In response to a few comments from the reviewers we chose to re-run the modeling for Lewis Glacier. In these re-runs we:

(a) Altered the bottom boundary condition which represents the internal glacier temperature to be more representative for a glacier that we are fairly confident is temperate. The bottom boundary temperature is now optimized over a range spanning 0°C to -2°C rather than -0.5°C to -5°C.
(b) Altered the range of possible fresh snow density from 400±50 to 370±60, which ranges from the upper limit of fresh snow density measured in the field to just below the lowest density measured for fresh snow in the field. The measured snow density at Lewis Glacier is very high so this alteration prevents the model optimization from exceeding the highest density field measurement.
(c) Increased the range over which surface roughness parameters being optimized were allowed to vary.
(d) The initial temperature field of the glacier sub-surface was set to be equal to the surface temperature, rather than an independent variable.
(e) Changed distribution of accumulation from being evenly distributed throughout the hours of a day with accumulation to being distributed evenly between the hours of the day for which RH>90%. Tests have shown that this makes little difference to the modeled mass balance, but it is logically a more realistic representation of the accumulation.

As a result of these changes, the specific values of modeled surface mass and energy balance have changed slightly, but the general patterns and conclusions are not altered.

We also altered the selection of sub-sampled period for the surface energy balance analysis because the previously selected 'dry' period of January 2011 represent a cold temperature anomaly in reanalysis data of air temperature for our site.

Response to reviewer 1:

General comments:

1) The reviewer argues that the validation is weak because:

(i) The surface height change only validates the ablation component of the model as the surface height change is used to derive accumulation input as well
(ii) The mass change is converted to a surface height change based on density assumptions within the model, that were optimized to match the measured surface height change, and is therefore not an independent validation
(iii) Validating against the independent variable of surface temperature is based on the assumption that the emissivity of the glacier surface is 1, and if it is less that 1, then the error on surface temperature computed from outgoing longwave could be in the order of a few °C

It is true that the SR50 data does not provide a fully independent assessment of model performance as it is the source of the accumulation input. However, at our site the nearby ablation stakes were measured only twice/year so is insufficient for statistical assessment of the model performance. The validation against surface temperature is provided as a second, independent validation because of this. This is the standard procedure in remote measurement sites, shown in numerous papers.

This was stated clearly in the text where we wrote: "Because modeled surface height increases are driven by the accumulation input derived from the measured surface height change, only the modeled surface lowering can be independently validated by the measured surface height change. Therefore, the modeled glacier surface temperature and the surface temperature computed from *LWO* measurements are also compared as a measure of model performance."

We have also added to the model validation figure a further independent validation against a single stake measurement within 5 m of the AWS, which is a fully independent test but provides validation at only a few points in time. This data comparison is now shown in Figure 4 and we write: "In addition we compare the modeled surface height change at the AWS to that measured at a mass balance stake within 5 m of the AWS site to provide a second independent evaluation. This stake was measured at 6 month intervals throughout the modeled period."

Glacier emissivity is commonly taken to be 1, and indeed the reviewer also finds this acceptable. For instance, Sicart et al., 2005, discuss the impact of the choice of emissivity briefly and conclude that using a value of unity is acceptable given the accuracy of the pyrgeometer sensors. The conversion of both measured LWO to surface temperature and the surface temperature calculation in the model used this value for emissivity so the choice of value does not influence the validation assessment. Surface temperature based on emitted radiation is associated with uncertainties of about 2-2.5°C, so the fact that the RMSD between modeled and observed surface temperature is well within this limit is an indication of satisfactory model performance.

2) A sensitivity assessment of the input parameters is required

We have added the results of a single parameter sensitivity study into Table 2. In this sensitivity study individual parameters are perturbed to the maximum and minimum of their plausible ranges, while all other parameters are held constant and we report the impact of the parameter perturbation on modeled mass balance.

We also clarified that the ranges specified in Table 2 do not represent error ranges, but rather the range of plausible values based on the underlying physics, the published literature and available field measurements. The goal was to constrain the parameters as far as possible on the basis of existing knowledge, but we have increased the ranges of the surface roughness parameters as advised by the reviewer.

3) Energy flux into the glacier

In our model results we presented the energy fluxes in the subsurface in two separate components, conductive heat flux, QC, and the penetrating shortwave radiation, QPS, which together (QC + QPS) form the energy flux into the glacier, often denoted as QG. Thus, in this approach QC+QPS must be close to zero, not QC alone as was assumed by the reviewer. QPS is not always considered explicitly in energy balance models but this is a useful way of presenting the data in that it allows us to see the driver of energy flux into the glacier. The QC that we report is calculated for the surface layer of the model, which in our case is the uppermost 0.09m of the glacier.

4) Representativeness of the study period in terms of temperature

In the absence of long-term air temperature observations at the site, and concerns that reanalysis data is not suitable for assessing temporal trends in climatological data we initially chose not to assess the representativeness of the measurements period in terms of temperature. However, we have now performed an additional analysis of ERA interim air temperature fields, and have decided that it will be valuable to add a figure illustrating both these patterns and those from the TRMM data.

Accordingly we have added text as follows: "The ERA-interim temperature data indicates that for the period of available measurements at Lewis Glacier, air temperature tended to be higher than the monthly average through February – June 2010, but only in May 2010 did the high temperature anomaly exceed the monthly standard deviation. Air temperature was anomalously low through January- April 2011 and in February 2012, and was generally lower than the monthly average from late 2010 until the end of the measurement period. The mean annual temperature cycle at 500 hPa level in the ERA interim data shows that over the period 1979-2012 April and May experience the highest air temperatures of the and July-October the lowest. Air temperature in February shows the greatest variability in this dataset. "

In response to adding this analysis we also altered the sub-sampled periods, as January 2010, which was initially chosen as a sub sample for dry conditions, represents a cold anomaly in the ERA-interim data.

Specific comments:

1. P 5182, line 7: comparable to South American tropical glaciers, for those located in the inner tropics (AR).

Now reads: "comparable to those experienced in the ablation zones of South American glaciers in the inner tropics"

2. P 5183 and following: EEA, MAM, IO, IOZM, ENSO... so many acronyms. It would help the reader to limit the number of acronyms along the text when possible.

Removed EEA, IO, IOZM and SST throughout the text but retained ITCZ, ENSO (because these are widely understood) and retained the MAM/OND seasonal shortenings.

3. P5184, line 11-13: it would have been useful to quickly provide some mass balance results from these studies to quantify the LG recession here.

Added: "Where concurrent data are available the geodetic and glaciological measurements of mass balance are within error of each other. Between 1934 and 2010 Lewis Glacier lost 16.67 ± 3.82 × $10^6 m^3$ (90%) of its 1934 volume and 0.394 ± 0.015 × $10^6 m^3$ (79%) of its1934 surface area (Prinz et al., 2011). Mean specific mass balance rate between 1934 and 1974 ranged between -0.22 ±0.40 to -0.54 ± 0.63 m.w.e. a^{-1}, after which mass balance became more negative, and was most negative between 1993-2004 when mean specific mass balance rate was -2.22 ± 0.44 m w.e. a^{-1} (Prinz et al., 2011), and the most recent glaciological measurements of annual mass balance are -1.40, -1.54 and -1.03 m w.e. for the mass balance years 2009/10, 2010/11 and 2011/12 respectively (Prinz et al., 2011, Prinz et al., 2012)."

4. P5185, lines 5-7: additionally to points (i) and (ii) (and probably more effective) is the effect of reduced precipitation on ablation (decrease in precipitation -> depletion of albedo -> increase in ablation).

Now reads: "In contrast to this energy balance response on Kilimanjaro, studies show that on South American tropical glaciers reduced atmospheric moisture reduces accumulation and can either enhance or reduce ablation rates: Ablation can be enhanced through the albedo feedback in which reduced precipitation results in decreased albedo and enhanced absorption of solar radiation or reduced as a result of (i) diverting energy from melting to the energetically more expensive ablation process of sublimation, and (ii) reducing incoming longwave radiation (Wagnon et al., 1999; Francou et al., 2003; Winkler at al., 2009)."

5. P5185, lines 9: the effect of ENSO is different between inner and outer tropics. In both cases, it is true that there is a warming usually observed in mountain areas, but in the outer tropics, also a precipitation depletion, which has the strongest impact on glacier mass balance.

The original text describes the impact of El Nino as follows: "The additional mass loss in the inner tropics results from changes in the spatial distribution of solid and liquid precipitation over the glacier surface, while in the outer tropics it is due to reduced precipitation (Wagnon et al., 2001; Favier et al., 2004a; Francou et al., 2003, Vuille et al., 2008)."

However, we changed the whole section to be even more clear, and it now reads: "The mass balance variability of the South American tropical glaciers is influenced by ENSO, with El Niño events resulting in more negative mass balance: Warmer temperatures during El Niño increase ablation and, in the inner tropics, also mean that a higher proportion of precipitation on the glacier surface is liquid rather than solid, while in the outer tropics, El Niño events are associated with reduced precipitation as well as warmer temperatures (Wagnon et al., 2001; Favier et al., 2004a; Francou et al., 2003, Vuille et al., 2008).

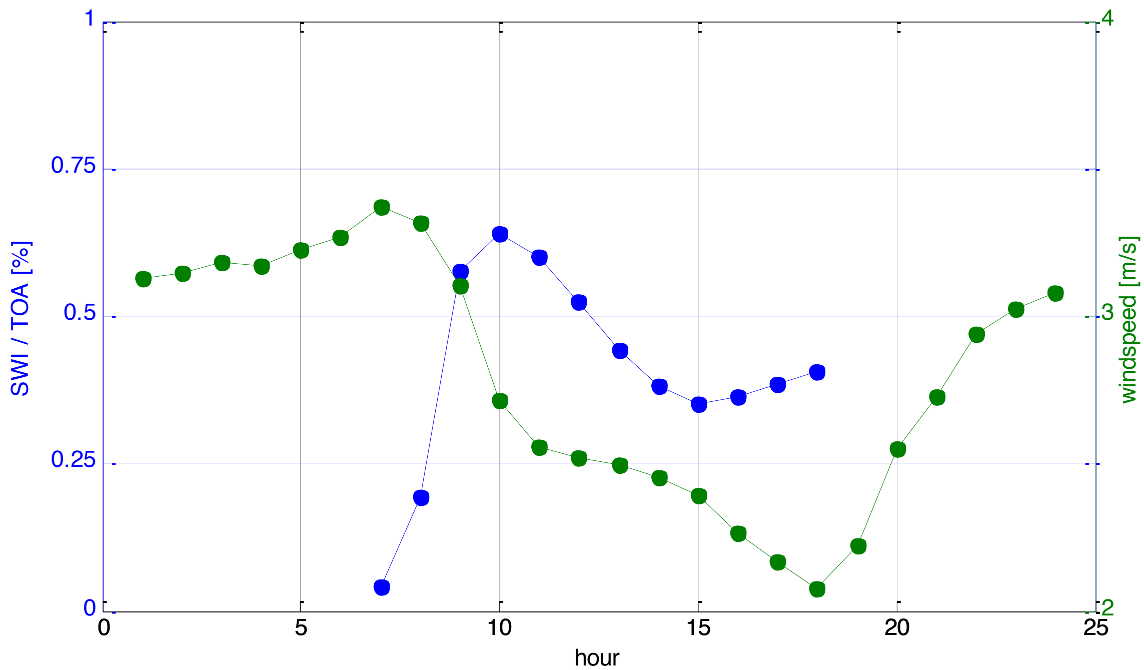
In the interests of being concise removed the following sentence which read "The relationship between ENSO and glaciers of the outer tropics is (i) often augmented by changes in the ablation regime as drier conditions are usually warmer and vice versa, and (ii) affected by instability in the circulation systems underlying the reduction in precipitation, which can disrupt the relationship in the boundary zone between inner and outer tropics (Vuille et al., 2008)."

6. P5186, lines 17-23: Considering that wind speed is low on this point site (Table 4), I assume that radiative heating of the Vaisala sensor may affect temperature measurements in a much more effective way than what is reported here. How can the author be sure that the sensor is barely affected?

Without concurrent ventilated measurements we cannot perform a formal assessment of the potential influence of radiative heating on the measurements. However, we think it is more justifiable to leave the values uncorrected than to apply a correction of unknown quality because:

(i) we use a radiation shield that consists of two layers of louvres to fully shield the sensor from all directions, so it is especially suitable for measurements over snow and ice surfaces

(ii) at Lewis glacier, the development of cloud cover coincides with the minimum wind speed from 11:00 onwards, as seen in the average diurnal cycle of SWI to top of atmosphere radiation shown below



(iii) it is difficult to identify readings that could potentially be affected by radiative heating as the offset between distribution of temperature measurements coinciding with windspeed above and below the 3.5 ms^{-1} wind speed threshold identified by Georges and Kaser (2002) is similar for both night and daytime, which implies that lower winds speeds are associated with higher air temperatures regardless of the degree of solar radiation.

(iv) finally, we followed the published method of Mölg et al. (2008), to assess the influence of the additional sensor membrane on the Vaisala instrument, and as deviations between these two datasets were in excess of the error in only 0.02% of the data were affected, we made no replacements.

We now expand on this and write: "Naturally ventilated air temperature measurements can suffer from radiative heating when insolation is high and windspeed is low and the sensor shielding is inadequate (e.g. Georges and Kaser, 2002), and in the case of combined temperature and relative humidity sensors the membrane that protects the sensors from contamination further impedes natural ventilation. The Met21 radiation shield used at LG combines white outer lamina and black inner louvres that prevent direct solar radiation reaching the sensor from any direction. The manufacturers specification reports that comparative studies between this type of radiation shield and an aspirated sensor indicate that at wind speeds <1m s^{-1} with intense solar radiation, measured air temperature can be 0.5°C above that recorded by an artificially ventilated sensor. At LG, times with low wind speed coincide with overcast skies, and so in the absence of concurrent ventilated temperature and humidity measurements required to explicitly evaluate any impact of radiative heating on the sensor we have assumed that the radiation shielding is adequate. Comparison of the 2°C-binned hourly air temperature (*T*) from the Vaisala sensor with that of a Campbell Scientific 107 thermistor installed within the same radiation shield (Mölg et al., 2008) indicates that differences between the two sensors exceed the error margins of the sensors in only 0.02 %

cases, so measured temperatures appear unaffected by the sensor membrane."

7. P5187 line 3: so any correction for SWI has been applied in case of riming or snow?

Yes, for 212 of the 30-minute data points (0.6%) SWI was corrected.

In the text we changed "*SWI* was screened for riming or snowcover on the upper sensor " to "At high elevation sites such as this, measured SWO can be higher than SWI when solar zenith angle is high, outside of these hours SWO in excess of SWI was taken to indicate snow cover or riming on the upper sensor and in these cases (0.6% of the data) measured SWI was replaced with SWI divided by the typical snow albedo over the measurement period."

8. P5187 line 6: How often has the AWS been visited and has any tilt been detected along the measurement period? This information is may be important to justify that no geometrical correction has been done.

The station was visited at least every 6 months and no tilting was observed. Clear sky days prior to the mast breaking on 20th July, follow the same hourly pattern of incoming shortwave radiation as the rest of the data series so we do not suspect any tilting prior to this mast failure.

Added "…without any geometrical corrections, as twice yearly visits to the AWS indicated that the sensors remained level throughout the measurement period."

9. P5187 line 6: how often are there gaps for V?

The anemometer was frozen, or disconnected for maintenance, for 42 of the 30-minute datapoints (0.1%) and we added this information in the text.

10. P5189, line 7: which density is used here to convert daily snow accumulation into accumulated mass? And the daily accumulated mass is divided into all hours of the day. However, the data logger is recording half-hourly values of SR50 sensor (p5186, line 15) so is there a more accurate way to distribute precipitation along the day? Considering that melting and accumulation are sometimes concomitant on such tropical glacier, this might have a non negligible impact on the results.

We had previously experimented with using derivations of hourly accumulation data and also daily data distributed only in hours with high humidity, but we found that, contrary to our expectations as well as those of the reviewer, these did not significantly impact the modeled results. We hypothesize that minimal effect of the timing of the precipitation may be related to the fact that we also use a daily derivation of albedo. Nevertheless as we were re-running the energy balance model with a more appropriate bottom boundary condition, for these re-runs we used daily accumulation distributed only within hours for which RH > 90%.

Daily accumulated snow height is converted into accumulated snow mass using a fixed value of fresh snow density that is one of the optimized model parameters. We changed the text to read: "Daily snow accumulation height is converted to accumulated mass using a parameterized value for fresh snow density and the daily mass accumulation is divided through all hours of the day for which RH > 90 %."

We used daily time scales despite the 30 minute availability of data as SR50 data is prone to spurious noise, which is minimized if we take data only from the surface height at midnight. This is described in section 2.1: "Daily snow accumulation height was computed as positive surface changes between successive midnight surfaces, taken as the mean of half-hourly measurements from 22:00-02:00 inclusive, which minimizes noise contamination of the signal."

We added the caveat that: "This approach might under-represent accumulation is snowfall in a given day is subject to significant ablation or compaction within the same day."

11. P5189, line 9: what is the glacier body temperature below 3 m? Is the glacier temperate? In Table 2 (last row), it

is said that “the ice was assumed to be near the melting point” although T(ice)=-3 °C at initialization? Why?

We did attempt to measure ice temperatures within Lewis glacier to a depth of 2m. However, the available sensors were old and of unknown quality as we had no history of their usage and they had not been recently recalibrated by the manufacturer. The sensors were assessed for accuracy and relative bias by (i) immersing them in iced water – all sensors read -0.16°C, and (ii) measuring room temperatures under the same conditions – mean standard deviation was 0.2°C. The ice temperatures recorded at Lewis glacier were between -0.16°C and -1.51°C. We chose not to present this ice temperature data, are unsure of the quality of it.

Consequently, as we did not have any measurements of the internal ice temperatures we first aimed to be cautious, by allowing the bottom boundary condition to vary over a wider range from -0.5 to -5°C. On reflection, this is probably too cautious in the light of the findings of Thompson, L. G. (1981) who measured firn temperatures of 0°C at 0.5m indicating that the ice was temperature in 1978, and in the light of the temperatures we measured. Thus, we recomputed the model runs allowing the bottom boundary condition a more restricted range of 0°C and -2°C, which is more reasonable in relation to the available evidence. In the repeat runs, initial ice temperature was set to be a fixed value equal to the surface temperature at the time of initiation as determined from the measured LWO. However, bottom temperature usually has little influence on the modeled mass balance (e.g., Table 3 in Mölg et al. (2009), and the sensitivity study now performed in this paper)

12. Table 2: in the last column, some references or field measurements are referred to support the results of the parameter optimization. For references, in some cases, it is easy to select other references that disagree with the parameter values obtained by optimization. For instance, there is a large scattering of z0 values (e.g. see the discussion regarding roughness parameters in Hock, Progress Phys. Geogr., 29(3), 362-391 2005 or see the standard deviation obtained for z0v and z0v on Kilimanjaro summit – table 2, Cullen et al, Ann Glaciol, 46, 227-233, 2007) and turbulent fluxes are very sensitive to z0 values. For field measurements, how fresh snow density measurements were conducted? It is not so obvious to obtain accurate fresh snow density measurements, although it is important to have reliable measurements to convert snow height increase into accumulated mass. The density of fresh snow reported here is 400 ±50 kg/m3 (from optimization) or 330 to 430 kg/m3 (from field measurements / the value of 420 kg/m3 is reported p5201, line 23). These values are very high, and, contrary to what is written in table 2, higher than the values reported by Sicart et al (2002) (i.e 250 kg/m3) which already were high. As a consequence and as already pointed out in General comments, a sensitivity analysis spanning a range of values much larger than the error range given in table 2 should be performed here, to know how sensitive the SEB results are to all model parameters.

Roughness values for fresh snow were taken from Brock et al. (2006), as stated. Both Brock et al. (2006) and Hock (2005) highlight the wide range of values possible, and the values quoted within these papers are also derived from different methods which makes it difficult to assess the most appropriate values. Cullen et al., 2007, do not specify what the surface type is on the Northern Ice Field during their experiment, but it was ice with shallow undulations, so we applied wider ranges to both old snow and ice roughness in our study in the light of their findings. The sensitivity study indicated that the modeled mass balance is relatively insensitive to changes in the roughness parameters at this site, but in the light of this comment when we re-ran the Monte Carlo optimization we allowed a wider range of roughness to be sampled for snow surfaces:

Roughness parameter [mm]	**initial runs**	**repeat runs**
z0h/m fresh snow	0.2 ± 0.1	0.5 ± 0. 4 (i.e. 80% range; 0.1 – 0.9 mm)
z0h/m old snow	4 ± 2.5	5 ± 3 (i.e. 60% range; 2 – 8 mm)
z0h/m ice	15 ± 5	15 ± 5 (i.e. 33% range; 10 – 20 mm)

Snow density was measured with a standard sampling tube and scale. Snow cover was not always present when we visited the glacier, so these data come from 10 snow sampling pits, which include 7 samples of fresh surface snow. The small depths of fresh snow generally found at Lewis Glacier mean that for some fresh snow samples the sample tube was only partially filled, but great care was taken to measure the proportion of the sampling tube that was filled as accurately as possible.

13. Table 3: the authors should provide the period covered here, and given that there are gaps in the data series, it would have been interesting to provide annual values, for a complete year (i.e. oct 10-11) as well as mean values for all data.

Done.

Caption now reads: "Summary of 30 minute meteorological data measured at LG AWS. Unshaded rows are computed for all 773 available days between 26 September 2009 - 22 February 2012 (missing data between 25 January to 2nd March 2010 and between 20 July to 29 September 2010) and shaded rows refer to a single annual cycle 01 October 2010 to 30 September 2011. This data was used to compute the daily values shown in Figure 2 and Table 4. Abbreviations are given in the text in section 2.1 and values in parentheses for the SWI parameter were computed on daytime values only."

14. P5191, lines 24-30 and Figure 3: there are data gaps for surface height, due to rotating mast or broken mast. After each data gap, the series starts again at 0, it would be worth mentioning it and also add an horizontal line at 0 on Fig3a to make it clear. Otherwise, we have the impression that it is a continuous series with some missing parts. Are there any ablation stakes nearby to help to reconstruct the surface height during the gaps? And if there are some, it might be interesting to show their records, to compare with SR50 measurements.

This choice of presentation was solely to minimize the vertical extent of the figure as it makes it easier to see the correspondence or mismatch between the model and measured surface height.

We now plot the real surface height evolution over the modeled period instead.

15. P5192, line 4: it might be useful to refer also to Fig 2a-h (and not only 2i) along the text in this section.

Done

16. P5192, lines 10-12: which period for the 800 mm value given by S Hastenrath? Obviously outside the TRMM data period, 98-2012, so the comparison does not match the same period, which is not so much a problem I believe, but worth mentioning it.

We changed the reference from the 1984 data source to the 2005 data source from Hastenrath in order to use data from all available rain gauge readings, and the text now reads: "Accumulating precipitation gauges at the elevation of Lewis Glacier indicate mean annual precipitation of *ca.* 870 mm (standard deviation of 270mm) from 1979-1995 (Hastenrath, 2005), which compares well with the mean annual value of 855mm for 14 full years of TRMM data (1998-2011) at the closest grid cell to Mt Kenya."

17. P5192, line 9 and Fig 3: -2.55 m is not clear from Fig 3. Is it without considering positive surface height changes, and only considering surface lowering? Because from fig 3, it looks like that the net surface lowering is rather closer to -2m than -2.55m

This apparent discrepancy is because the initial figure representation did not include the surface height change ongoing during the data gaps. We now plot the real surface height evolution over the modeled period instead, so this is no longer an issue.

18. Fig4 and Fig5 : it might have been useful to provide somewhere in the paper (may be above Fig4?) the energy balance equation with the flux notations used in these figures, for sake of clarity. With this equation, we could have understood better the sign of the energy fluxes, and especially the sign of QM (negative when there is melt). On

overall, I think both figures are not very convenient to read. It would have been more convenient to plot on the same subplot all terms of the energy balance equation, so that the reader can visualize easily QM, as the algebraic sum of all the other terms of the same subplot. For instance, on the first column (i) of Fig5, adding Net SW and Net LW would have facilitated the visualization of full SEB, and the resulting term QM.

The sign convention used is conventional in that all fluxes are positive towards the surface, and QM is therefore negative as it consumes energy (i.e. removes energy from the surface).

We choose not to plot the radiative fluxes alongside the remaining fluxes as the scale difference between radiative and non-radiative fluxes makes it difficult to see the variation of the non-radiative fluxes.

19. Table 3 and Table 4: It could have been interesting to provide mean values for all the energy fluxes of the energy budget equation, and not only SWI and LWI. (not only for LG, but also for the other glaciers in Table 4)

We prefer not to do this as the models used to compute the surface energy balance terms are not the same across all sites and so providing a direct comparison here could be misleading.

20. P5195 line 2: how can you explain that QS <0 between 17:00 -21:00? It looks strange because usually, at sunset, surface temperature decreases more rapidly than air temperature (due to the energy loss through net LW), making the near surface gradient positive and so is QS.

The slightly negative QS values after sundown are most likely to be a model artifact caused by the fact that the modeled glacier temperature field and surface temperature shows a slight lag compared to the surface temperature derived from measured LWO. In general though the agreement between modeled and measured surface temperature is good and the correlation between them at the hourly timescale is 0.77.

Measured air temperature and glacier surface temperature derived from the LWO record are within error of each other, but the LWO-derived surface temperature also exceeds the measured air temperature for a short period in the early evening.

We have removed the reference to the time period from the text as this may be misleading, and we now write: "The additional mean surface energy contributions from sensible heat flux is 8.4 W m^{-2} and mean sensible heat flux is almost always positive."

21. P5195 line 27: it is Fig5c-e and not 4c-e

Thank you, this has been corrected.

22. P5197, line 5. There is still a 40% difference in LWI between dry and wet seasons. Consequently, there is not "only a very slight" difference in LWI at seasonal time scale. And this difference is almost as high as for outer tropics glaciers ZG and ARG (P5199 lines 9-11); may be a short discussion regarding this point and a comparison with other inner tropics glaciers could be interesting.

Sorry, this reported number was an error on our behalf, as can be seen by comparing the LWI bars plotted in Figure 5 in the original submission. The 40% change is a reduction in the magnitude of net longwave, not a 40% increase in the incoming longwave as we originally wrote. In fact mean incident longwave is increased by only 11% in the wet conditions and 14% in the warm/wet extreme compared to the all time mean.

This has been corrected in the text, which now reads: "During the wet and warm/wet extreme conditions at LG, *LWI* is only slightly elevated in comparison to the mean *LWI* over the whole period, however, enhancement in *LWI* from the clear/dry extreme to the wet and warm/wet conditions is 42 % and 46% respectively, which is between the upper limit of cloud cover *LWI* enhancement for mid-latitudes and the humid season enhancement typical for outer tropical glaciers (Sicart et al., 2005)."

We also added: "At LG, the most evident perturbation of *LWI* occurs during brief intervals of sustained clear sky conditions which indicate that the typical cloud conditions serve to elevate atmospheric longwave emissions by between 14 - 45% compared to that of clear dry atmospheric conditions, while at ZG the presence of clouds during the humid season increases *LWI* by >50 %." at the end of section 4.1.

23. P5198, line 26: Wagnon et al (1999) deals with SEB on Zongo Glacier and does not support the statement here, which concerns only AG.

Thank you, this was an error and has been removed.

24. P5200, line 3: Table 3 should be removed from the () since it does not deal with lapse rates.

Done.

25. Fig7 : blue dots are barely visible

Dots have been enlarged and are now in red for clarity

26. P5202, lines 1-2: same as comment 12. I agree that due to temperature, fresh snow density might be higher at LG than at KG, but the difference reported here is rather high 420 kg/m3 against 255 kg/m3. How was these densities measured? How many measurements?

See response to comment 12.

27. P5203, lines 12-16: Do the authors have an idea of the elevation of the rain-snow limit? Actually, the mean temperature recorded at the LG AWS site is close to 0°C, and the rain-snow limit might rise in a near future at the elevation of the glacier, which will severely affect its life expectancy.

No, we don't have any information on the rain-snow limit. The glacier is so small that it seems unlikely that different parts of the glacier experience solid or liquid precipitation.

28. P5203, line 26: Wagnon, 1999 is Wagnon et al., 1999

Thank you, this has been corrected.

29. P5204, lines 20-22: looking at Fig5 (ii) I believe that the net LW difference between dry and wet months play a significant role in the SEB, and in turn, on the ablation melting. This effect is probably not as important as on outer tropics glaciers of South America, but it is significant on LG which shows a kind of intermediate behaviour between AG, EG, and ARG or ZG.

We have added: "ARG and ZG (Favier, 2004a) and KG (Mölg et al., 2008) show clear seasonal offsets in net longwave radiation in the order of 50-60 W m^{-2}. Seasonal offsets in net longwave radiation are more muted at AG, where differences are in the order of 20Wm^{-2}. At LG the magnitude of monthly mean net longwave radiation is generally most negative during the JF season and least negative during ON (Figure 5a). The range in monthly mean net longwave is greater at LG than at AG, and approaches that of the seasonal offsets at KG, ARG and ZG, although the maximum and minimum net longwave conditions at LG are expressed within single months rather than sustained seasons. " and "At LG, although the net longwave does vary throughout the year, this is more than compensated for by concurrent changes in the net shortwave, and the correlation analysis indicates that melt energy is more influenced by the shortwave than the longwave component of the radiation fluxes."

References:

Georges, C., & Kaser, G. (2002). Ventilated and unventilated air temperature measurements for glacier-climate studies on a tropical high mountain site. *Journal of Geophysical Research*, *107*(D24), 4775. doi:10.1029/2002JD002503

Mölg, T., Cullen, N. J., Hardy, D. R., Kaser, G., & Klok, E. J. (2008). Mass balance of a slope glacier on Kilimanjaro and its sensitivity to climate. *International Journal of Climatology*, *28*(7), 881–892. doi:10.1002/joc

Mölg, T., Cullen, N. J., Hardy, D. R., Winkler, M., and Kaser, G. (2009) Quantifying climate change in the tropical mid troposphere over East Africa from glacier shrinkage on Kilimanjaro, J. Climate, 22(15), 4162-4181.

Sicart, J. E., Wagnon, P., & Ribstein, P. (2005). Atmospheric controls of the heat balance of Zongo Glacier (16 degrees S, Bolivia). *Journal of Geophysical Research*, *110*(D12106). doi:D12106 10.1029/2004jd005732

Thompson, L. G. (1981). Ice core studies from Mt Kenya, Africa, and their relationship to other tropical ice core studies. *IAHS Publication* (Vol. 131, pp. 55–62).

Response to reviewer 2

General comments:

The only general comment on the paper is that we ought to exercise more caution in making generalizations based on the short dataset available, and the apparent variability of the seasonality.

This is a good point, and we have reviewed the text for cases when we should re-assert this caveat. For instance we removed reference to the sub-sampled wet and dry conditions as 'seasons' as we are not sure if they represent 'typical' seasonal conditions.

Specific comments:

A. p. 5184, line 14: My interpretation of "short-term" here would be roughly 6 months or less (i.e., less than an annual cycle). If measurements exist spanning a longer time period, I suggest mentioning a time period – as they could prove useful for further work.

From the published literature it appears these studies consisted of only a few days at a time, 2-15 April 1960, 14 days in 23 July – 5 August 1975 and 5-8 February 1983 and 27 February – 02 March 1983.

We now write: "Although periods of a few days measurements of some meteorological variables exist from April 1960, July-August 1975 and February-March 1978 (Platt, 1966; Davies et al., 1977; Hastenrath and Patnaik, 1980; Hastenrath, 1983)."

B. p. 5186, beginning of section 2.1: I suggest a relatively-recent, aerial oblique view of Lewis Glacier, in the context of Mt Kenya's upper slopes would be helpful as figure 1. This would nicely show the topography and setting for the study site as only a photo can, and given that the paper reports new data from a new site I think it would be warranted.

A photo of Lewis Glacier has been added as an inset within Figure 1.

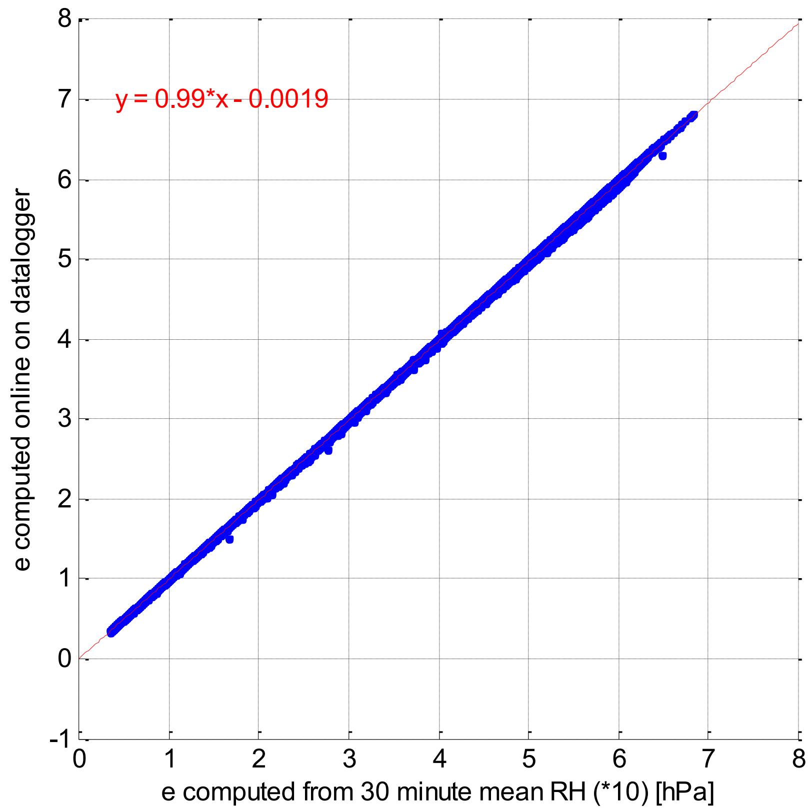
C. section 2.1, second paragraph: The discussion of T/RH should begin by informing the reader – in text or the table -as to what type of radiation shield houses the sensors. I assume it is a multiplate, naturally-ventilated shield, in which case errors due to radiation loading likely overwhelm those due to the sensor membrane. These errors would impact both the thermistor and the Vaisala sensor. Elsewhere the paper reports a median wind speed of 2.5 m/s, and diurnal speeds that are ~1 m/s lower. Under such conditions over snow or ice, there will be considerable error in maximum daily temperatures for both instruments. However, this error will likely only influence daily means on the calmest days with fresh snowcover.

Shield type was added in station detail in Table 1 and in the text.

See response to Reviewer 1 – specific point #6 for details on our response regarding radiation loading.

D. Line 25 of the same paragraph: please state whether e is computed after each T and RH measurement, prior to averaging, or done from the 30-min averages of each in post-processing. With rapidly-changing values of both due to large diurnal fluctuations, these will yield different values of mean vapor pressure.

Initially our AWS recorded only 30 minute mean RH from which we computed 30 minute mean vapor pressure using 30 minute mean air temperatures, on reflection we also thought this is not correct and so the AWS program was altered in March 2011 to additionally compute vapor pressure at each 1 minute data scan and store a 30 minute average of vapor pressure. Thus both methods of computing 30 minute averages of vapor pressure are available for 40% of the record (n = 17160 30-minute average values for comparison of the methods). We found that the relationship between the 30 minute averages of vapor pressure computed from 1-minute vapor pressure values as compared to from 30 minute means of T and RH are almost perfectly 1:1.



E. p. 5187, line 1: why not just delete negative values of SWI and SWO rather than use a TOA time series to exclude nocturnal readings? Otherwise, how can a TOA time series account for cloud reflection, diffuse radiation and other effects of low sun angle?

Both positive and negative values exist in the nocturnal readings. We excluded them all by making a time series of top of atmosphere radiation, and extended daytime by one hour in each direction to account for dawn and dusk light and then set all times classified as night time to 0 in both shortwave series. We then set all remaining values that were <1Wm^{-2} to 0 as well. Arguably the second step would have sufficed.

We now write: "Nocturnal values of *SWI* and *SWO* were set to 0, as were all value pairs for which *SWI* < 1Wm^{-2}."

F. same page, line 15: As I'm sure the authors are aware, snow cover on the radiometer dome can create a situation where SWI is <35% of the clear sky value. Perhaps a model should be considered which also looks at SWO before defining conditions as overcast, although there may not be much difference.

The data was screened for riming or snowcover effects prior to assessing the sky condition. This is now described more fully in the text where we write: "At high elevation sites such as this, measured SWO can be higher than SWI when solar zenith angle is high, outside of these hours SWO in excess of SWI was taken to indicate snow cover or riming on the upper sensor and in these cases (0.6% of the data) measured SWI was replaced with SWI divided by the typical snow albedo over the measurement period."

G. p. 5189, line 13: It would be useful to indicate (perhaps parenthetically) what the proportion was of "input parameters that were poorly constrained by field data", and thus optimized.

We added a further reference to Table 2, in which all the parameters that were optimized are listed.

H. section 3.1, second sentence: I suggest a different word than "clear" as this is too vague, in light of the relatively-short measurement period. In subsequent text there are numerous examples cited of seasonality in many of the variables. While I appreciate that the authors wish to stress that low latitudes have generally less seasonality than many readers may be aware, I suspect that monthly or daily means over >3 years would reveal a greater degree of seasonality. The word "clear" seems too subjective.

We changed 'lack clear' to 'weak and inconsistent'.

I. p. 5190, line 26: I suggest that RH > 99% may be too precise in defining saturation conditions, especially with sensor accuracy when new of +/-3%. This is a minor point perhaps, but would saturation conditions be reached in 5% more sampled days with a RH value of 97%? 10% more?

Good point. We recalculated this with a RH > 95% threshold. With this criteria for saturation >55% of days reach saturation between 16:00 -18:00 and > 40% of days reach saturation between 14:00-19:00.

Now reads: "Saturation conditions at the AWS (defined as *RH*>95 %) can occur at any time, and are reached in >40 % of the sampled days between the hours of 14:00-19:00 ..."

J. p. 5191, line 26: The meaning of "enhanced accumulation" is not clear. Does this mean greater than normal (average), or no accumulation? I believe that the 2011 long rains failed completely, bringing widespread drought to much of Equatorial East Africa (termed "catastrophic" by some). See also p. 5201 line 27; what is "slightly elevated" accumulation? On Kilimanjaro, the 2009 short rains were distinctly above normal.

Good point, the way we had state this was unclear. We meant accumulation rates and amounts elevated above the all-month mean and this is now stated explicitly.

We now write: "Visual inspection of daily snow depth accumulation rate averaged over weekly and monthly intervals show that only May of the 2010 long rains brought accumulation rate above the all-month mean rate at the AWS, and that accumulation during the long rains in 2011 remained below the all-month mean rate. In contrast, of the 3 short rains periods sampled, at least 2 months of each brought snow accumulation above the all-month mean. Additionally, accumulation was above the all-month mean during the dry season months of January 2010 and August and September 2011, when the accumulated snow depth equaled that of ON 2010." and "These ratios are partly affected by the strength of the wet season at each site: at KG, the OND 2009 rainy season brought accumulation only slightly elevated above the all-month mean, and, at LG, the MAM 2010 season was relatively poorly expressed, with accumulation concentrated in May."

K. p. 5193, ~line14: As an example of the caution required in generalizations based on limited time periods (i.e., General Comment above), October 2010 was chosen to represent "typical wet season conditions" – yet on Kilimanjaro this was a rather dry interval of little precipitation, low albedo, and high ablation.

We agree that it is difficult to find representative conditions within this short and variable record. We chose these 'representative' periods on the basis of finding a period of high humidity for 'wet' conditions and low humidity for 'dry' conditions on the basis of the daily mean data given in Figure 2. The 'standard' conditions are also not anomalies in the TRMM precipitation and reanalysis temperature data.

We provide more qualification of this in the revised text where we write: "In order to assess how surface energy balance differs between the seasons it would be desirable to be able to identify typical wet and dry season conditions. However, this is challenging given the short and highly variable meteorological records. On the basis of the temperature, cloud, and humidity conditions (Figure 2) and on the representativeness of the measured period in comparison to the mean monthly conditions, sub-samples from 1 - 20 October 2010 and 1 - 20 July 2011were selected to represent 'wet' and 'dry' season conditions, respectively. In addition, 26 April – 15 May, 2010, (20 days) and 19 January – 7 February, 2010, (20 days) were selected to represent extremes of warm/wet and clear/dry conditions, and were chosen on the basis of coinciding with thermal optimum and humidity minimum in the measurements period respectively."

L. p. 5197, line 23: why not use Hastenrath (2005) to assess the relative reliability (1981-90), or to corroborate the short record from LG?

We wrote: "at KG, the long rains appear to be more reliable and bring the most rain, while the short rains are variable and, thus, exert a strong control on inter-annual glacier mass balance (Mölg et al., 2009a); conversely, in the short record available here, the short rains at LG appear more reliable than the long rains."

We now add: "In the historical precipitation records from rain gauge measurements (1978-1996), MAM appears to be more variable than OND with mean MAM precipitation of 314mm (st. dev. 139mm) and mean OND precipitation of 302mm (st. dev. 100mm)."

M. same page, line 28 referring to JF snowfall on Kilimanjaro: this is unclear and/or inaccurate, because the JF period is fundamentally different from the JJAS dry season there. It is much shorter than 2 months in duration, which may account for the notion that is comparable with wet seasons; often the short rains extend into January, for example.

We now write: "During the shorter JF dry season on Kilimanjaro, the short rains often extend into January and snowfall at the summit of Kilimanjaro is comparable with that of the regional wet seasons (Chan et al., 2008). Higher JF accumulation here could also be a result of moisture supply from the northern edges of the tropical rainfall belt in its most southerly position, augmented by enhanced Atlantic moisture transport to east Africa in some austral summers (e.g. Whittow, 1960; McHugh, 2004)."

N. p. 5198, lines 13-18: this is a wonderful sentence, beautifully summarizing the situation in EEA!

Thank you.

O. p. 5204 at top: I question the wisdom of stating that 65% of mass loss on KG is due sublimation, consuming 94% of the energy for ablation. These values are from point models representing short intervals of time. They may indeed be accurate for a particular point over a particular interval, but 10 m in any direction over a different time period one is almost certain to get differing values. So is it reasonable to characterize and generalize these important processes with such precision? I don't have an answer, for the numbers are what the robust model provides. Perhaps some form of caveat could be used in such cases, e.g., "point modeling suggests that" or "about 2/3" or "most of the energy"?

We now write: "….and previous studies found sublimation to be responsible for two thirds of mass loss and consumption of almost all the energy available for ablation."

P. Table 4 shows the periods from which data from other AWS are used, and it appears that these are ~1 year for all the South American sites. These data sources are discussed on p. 5187 in the final paragraph (i.e., line 28). Since one year is a very short interval for sites with high interannual variability, please consider adding a caveat to this effect. Otherwise, the reader may not realize that only short portions of the relatively-long ZG and AG records are used.

To this sentence we added that: "…, and some consider only one year from a longer record,…"

Technical Corrections

1 p. 5187, line 28: Table 4 is referenced after Table 1 but prior to Table 3

Table 4 has now been promoted to Table 2.

2 consider more precisely specifying what “ERA-interim” wind fields are, although a quick Google search will reveal this to any reader

We added a further explanation.

3 p. 5194, line13: I think you need a comma after “freezing” – unless I am misinterpreting the sentence

Done.

4 p. 5205, line 14: Especially in the Conclusions section, the clarity of wording is important. I suggest “…meteorological conditions high on Mt Kenya…” rather than “…on the summit of…”, even though the station is within 400 m of the summit. Likewise, I suggest “…little variability on an annual timescale, in accordance…” to clarify that it is temporal variability being discussed. Finally, I think a comma after “…regional hygric seasonality” would be helpful.

Done.

5 p. 5205 line 18: I hope there is a typo in that “…whereby JJAS (JF) is…” should be “…whereby JF (JJAS) is…” – because JJAS is decidedly more arid on Kilimanjaro than JF. Indeed, if based on mean monthly vapor pressure, the short dry season on Kilimanjaro can be more narrowly defined as February only.

Thank you – we have corrected this.

6 Table 2: How can the permissible value range for “% of refreezing meltwater forming superimposed ice” be 0.3 +/-20%?

The range is ±20% of this central value (0.3), so the range is 0.24 – 0.36.