertainties are likely to be small for smaller glaciers with quicker response times, but larger for large debris-covered glaciers with thick termini.

Important point of discussion: the parameter set of the initialization run is different than for our “best” set (run5). This is a classic “chicken-egg” problem. The calibration depends on the initial ice conditions and the initial ice condition determines the calibration. One possible option would be to run all parameter sets for an initialization period of suitable length before running the 1961-2007 run.

However, we feel that this would be beyond the scope of the current study, and we recommend in the discussion that it be considered for future research with additional parameter sets (see response to your point 2 below).

We have tested the initialization setup by continuing to run the model for two additional 47-year periods. In Figure R1 we compare the modeled ice thicknesses at the end of the first initialization run (as used in this study) and at the end of the third. There are some slight differences in thickness (mean increase of 3.2 m), but the majority of the ice thickness differences are near zero, except at the glacier termini. The initialization itself may be sensitive to the applied temperature trend, and the secular trend of increasing temperatures over the period 1961-2007 may introduce errors as well during the initialization. An alternative option would be to use the 1961-1970 historical climate for the initialization, and to repeat until the glaciers show zero mass change.

We have included Figure R1 in the manuscript, edited the text to clarify our initialization procedure, but have left the initial conditions as described in the original paper:

“To simulate the observed climate in the region prior to 1961, temperatures in the initialization run are decreased by -0.025 C/yr (Shrestha and Aryal, 2011) for a total decrease of -1.2 C over the 47-year initialization period. Mass change at the end of the initialization period is close to zero, indicating that near-equilibrium conditions have been realized, and additional runs of the initialization period yield relatively small changes in glacier volumes. However, it is possible that there are significant uncertainties in our estimates of initial (1961) thicknesses and extents, given the forcings and parameter set used, and the lag in glacier geometry responses to climate forcings.”

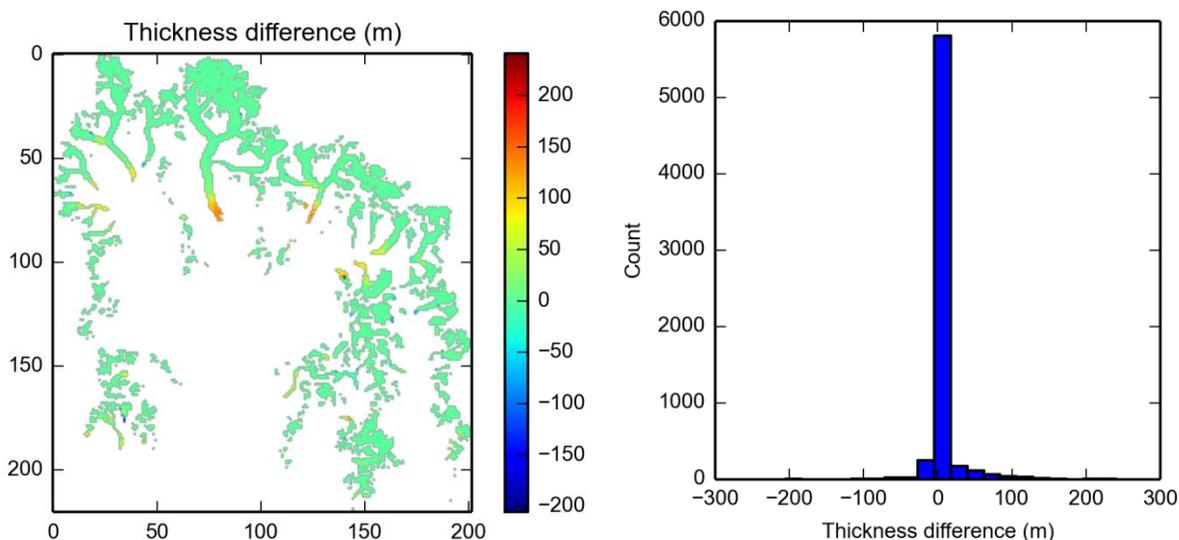


Figure R1. Left: differences in modeled ice thickness between the end of the original initialization run (47 years), and after an additional 94 years of simulation with $dT = -1.2$ C. Right: histogram of differences in modeled ice thicknesses. This figure is included in the revised manuscript.

- The "Monte-Carlo"-type calibration: during the calibration, there are 6 degrees of freedom (i.e., 6 parameter values are being optimized). 20 parameter sets to choose from is then a very low number

(e.g. if you want to cover every combination of parameters with just two values for each parameter, you would need $2^6 = 64$ parameter sets/calibration runs). My question therefore is: what part of the parameter space is covered in the 20 runs, how did you decide on the parameter sets you tried, and how do you make sure that the "best set" is within the part of the parameter space that you cover? Fixing this could be numerically very expensive (e.g., covering the parameter space with 3 parameter choices in each direction leads to the order of 1000 calibration runs) – but would it be possible to do the validation for each of the 20 calibration runs? This could give a good indication on the robustness of the calibration (i.e., one would assume the four runs with the best overall scores in the calibration to also perform well in the validation)?

Response: Initial values for parameters were taken from previously published studies, and the actual values used were chosen at random based on plausible ranges. Table R1 below provides all 20 parameter sets used in the calibration runs. To explore the parameter space, parameter values expressed as standardized Z-scores $[(y_i - \bar{y})/\sigma]$ are shown in Figure R2 below. Some values are up to 2 standard deviations above the mean, but there are no major surprises. To conduct a larger Monte Carlo simulation (100 sets would probably reveal the same patterns as 1000) with the distributed model would be useful to examine the true parameter sensitivity and ensure that our chosen set is the "best", but given the computing requirements, we recommend that this be (a) combined with initialization runs (see response above) and (b) conducted for multiple study sites where suitable calibration data exist.

Table R1: Parameter values used in the calibration runs. This table is included in the revised manuscript.

Run ID	DDFc	DDFd	DDFk	DDFs	R	Hc
1	10.1	2.4	5.7	5.1	965538934	5948
2	9.8	3.7	6.8	4.6	862185519	5974
3	9.2	4.1	8.5	3.6	1326340408	5544
4	8.8	1.7	5.3	5.7	2115148902	6392
5	9.7	4.6	8.6	5.4	1507211339	6268
6	8.9	1.9	6.8	4.3	1757035837	5712
7	9.3	3.6	7.3	6.6	1602852068	5810
8	8.9	2.0	7.0	5.3	1891517886	7175
9	9.3	2.9	8.2	5.7	965461867	6663
10	8.1	3.1	9.0	5.8	1966902971	6339
11	9.3	4.1	7.0	5.1	2119160369	5804
12	10.1	3.3	6.4	4.7	1183544033	5774
13	10.2	2.2	5.7	5.1	2027971886	5960
14	9.3	5.2	6.6	6.4	1642592045	5887
15	8.5	3.2	6.7	3.9	1674708607	5466
16	8.1	4.3	4.2	5.5	1278943171	6877
17	10.2	3.5	5.4	5.6	1687134148	6314
18	10.7	2.0	6.2	5.3	1920883676	6270
19	7.6	2.9	7.2	4.6	2402645369	5586
20	10.8	3.5	6.0	6.4	1885850339	5673

Mean	9.3	3.2	6.7	5.2	1639181469	6072
St.dev	0.87	0.98	1.23	0.80	428282810	459

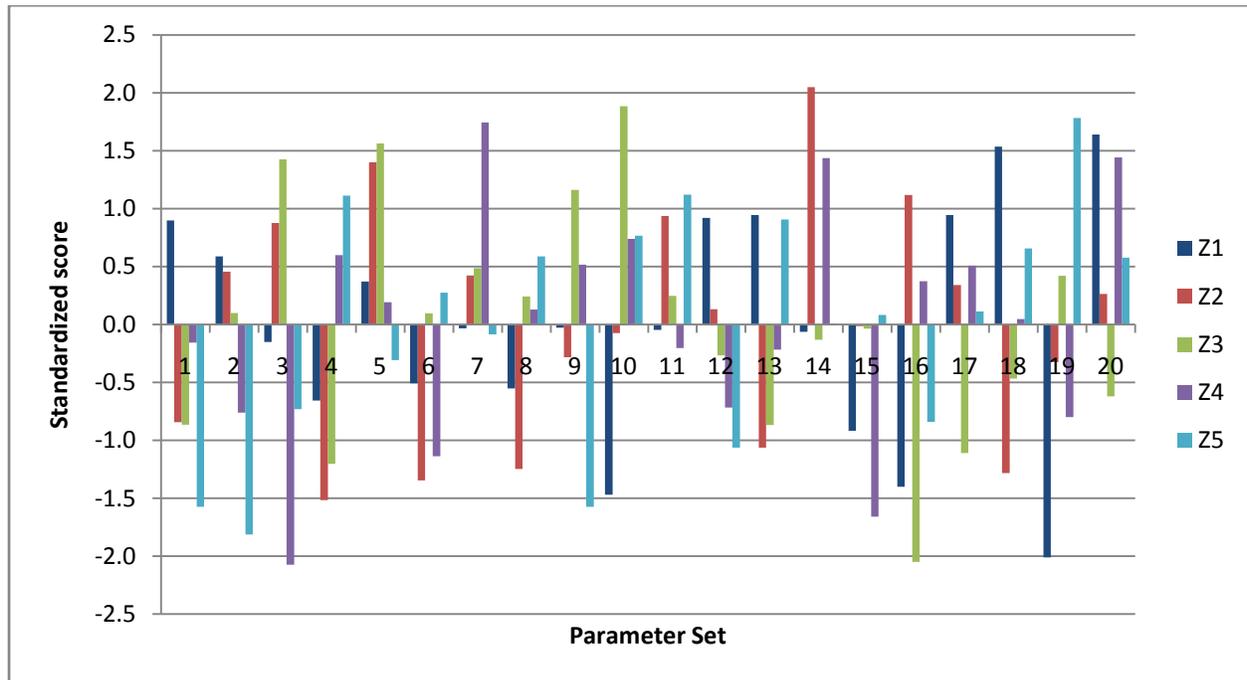


Figure R2: Z-scores for 20 sets of model parameters (Z1 = DDFc, Z2 = DDFd, Z3= DDFk, Z4 = DDFs, Z5 = R, Z6 = Hc). This figure is *not* included in the revised manuscript.

- The rationale behind the application of the CMIP5 scenarios is unclear. Why use linearized anomalies instead of the full anomalies that are readily available? Linearizing symmetrically around the year 2000 will have very significant consequences for the estimated differences between the scenarios (e.g., T change is not at all linear, and the different rates of T change from 1975 to 2035 do not reflect the difference between the RCP4.5 and RCP8.5 scenarios – this probably explains why the two scenarios are practically identical in Table 5, and hardly different in Fig. 14), but also between the ensemble members of each scenario. It should be straight forward to take the full T and P time series, and I strongly recommend to do so.

Response: Our stated goals were to examine the sensitivity of glaciers in the regions to temperature and precipitation changes. Given the uncertainties in the model and the initial conditions, we have been reluctant to do full “projections”, and used the linearized T+P anomalies as guidance as opposed to prescribed T+P scenarios which may only reflect CMIP5 model structure and variability. As requested, we have re-run the future scenarios using the full CMIP5 anomalies from 8 end-members (Figures R3 and R4 below). Our approach extracts decadal T+P anomalies relative to the 1961-1990 historical climatology, and applies them to the daily downscaled climate fields in order to retain natural climate variability. We have modified the scenarios of future volume change (Figure 14 in the manuscript; Figure R5 below), modeled area changes at different elevations (Figure 15; Figure R6 below), and modeled ice depth distributions at different time steps (Figure 16; Figure R7 below). Table 5 has also been revised slightly with the new volume scenarios (Table R2 below).

We do not include the T+P anomalies in the revised manuscript, but would mention here that the precipitation anomalies do not show any significant trends, and vary between 0.4 and 1.8 times the baseline period. Temperature trends are strong in all months in the RCP4.5 and RCP8.5 scenarios, with greater temperature increases projected for the RCP8.5 scenarios. Ensemble mean temperature increases to 2100 are as large as +8C in the late-winter and early-spring (January – April).

Volume change scenarios (Figure R5) are slightly different from those given in the original manuscript (Table R2 below), with increased variability and a greater rate of volume loss by mid-century. Scenarios of volume change for the year 2100 are nearly identical for RCP4.5, and are more negative for RCP8.5. These figures are updated in the abstract and conclusions of the revised manuscript.

The patterns of ice loss (Figures R6 and R7) are similar to those shown in the original manuscript, though the use of the raw T+P anomalies results in stronger mass loss for the RCP8.5 warm/dry scenario.

Table R2: Glacier volume change scenarios, relative to 2014. X = mean, SD = standard deviation. Revised values are in bold, and values in the original manuscript are given here in *italics*. This table is included in the revised manuscript, with the revised values only.

Scenario	X2050 (<i>old</i>)	SD2050 (<i>old</i>)	X2100 (<i>old</i>)	SD2100 (<i>old</i>)
RCP4.5	-39.3 (<i>-45.4</i>)	16.8 (<i>9.0</i>)	-83.7 (<i>-83.2</i>)	11.2 (<i>8.7</i>)
RCP8.5	-52.4 (<i>-51.9</i>)	14.5 (<i>9.8</i>)	-94.7 (<i>-88.7</i>)	4.2 (<i>7.3</i>)

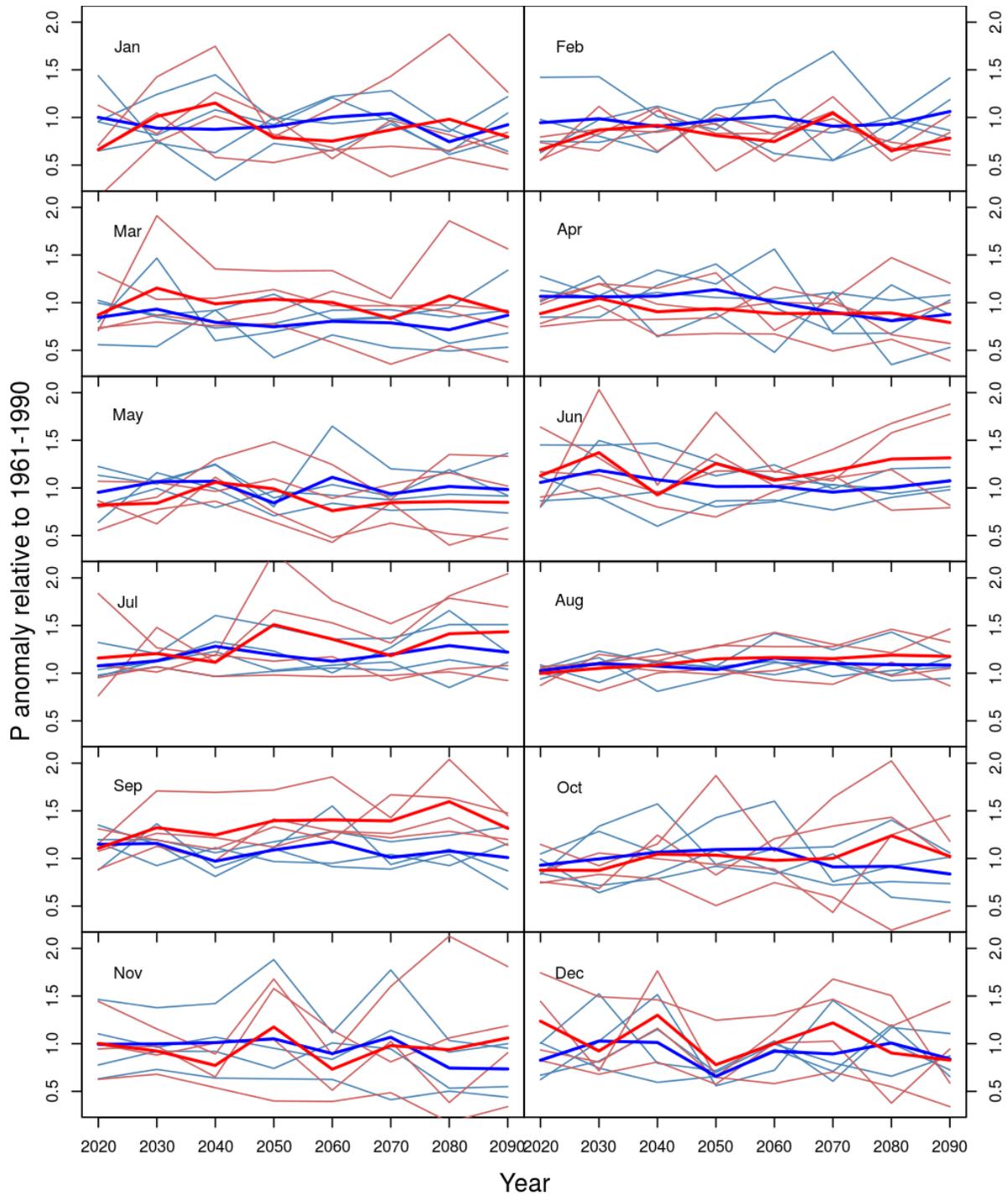


Figure R3: Decadal precipitation anomalies from 8 CMIP5 end-members, relative to 1961-1990 climatology. RCP4.5 scenarios are given in light blue, with composite mean in thick blue, and RCP8.5 scenarios are given in light red, with composite mean in thick red. This figure is *not* included in the revised manuscript.

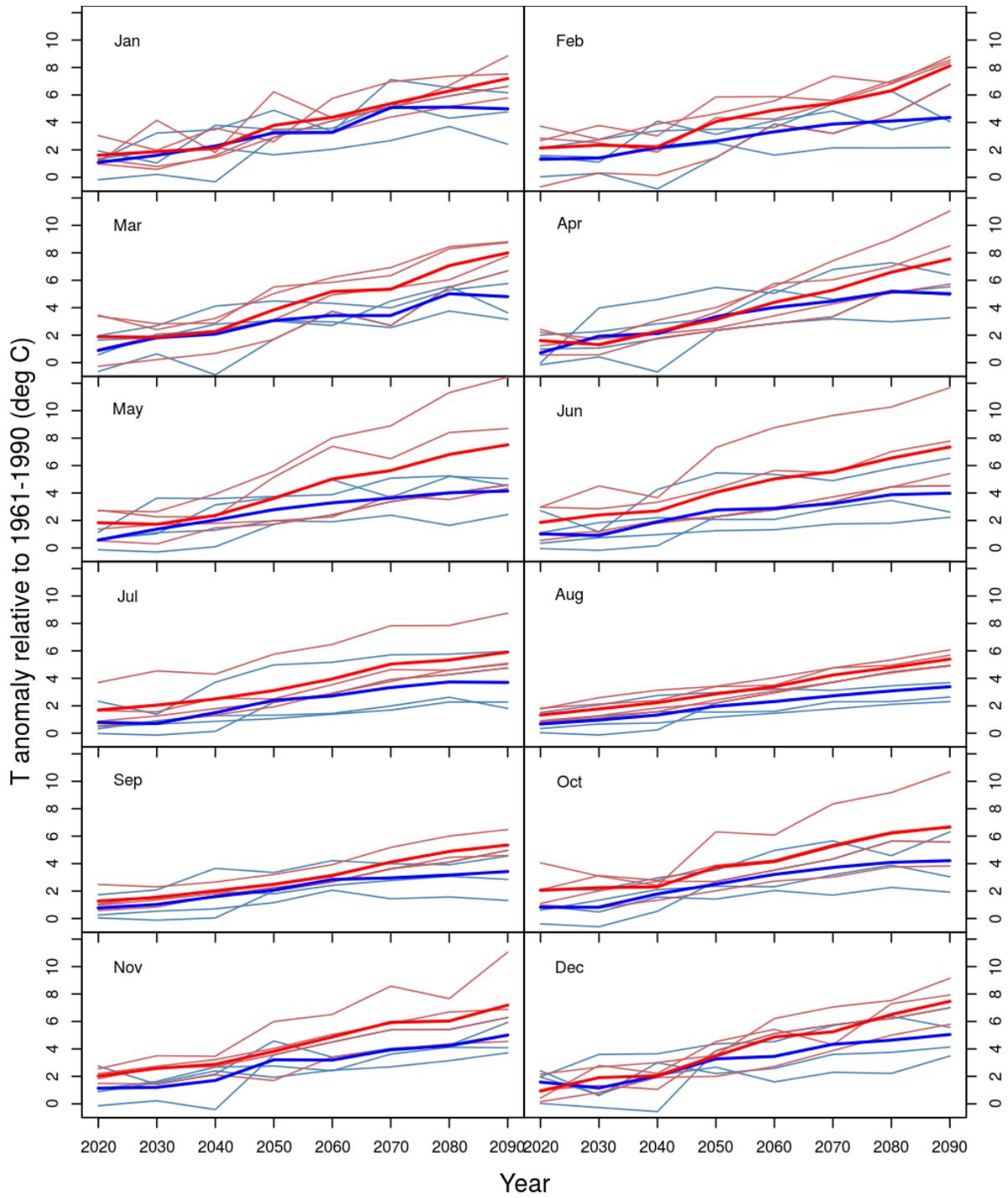


Figure R4: Decadal temperature anomalies for CMIP5 end-members, relative to 1961-1990 climatology. RCP4.5 scenarios are given in light blue, with composite mean in thick blue, and RCP8.5 scenarios are given in light red, with composite mean in thick red. This figure is *not* included in the revised manuscript.

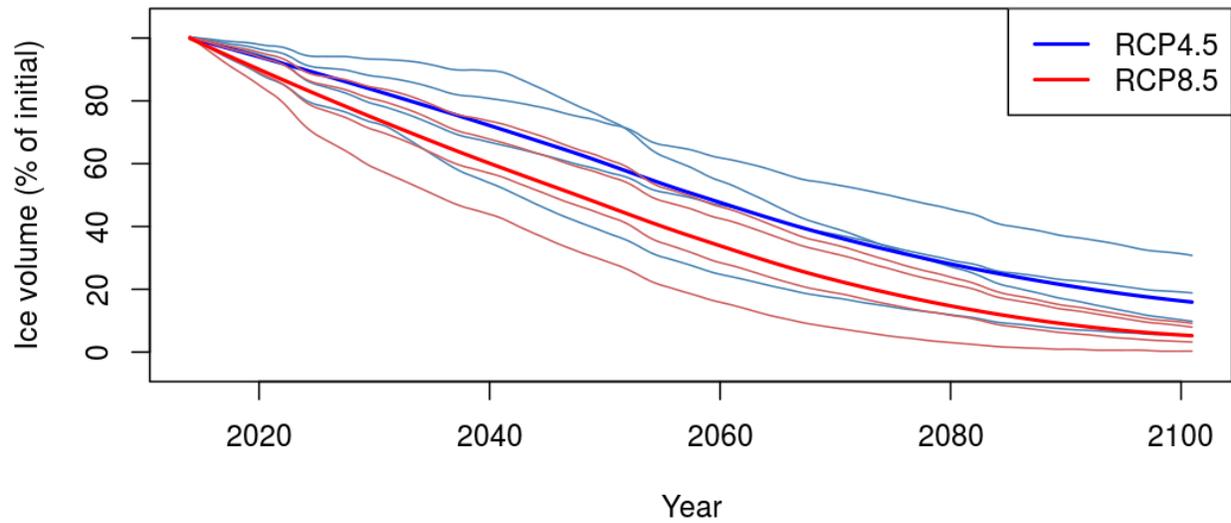


Figure R5: Revised ice volume scenarios for the Dudh Kosi basin, using decadal temperature and precipitation anomalies from eight CMIP5 end-members. RCP4.5 realizations are given as thin blue lines, and the ensemble mean is in thick blue. RCP8.5 realizations and ensemble mean are given in red. All realizations are smoothed with a loess filter (span = 0.05) and scaled to the initial ice volume. This figure is included in the revised manuscript.

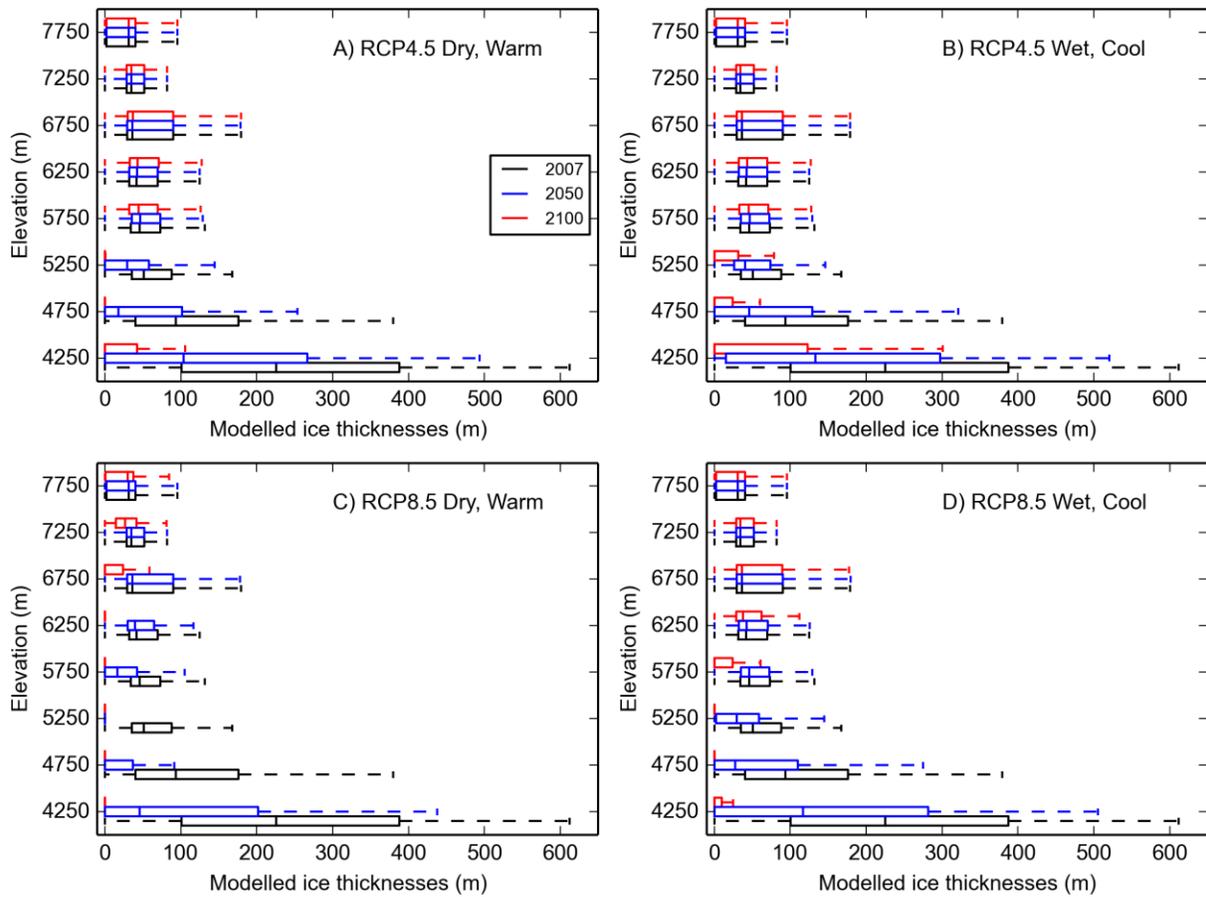


Figure R6: Revised ice thickness distributions for wet/cool or dry/warm RCP4.5 and RCP8.5 scenarios. This figure is included in the revised manuscript. Temperature and precipitation anomalies to 2050 will be given in the legend for each subfigure.

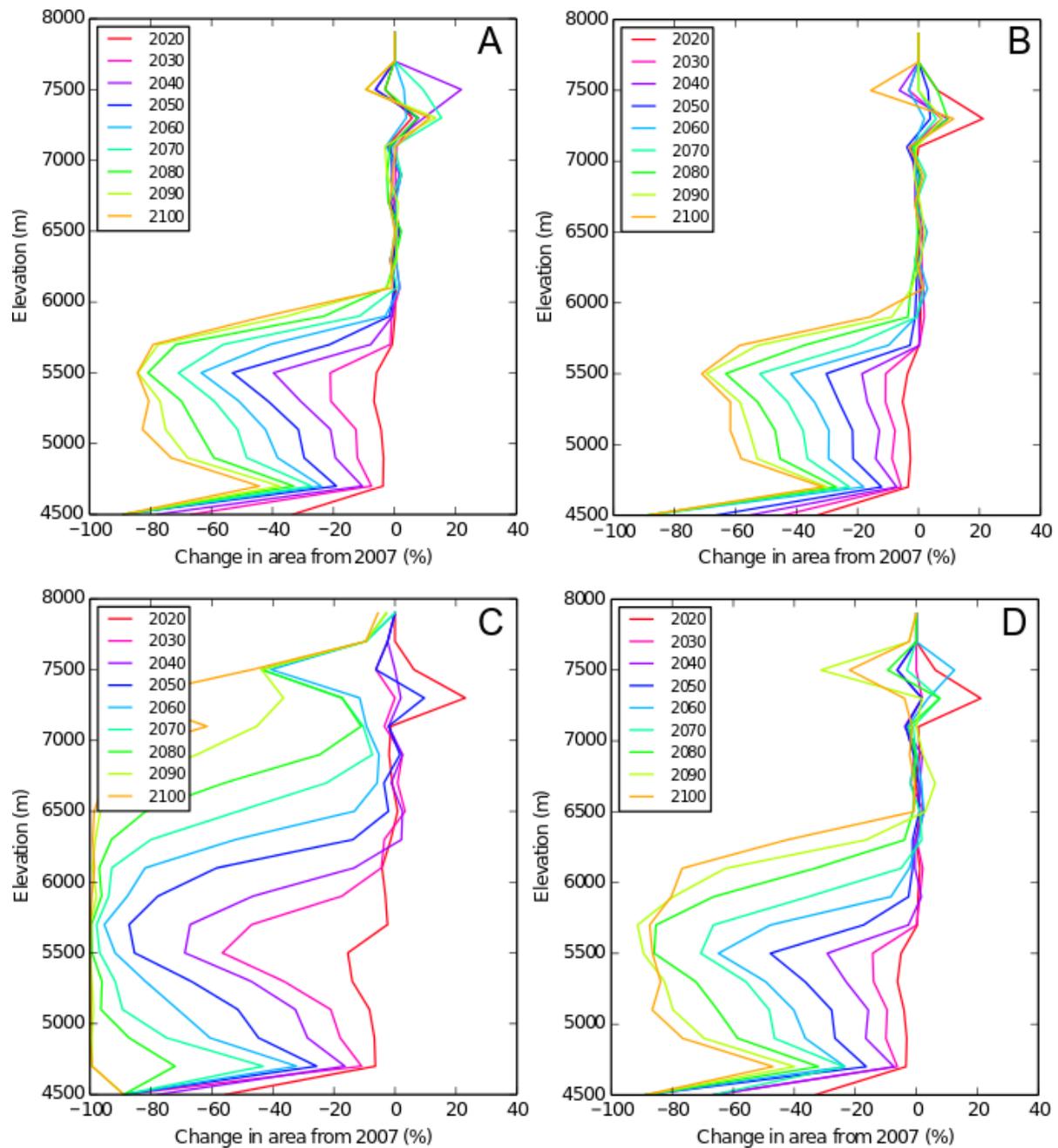


Figure R7: Revised modeled changes in glacier area from 2007 versus elevation for (A) the dry/warm RCP4.5 scenario, (B) the wet/cool RCP4.5 scenario, (C) the dry/warm RCP8.5 scenario, and (D) the wet/cool RCP8.5 scenario. This figure is included in the revised manuscript.

Specific comments:

- In the abstract: clarify the difference between the geodetically derived estimates of net glacier mass change you use for calibration, and the remotely-sensed observations of decadal glacier change you use for validation.

Done.

- Not sure about TC policy, but I would try to avoid "complicated" abbreviations like APHRODITE and EVK2CNR in the abstract – or at least introduce them here as you did with "ELA".

We have removed the complicated abbreviations from the abstract.

- P5376 L16: I'm not sure what you mean by calling the RCP4.5 and RCP8.5 scenarios "end members"?

Here the 'end members' refer to the warm/cold, dry/wet scenarios that were identified as end members by previous studies (Immerzeel et al., 2013). We have clarified the text.

- P5377 L15: Fig. 2 is referred to before Fig. 1.

Thanks, fixed.

- P5379 L10: abbreviation "ELA" already introduced in the abstract.

Thanks, fixed.

- P5379 L24: change "Results from CMIP5 ensembles..." to "Results from the CMIP5 ensemble...". Also: introduce the abbreviation CMIP5 (earliest mentioned in the abstract).

Done.

- P5380 L9: see previous comment

Fixed here also.

- P5381 L10: it is unclear here what is meant by "downscaling ensemble", as the CMIP5 data base to my knowledge does not contain downscaled ensemble members. I think this term "downscaling" is not appropriate here. It could either be used to describe dynamical downscaling as e.g. in the studies of Dimri et al. and Kulkarni et al. (both cited in the manuscript), or it could be statistical downscaling. Instead, you are simply applying the anomalies from the CMIP5 ensemble members.

Yes, we are confusing our terminology here. This now reads "We then develop first-order scenarios of future glacier change in the Everest region with calibrated model parameters and a suite of prescribed temperature and precipitation changes from CMIP5 end-members."

- P5381 L14: change "year" to "years".

Changed.

- P5382 L2: 14 x 14 grid cells?

Now reads: "...we extract a 196 (14 x 14) grid cell subset..."

- P5382 L4: are there any significant gaps in the SRTM DEM in the area? If so, how do you treat them?

We use the gap-filled SRTM data (V4), which was produced with a void-filling algorithm documented by Reuter et al. (2007).

- Fig. 3, caption, L3: delete "for".

Deleted.

- Fig. 3, bottom panel: I assume all points are significant? Please indicate.

All temperature/elevation regressions are significant. This is now indicated in the caption.

- Fig. 3, upper and middle panel: add _ to vertical axis.

Done.

- Fig. 4: add _ to vertical axis.

Done.

- P5382 L16: introduce the abbreviation "EVK2CNR" (earliest mentioned in the abstract).

Done.

- Fig. 5: the daily precipitation almost certainly is not normally distributed, and therefore showing +/- 1 standard deviation is not that helpful. Better show percentiles (e.g., 10th, 50th (median) and 90th).

Figure 5 has been modified as suggested, with 10th, 50th (median), and 90th percentiles of precipitation (Figure R8 below)

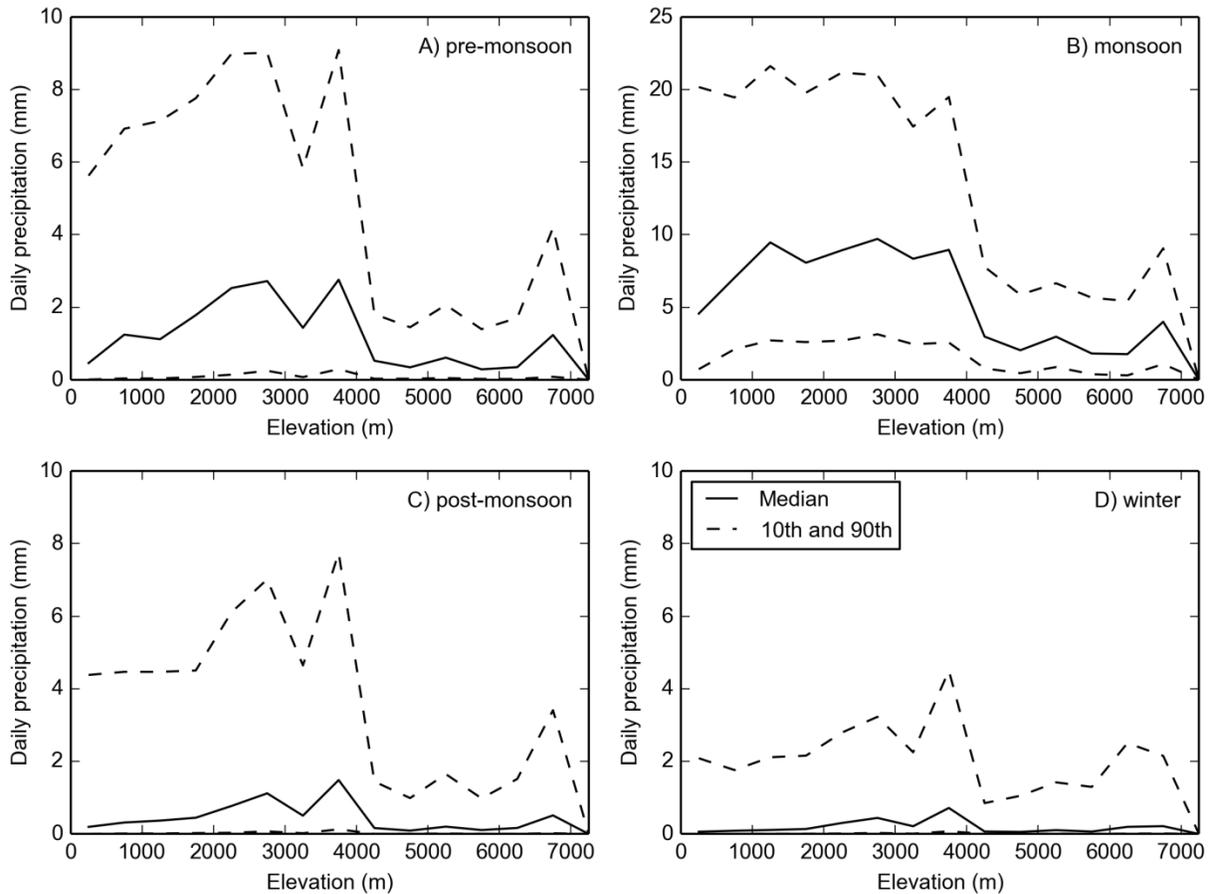


Figure R8: Revised figure of elevation-precipitation relations for pre-monsoon, monsoon, post-monsoon, and winter seasons.

- Fig. 6, caption: change "modelled" to "predicted", both for consistency with the axis label, and since you did not really model precipitation.

Changed.

- P5383 L4: Solid precipitation may (and often does) occur at temperatures slightly above freezing – have you tried threshold values of 1_C? Also, please add the unit here and in Fig. 6.

Yes, in hindsight a higher threshold temperature for solid/liquid precipitation transition, or some estimate of dewpoint temperature (e.g. Marks et al., 2013) might have been more appropriate. While a 1C threshold is physically more realistic, we do not anticipate significant differences in the modeled output. The calibration procedure might result in the selection of a different parameter set that compensates for the higher threshold temperature.

- Eq. 4: perhaps you can explain a bit the rationale behind this - why would you expect highest ablation rates on flat surfaces? Shouldn't this be a function of the day of the year (assuming that you want to capture short wave radiation effects here), i.e. a measure of exposure relative to the sun?

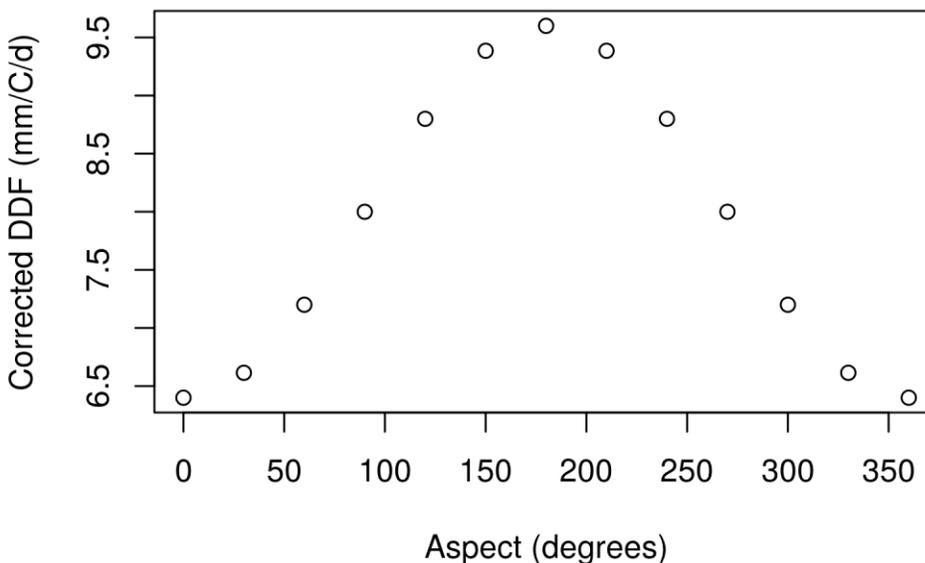
Eq. 4 only deals with aspect-dependence of degree-day factors, and as we are using a degree day approach we do not calculate energy fluxes on horizontal/inclined surfaces.

- Eq. 4 and/or Table 2: please ensure consistency in notation, i.e. K (Eq. 4) vs. ddf in Table 2.

Thanks, this is fixed.

- P5384 L8: How was $R_{exp} = 0.2$ determined?

$R_{exp} = 0.2$ was the value used in Immerzeel et al. (2012), which we now refer to in the revised manuscript. For a base degree day factor of $8.0 \text{ mm } ^\circ\text{C}^{-1} \text{ d}^{-1}$, this gives a corrected degree day factor between 6.4 (north) and 9.6 (south).



- I don't think Table 1 is referred to in the text.

This has been fixed in the revised manuscript.

- Figures generally, but particularly 1 and 10: the labeling is really small, please increase the font sizes a bit.

We have tried to increase the font sizes – I think this was partly a function of the scaling used in the TCD paper.

- P5385 L3: "roughness" (typo).

Fixed.

- P5385 L17: "steeper cells receiving a greater share of the ice flux": how exactly is this partitioned? **Ice being transported out of a specific cell is distributed to all downstream cells proportional to the slope with this downstream cell.**

- P5286 L1: "daily timesteps using a 0.2 m w.e. threshold, which represents the average seasonal snowfall depth": why are you using a threshold? I'm not sure I get what you want to say here. **The 0.2 m w.e. threshold refers to the classification of a grid cell as either glacierized (>0.2 m w.e.) or simply snow covered.**

- Eq. 8: please refer to Eq. 6, and add an explanation why you are using tau_0 instead of tau_b here. **Reference to Eq. 6 added, and tau_0 is used here since we do not know tau_b (which requires an estimate of thickness).**

- P5386 L23-27: I find the argument for a special Khumbu glacier melt factor weak: what constitutes "anomalous ice flow velocities"? The model will always show different success at different glaciers, so adapting special parameters for individual glaciers seems arbitrary (and potentially distorts the validation: what if there are more "anomalous ice flow velocities" which you don't recognize, because of lacking observations?).

The model failure at Khumbu Glacier may be related to the inability of the Weertman sliding model to capture the extreme ice deformation in the icefall system that leads down to Khumbu Glacier. The use of the optimized DDFd value thus resulted in unrealistically high flow velocities and ice extent in this region. To compensate for this we used a separate DDF for the Khumbu tongue. Improved representation of the glacier dynamics, or possibly testing a greater range of calibration parameters, may help avoid this admittedly arbitrary correction. We discuss this further in the revised text.

- P5387 L15-18: I think this score is biased – or what would happen if all the glacier polygons are ice covered, but additionally many more? Perhaps you did it, but didn't mention it, that there also should be a penalty for ice cover outside the polygons.

Yes, this score captures both ice covered and ice-free areas (see the study area figure with the polygons used). We have clarified this in the revised manuscript.

- P5387 L18-20: I don't understand how this score is calculated, perhaps just because of the example numbers given are a bit ambiguous: If the mass balance during 1992-1998 had been 0, what would the score of a modeled mass balance of -0.02 be? Perhaps give equations for the scores.

The MB score is defined as:

$$\text{MB score} = \text{ABS}((\text{modelled mass balance} / -0.4) - 1)$$

So a modelled mass balance of 0 m w.e. γ -1 would result in a score of 1. By definition this means indeed that this may yield unrealistic results if the observed mass balance would have been close to 0. However this was not the case and the simulated mass balances ranged from -0.12 to -0.39 and given the observed MB of -0.40 we conclude that all computed scores are realistic.

- More generally, concerning the scores: why multiplying instead of adding the scores? By multiplication, a single very good score will lead to a very small totalscore even if the three remaining scores are disastrous.

This is a good point. If the four scores are added, run 5 has the lowest score overall although originally, while multiplying, it was the second best score. We have changed the text and table to reflect this.

- P5390 L3: introduce "SD".

Done.

- Fig. 8: add C to vertical axis of lower left panel.

Done.

- P5390 L29: add unit to T.

Done.

- P5391 L4&5: change "modelled" to "predicted".

Done.

- P5391 L18-19: so you actually do weight the scores (cf. P5387 L14), because you don't take run with the lowest overall score. Perhaps this would change with an additive overall score?

As indicated above, we have changed this to additive scores, and selected run 5 because it has the lowest score. Run 5 also provided improved performance with respect to glacier mass balance and velocity, which we qualitatively weight as being more important in this modeling setup.

- P5391 L21-22: I think you can't call the parameter set of run 5 the "final calibrated degree-day factors", because you had on average less than two values to choose from (see general comment above).

We will describe values from run 5 as the 'selected parameter set'.

- It's such a pity your model run goes to 2007 when the observation starts in 2008...any chance on having an updated atmospheric data set (not necessarily now, but later)?

Agreed! We hope the APHRODITE data set will be updated in order to do a full comparison with observed mass balances.

- P5393 L20: please mention the minimum slope already somewhere close to Eq.8.

Done.

- P5394 L16: now I understand the threshold mentioned earlier (P5286 L1) –please add an explaining sentence there.

Done.

- P5395 L2-5: Depending on how the initialization run was made, I can also imagine this as an explanation for the differences (particularly since they are larger in the first decade, but that could of course be coincidence...).

True, we have added text to this effect in the discussion.

- P5395 L7: perhaps rather say something like "extracted from members of the CMIP5 ensemble that capture the range of results in the respective scenarios".

Done.

- P5397 L5: "ensemble" (singular).

Done.

- P5397 L17: please replace "highly significant" with "good" or something similar, or "matches" with "correlations" – which would slightly change the meaning as well.

Done.

- P5397 L18: change "model" to "predict" or "downscale".

Done.

- P5398 L10-17: here would be a good place to also (qualitatively) discuss the impact of having a stationary debris cover in your model – will this lead rather to an over- or underestimation of the projected changes?

New text added: "Our assumption of stationary debris cover may be incorrect in the long-term, as glacier wastage typically leads to increased debris concentrations and the development of a debris cover. However, the median glacier slope above 5500 m is greater than 20° (Figure 7), and the development of debris cover on such slopes is unlikely (cf. Fig 3b, Scherler et al., 2011) as de-glaciation proceeds. "

- P5398 L22-24: I don't think this is a good justification for the linearization, as (i) it would also be possible to keep natural variability constant without linearizing (e.g., apply ensemble mean anomalies); and (ii) using an anomaly coupling of the full time series will retain the characteristics of the observed monsoon, by removing the biases that the GCMs have in the present day.

This justification is removed as we now use the full T+P anomalies.

- P5399 L16: b instead of b_a , and why >0 if in Eq. 13 is $-H$?

Fixed the subscript, and mass balance at the terminus is negative.

- P5399 L21&22: "years" (plural).

Fixed.

Reviewer #2: Ann Rowan

Shea and coauthors apply a glacier mass balance and ice flow model to glaciers in the DudhKoshi catchment in the Nepalese Himalaya. The topic is important and relevant to the scope of The Cryosphere. The glacier model has previously been applied elsewhere in Nepal, but some major issues with the glacier model application concern the representation of glaciers where thick supraglacial debris modifies both mass balance and dynamics. The manuscript is generally well written, although difficult to follow in places due to the large number of terms and parameters discussed in each section and for both models.

1. The model is applied with different degree day factors for clean ice, snow, and debris-covered ice (with a different value used for Khumbu Glacier) to represent different amounts of melt expected between clean ice and debris-covered ice across the study area. However, the dynamics of these glaciers are treated as whole, and include active ice, debris-covered ice that is likely to include large stagnant areas, and tributary glaciers that at present are dynamically detached from larger glaciers (e.g. Changri Nup and Shar). Calibrating the model to the area of these glaciers seems to be matching simulations with the Little Ice Age climate rather than representing 1961–2007.

Response: This comment raises some interesting points of discussion. Our work in Langtang Valley (Immerzeel et al., 2014) indicates that debris-covered glaciers that are completely detached from their accumulation areas still exhibit glacier flow. And while it is true that the large debris-covered tongues can include a mix of stagnant ice, it is not clear from Quincey et al. (2009) that any of the glaciers in the Everest region are completely stagnant. Centre-line profiles (cf. Figures 4 and 6, Quincey et al., 2009) indicate that velocities are non-zero along the length of nearly all glaciers surveyed. Modelling the dynamics of debris-covered glaciers is a complex and challenging task, and our discussion previously raised this point, but we have added additional text to the section on “Glacier Models” to make this clear. While the current extents of large debris-covered termini are not dissimilar from the Little Ice Age maximum extents, calibration of the model to the selected area helps to avoid model parameters that would lead to unreasonable glacier extension or wastage.

2. Four sets of observations are used to calibrate the model by comparison of results from four large debris-covered glaciers; (1) terminus position, (2) glacier area, (3) mean flow velocity, and (4) mean basin-wide mass balance. Of these four values, both (1) and (2) are unlikely to change over the next century even with considerable mass loss as these glaciers are downwasting rather than receding. Mean flow velocity (3) is not a particularly helpful calibration value as velocities vary from 40 m per year to zero over relatively short flowline distances (Quincey et al. 2009). A better calibration/validation would be to see the variations in velocity across each glacier, and ideally similar ice flow rates, reproduced by the model.

Response: Of the four observations chosen for calibration, only two refer to the large debris-covered glaciers: terminus position and mean flow velocity. The others are basin-wide. Downwasting will eventually lead to retreat of the glacier termini (e.g. Immerzeel et al., 2014), so we disagree with the reviewer on this small point. The challenge from an observational viewpoint is to determine what is glacier and what is not on these large debris-covered termini.

As suggested by the reviewer, we compare the modeled ice flow velocities against observations from Quincey et al. (2009), and a recently accepted manuscript (Dehecq et al., accepted in RSE). We caution again that the model is not meant to be an explicit representation of glacier dynamics! It is intended to simulate mass balance and mass redistribution. A map view of modeled velocities from

1961-2007 is shown below in Figure R9, with the location of centerline profiles on Ngozumpa and Khumbu glacier (these may not match those used by Quincey et al., 2009). Averaged over the entire period (1961-2007) we achieve a reasonable representation of glacier flow in the basin, except for the termini of debris-covered glaciers where velocities are overestimated. The overestimation of velocities at the terminus is a function of the ill-defined ice thicknesses at low slope angles, and reflects short-lived rapid flows. Averaged over 2000 – 2007, our model shows stagnant debris covered tongues, and a slight decrease in velocities of the ramps that descend from the accumulation areas.

Centerline velocity profiles of Ngozumpa and Khumbu glacier (Figures R10 and R11) show that velocities on glacier termini are typically less than 10 m/yr, with increases into accumulation areas and steeper regions. This pattern and magnitude corresponds with the feature-tracking observations of Quincey et al. (2009) and Dehecq et al (in review, RSE) and we are satisfied that the simplification of glacier dynamics (i.e. the whole ice mass moves with the same velocity and deformation is neglected) is suitable for the current study, the goal of which is not to physically simulate the ice flow but rather to have a basic ice-redistribution model to allow glacierized areas to shrink, in order to provide realistic glacier response to temperature or precipitation changes.

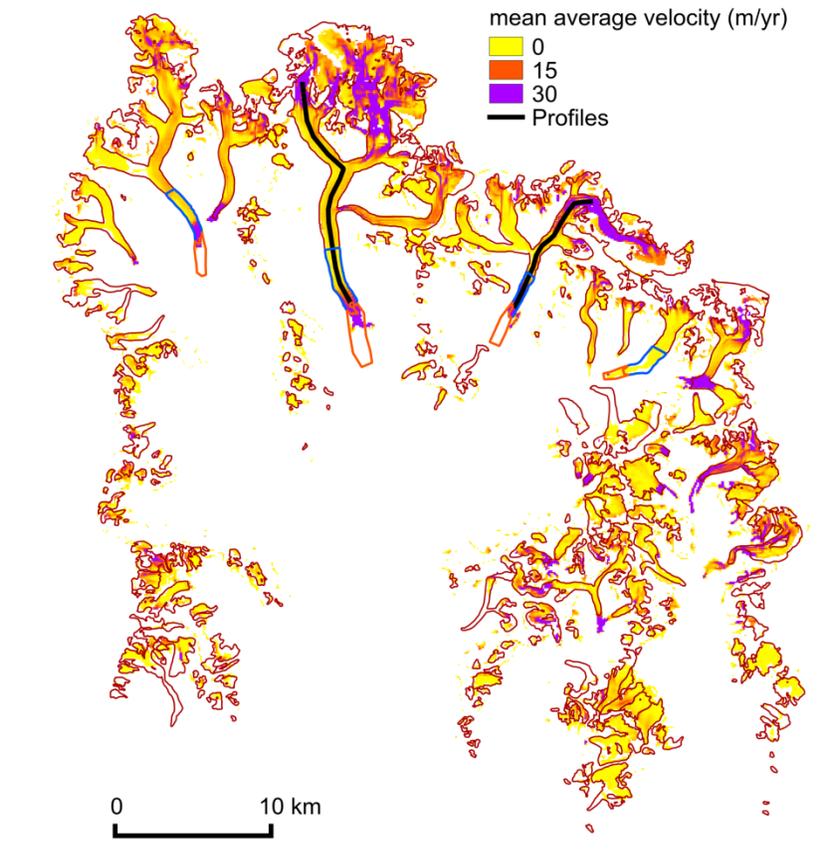


Figure R9: Mean annual velocities (1961-2007) and centerline profile locations for the calibrated run. This figure is not included in the revised manuscript.

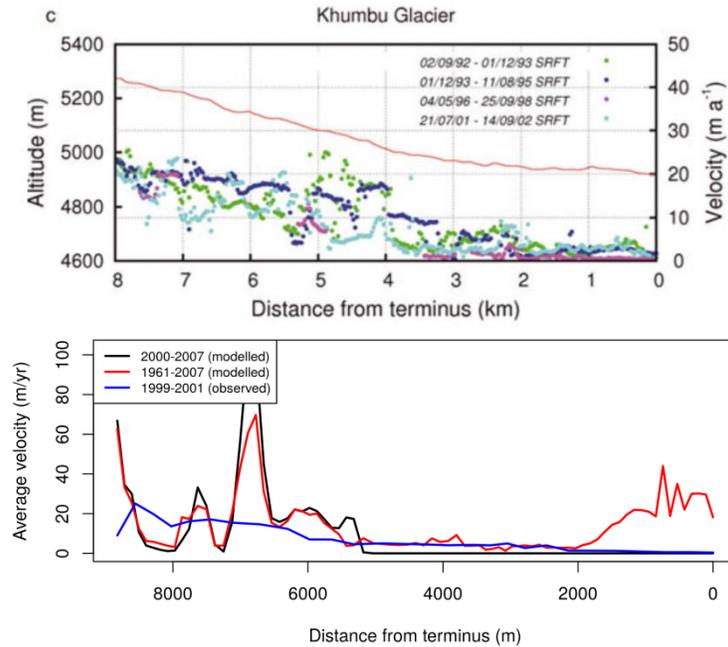


Figure R10: Centerline profiles of modeled and observed mean annual velocity for Khumbu Glacier. Results from Quincey et al. (2009) are given in the top panel, and observations from Dehecq (in review) are given in the bottom panel. Note that y-axis velocity scales are different.

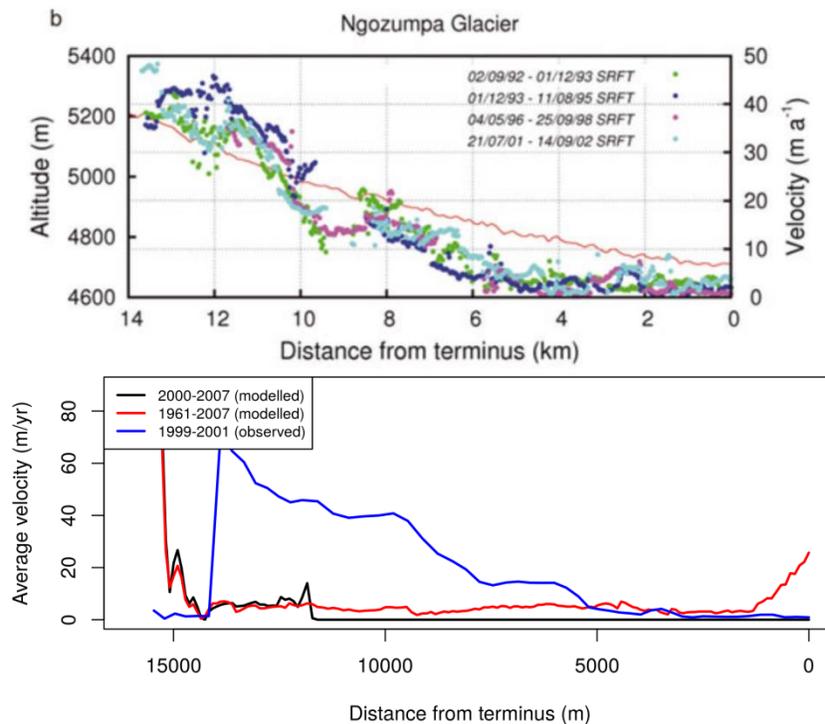


Figure R11: Centerline profiles of modeled and observed mean annual velocity for Ngozumpa Glacier. Results from Quincey et al. (2009) are given in the top panel. Note that y-axis velocity scales are different between both panels

3. The model outputs are validated against four independent datasets, although this validation is rather difficult to follow as written. The fit between calibrated model outputs and decadal glacier extents does not appear to be presented in Section 3.2 or elsewhere. Ice thickness validations are only presented for two glaciers and there is some mismatch in each case, the impact of which on the results should be quantified.

Response: We have attempted to clarify the validation in the text by reorganizing the subsections, and the fit between modeled and observed extents is now included in Section 3.2. In addition to the comparisons of ice thickness at Mera Glacier and Changri Nup Glacier (Figure 13 in the manuscript), we compared our modeled Khumbu Glacier thickness with observations from Gades et al. (2000) in the text (P5393 L20-25). We are not able to directly quantify how the errors in modeled thickness will affect our overall results. We have attempted to incorporate as much field data as possible in this study, and hopefully with future field work and collaborative networks a greater degree of model validation can be done.

4. The validation also refers to ICIMOD glacier outlines from 1980–2010 (Bajracharya et al., 2014a) that are not presented in the current manuscript or visualised in the cited report. An image of or a link to these data should be included to allow the reader to compare these with the model outputs.

These data are available at rds.icimod.org, and a link is now provided in the manuscript and the reference to Bajracharya et al. The 2010 glacier extents are given in the study area figure (Figure 1), as well as the maps of model output (Figure 10). As indicated in the manuscript (P5388), we do not use the 1980 extents as snow appears to have been misclassified as ice in the 1980 inventory.

Specific comments:

- P5376 L17–19: This comment is unclear, why does the stated glacier elevation result in sensitivity to temperature and ELA change?

As the majority of glacierized area in the basin lies between 5000 and 6000 m, increases in temperature and upward shifts in the current ELA (which is between 5500 and 5800) can result in large increases in ablation area and melt.

- P5377 L15: Figure 2 referred to before Figure 1.

Fixed.

- P5377 L22–26: Please also give absolute values for precipitation as well as temperature here. What fraction of annual precipitation is snowfall?

The fraction of annual precipitation that falls as snow will vary depending on elevation and how liquid/solid precipitation ratios are defined. Similarly, the absolute values of precipitation varies with elevation and location, and we have not done an in-depth analysis for this study. It appears we should have done a separate paper on the downscaling routine first!

- P5377 L26–P5378 L3: The range scale distribution of precipitation described by Bookhagen and Burbank is not relevant to scale and elevations described in the current study, please remove this text.

Done.

- P5378 L15–17: Does the value for decrease in glacier extent refer to both clean-ice

and debris-covered glaciers?

Yes, this is now specified in the revised manuscript

- P5379 L3–4: Explain what is mean by “mass balance amplitude”

Mass balance amplitude refers to the rate of mass turnover (Cogley et al., 2011).

- P5379 L24: What is the elevation of the rain–snow threshold for the study area?

We had not looked at this previously, so thank you for bringing up this important metric. Based on the calculated 1961–2007 daily temperature gradients and intercepts from APHRODITE and SRTM data (Figure 3 in the manuscript) and the day-of-year bias correction, we have estimated the daily elevation of the OC isoline ($Z_{T=0}$), which in this study represents the rain-snow threshold and the melt onset threshold. Monthly boxplots (Figure R12) show that mean $Z_{T=0}$ varies 3200 m in the winter to 5600 m in the summer. This corresponds to field observations from Langtang (Shea et al., accepted) and from the Khumbu (Wagon, Response to Reviewer (C. Mayer) in TCD; <http://www.the-cryosphere-discuss.net/7/C1879/2013/tcd-7-C1879-2013.pdf>). However, our derived $Z_{T=0}$ may be an overestimation of the true $Z_{T=0}$, as APHRODITE data are based on valley station data from much lower elevations. Temperature gradients at higher elevations will likely follow the dry adiabatic lapse rate due to the lack of moisture, resulting in a more rapid decrease in temperature with elevation.

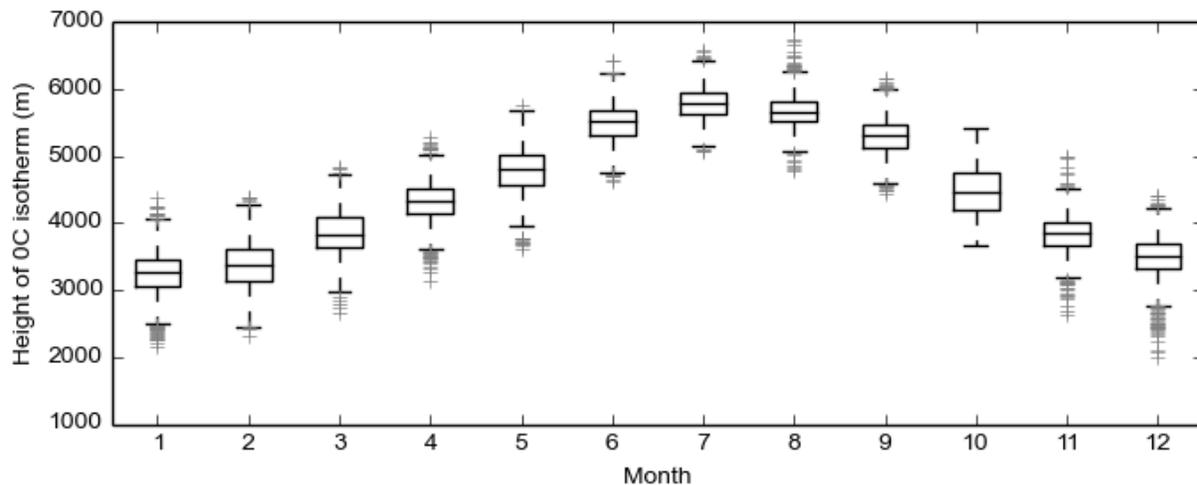


Figure R12: Boxplots of monthly temperature gradients derived from SRTM, APHRODITE, and bias-correction.

- P5382 L 10: The temperature bias values range between 3_C and 6_C. Please state this in this paragraph.

Done.

- P5384 L10: The glaciological reasons for the use of a different DDF for Khumbu Glacier are unclear.

Please see our response to Ben Marzeion above.

- P5386 L20–27: When the calibrated model is run forward from the 1961 starting point, does it accurately reproduce the present-day glaciers when this year is reached?

The model is run from 1961-2007 for calibration with glacier extents, and validation with glacier change statistics. So yes, it reproduces the present-day glacier extents and total area reasonably well as a result of the calibration procedure

- P5391 L1–7: Do the measured precipitation values include snowfall as well as rainfall? If not, this may account for the overestimation of modelled precipitation at lower elevations, as higher stations would underestimate total precipitation to a greater degree.

For comparison with the EVK2CNR data we consider only days where liquid precipitation is expected ($T > 0$), as these sites record precipitation with tipping buckets (P5383L4, original manuscript and also specified in Fig 6 caption).

- P5393 L21–25: The authors refer to the comparison with their ice thickness estimates For Khumbu Glacier and those GPR data presented by Gades et al. (2000), but these results are not presented.

The results of Gades et al. (2000) were presented, but not in a graphical format.

Short Comment #1: Franco Salerno

The modeling efforts carried out in this paper to forecast the glaciers behavior are in my opinion very important in the actual scientific debate. Therefore the paper faces an interesting topic. In general I think that from one side the paper shows real high modeling potentiality, but from another side the methodological aspects presented here are not easy to follow. I think the paper needs a review focused to increase the readability of the method section in order to 1) the scientific rigor can be assured and 2) the followed approach can be replicable to other regions. I fully agree with one of the comments posted by B. Marzeion (Referee), that model outputs depend on the input parameters. The authors should make it clear how model initialization works. It is not clear which part of the model is based on methods presented in previous papers and which parts are new.

The authors refer to the ICIMOD report for calibrate/validate glacier parameters that is not a peer-reviewed dataset. Why not the authors use the recently published dataset on terminus, surface changes from Thakuri et al (2014; Supplement) that provides a complete dataset from 1960s to 2011 for the Sagarmatha National Park?? In particular for the initialization of the 1960s (as required by the reviewers) this study could be useful if the model runs are for each glacier separately (if I understood the method followed). In their paper Thakuri et al, 2014 analyze about 400 km² of glacier surfaces.

As referenced in the manuscript, the methods of Bajracharya et al. have been published in the Annals of Glaciology, and the ICIMOD glacier dataset has been ingested into GLIMS and the Randolph Glacier Inventory. We reference the results of Thakuri et al in the section on “Modelled and Observed Glacier Extents and Shrinkage”, but neglected to compare our approximated 1961 area (499 km²) with their estimate of 404.6 km². While the order of magnitude is correct, there are a large number of possible explanations for this discrepancy that are out of the scope of our study here. The main reason would probably be differences in the glacier classification between Bajracharya et al. and Thakuri et al., although it is also likely that the glaciers do not expand quickly enough in the initialization run.

Concerning the meteorological forcing,

1) the authors carried out the downscaling of APHRODITE through the EVK2CNR data. In this way they got good performance on mean values (mean bias and root mean square error). However I am not sure that the adopted procedure can be considered as an independent validation (as declared) due to the fact the same stations were used to calibrate the mean bias correction.

Yes, we were not careful enough with our terminology. We have clarified that this is a semi-independent validation, as we use the mean bias from all stations to correct the APHRODITE temperature fields (which are fully independent of the EVK2CNR data).

2) the authors do not have the availability of long series for discerning adequately the different gridded reanalysis existing climatic data set. However Yatagai et al., 2012 developing APHRODITE, underlined some discrepancies with GPCP in Nepal during the last decades. Furthermore Yao et al, 2012 underlined the monsoon weakening in the Himalayan region, confirmed by Salerno et al. (on TCD) and inferred by Wagnon et al., 2013. How do the non-stationarity of GPCP precipitation affect the results of this study??

This is a good question that we neglected to discuss in our paper. We had not seen the Salerno TCD paper prior to submission of our own article. But a weakened monsoon would result in reduced accumulation at all altitudes, and a subsequent increase in melt as you and co-authors have suggested.

3) Finally, the 500 m elevation bins used to estimate daily precipitation vs elevation might be too rough. As highlighted in the discussion this is a major source of uncertainty for the glacier mass balance. The precipitation gradient estimated on the base of the observations in Salerno et al. 2014 may help to resolve some of the miss-fit (for example at Pheriche (fig. 6)), and better constrain the overall calibration process. Furthermore I provide some suggestions and comments considering the recent paper of Thakury et al., 2014 analysing changes for different glacier parameters since 1962 in the same regions and considering the paper of Salerno et al (on TCD) analyzing temperature and precipitation trends for the last twenty years.

We feel that using a finer vertical resolution for the precipitation gradients would be stretching the APHRODITE set beyond its limits. Our temperature and precipitation gradients appear to correspond with those presented in Salerno et al (2014), which is encouraging.

p5377 (23-25) Salerno et al (on TCD) report that nearly 90% of the annual precipitation falling in the months of June to September. Considering that the mean daily temperature during these months is above 0C. On a yearly basis, this probability reaches 20% of the annual cumulated precipitation.

Our statistics were derived from the APHRODITE dataset, so there will be some differences with high-altitude observations.

p5377 (26) Salerno et al (on TCD) report that the precipitation sensors at these locations are tipping buckets usually used for rainfall measurements and may not fully capture the solid prec. Therefore, prec is probably underestimated, especially in winter.

Yes, we agree and consequently we only compare predicted and observed precipitation for days when $T > 0$ (supposed to experience rainfalls only).

p5378 (1-3) Salerno et al (on TCD) report the precipitation gradient along the altitudinal transect until 5600 m. They observe a clear rise in precipitation with elevation until approximately 2500m.a.s.l. At higher elevations, they observe an exponential decrease

Our revised figure (see response to Marzeion above) shows a very similar pattern in the pre-monsoon and monsoon periods, though our decrease occurs after 3500 m.

p5378 (1-3) I suggest to report here the study just published few months ago by Thakuri et al., 2014 reporting the surface area loss on the south side of Mt. Everest from 1962 to 2011 considering five intermediate using optical satellite imagery. They found an overall surface area loss of 13.0 3.1%.

The Thakuri et al. (2014) study was referenced in the discussion, but we have added it here as well.

Comment: this result in term of surface area loss is strictly comparable with the decrease of 15.6% in term of ice volume observed in the manuscript for the same period. I will underline in the discussion section that the general statement "surface area loss in debris coverage glaciers is not representative of the real mass loss" is rejected.

We would politely disagree: ice volume and area change statistics are *not* strictly comparable, as glacier thinning and mass loss can occur without significant changes in glacier extent (though this is probably only true for short periods).

p5379 (4) Here I would report data for precipitation for north and South slopes of Mt Everest. At Pyramid station (1994-2013) the total annual precipitation is 463 mm (Salerno et al., TCD). In the north I would cite Yang et al., 2006.

This comment appears to refer to another location in the text.

p5379 paragraph 1.2. I think that in this section needs to be specified for the readers that in the Dud Koshi the shrinkage of the glaciers in south of Mt. Everest is the lowest of the region as reported by Thakuri et al., 2014 using surfaces and terminus. The same considerations could be reached considering the mass balance differences reported by other authors (e.g. Yao et al., 2012).

We are not trying to place the observed changes in a regional context, but the Thakuri et al. (2014) paper is now included here.

p5379 (17-20) Salerno et al., 2014 (on TCD) find just higher trends for minimum temperature.

We have added a reference to Salerno et al (2014) here.

p5379 (2-6) Salerno et al., 2014 (on TCD) find a huge decreasing trend at high elevation.

This will have potentially large implications for glaciers in the region, as your paper notes.

p5379 (5-22) Salerno et al., provides the precipitation gradient

We do not fix precipitation gradients but use daily $P = f(\text{elevation})$ functions.

Yatagai et al., 2012 APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges

Wagnon et al., 2013 Seasonal and annual mass balances of Mera and Pokalde glaciers (Nepal Himalaya) since 2007

Salerno et al. (on TCD) Weak precipitation, warm winters and springs impact glaciers of south slopes of Mt. Everest (central Himalaya) in the last two decades (1994–2013)

Yang et al., 2006. Climate change in Mt. Qomolangma region since 1971. J Geographical Sciences

References:

Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., & Zemp, M. (2011). Glossary of glacier mass balance and related terms, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2.

Dehecq, A., Gourmelen, N., & Trouvé. Deriving large-scale glacier velocities from a complete satellite archive : Application to the Pamir-Karakoram-Himalaya. In review, *Remote Sensing of Environment*.

Marks D, Winstral A, Reba M, Pomeroy J, & Kumar M (2013) An evaluation of methods for determining during-storm precipitation phase and the rain/snow transition elevation at the surface in a mountain basin. *Advances in Water Resources* 55:98-110 doi:10.1016/j.advwatres.2012.11.012.

Reuter H.I, A. Nelson, A. Jarvis, 2007, An evaluation of void filling interpolation methods for SRTM data, *International Journal of Geographic Information Science*, 21:9, 983-1008.