Response to review comments by referee 1

We thank the referee for making this thorough and useful review of our paper. Below we give a point-by-point response to the review comments.

The page and line numbers indicated by the referee do not correspond to the version of the manuscript on the discussion site. Nevertheless, we are confident that we have identified the parts of the text that the comments are concerned with.

General Comments:

Page 7 line 5, Page 12 line 24. "nominal" usually refers to the at launch calibration in a lot of the AVHRR papers published. I would persuade the authors from using this term to describe the existing calibration used to compare the inter-calibrations used in this study. I suggest the following.
1) "For MODIS the nominal calibration" replace with "For MODIS the collection 5 calibration is provided …"
2) "For AVHRR we use the Heidinger et al. 2010 calibration referred to in this study as the "Heidinger calibration"
3) "For SEVIRI the operational calibration by EUMETSAT is used and referred to as in this study as the EUMETSAT calibration"

Agreed. We will remove the term 'nominal'.

Page 7 line 19 and Page 12 line 18. The word "re-calibration" to identify the regression slopes defined in this paper is confusing. It implies an iterative process. Also many papers define calibration as an absolute calibration method, such as deserts, and intercalibration as the calibration referenced to a well-calibrated contemporary sensor. I suggest to use the word "inter-calibration" to describe the calibration method and the calibration slopes in this paper. Some sections of the text already follow this standard, for example section 4 heading is "Inter-calibration of SEVIRI and MODIS" in section 6.

The reason we used the word 're-calibration' is that we report gains with respect to reference calibrations (termed 'nominal' before) rather than to pixel counts. However, since 're-calibration' may be confusing, we will switch to 'inter-calibration'.

Section 3.1. The correcting for differences in spectral response. In this description the first paragraph has to mention clearly that the only spectral correction being made is the atmospheric correction. Also state that the spectral correction is derived from two components, surface spectral signature and atmospheric absorption. The underlying surface is assumed to be Lambertian and a constant spectral reflectance based on the measured reflectance with the atmospheric absorption removed. The absorption is only computed above either the cloud effective radiating temperature or surface. Essentially Page 10 line 20 paragraph and page 15 line 22 sentence, should go first to divide the spectral correction into two components, surface spectra and atmospheric spectra.

We will follow these suggestions, and state the two components at the beginning of Section 3.1.

I am confused about the TCWV. The description indicates that the water vapor absorption was computed in a look up table (Page 9 line 15). Yet in the next paragraph (Page 9 line 23) the TCWV was computed using ERA-Interim climatology. Then the water vapor above the cloud was determined from the McClatchey tropical profile (Page 9 line14). Please clarify how the water vapor absorption was determined.

What we intended to explain is: (1) TCWV is an independent variable in the LUT, and (2) the actual TCWV for a pixel (needed to read the actual SRF correction from the LUT) is taken from ERA-Interim. The RT simulations performed to generate the LUT take into account the above-cloud column water vapour, which is calculated using TCWV and cloud-top height in combination with the McClatchey vertical water vapour profile shape. We will explain this more clearly in Section 3.1 in the revised version.

Page 10 and line 14 then the MODIS ch18/ch17 ratio is well correlated to the water vapor absorption in Figure 2. Please clearly state the intent of the sentence on Page 10 line 14 "Indeed, a good linear correlation is obtained, demonstrating the validity of the water vapour correction both for clear and cloudy pixels." Is the spectral atmospheric correction algorithm used in this study validated by Figure 2. Or was this a separate study of simulated MODIS SEVIRI ch2 reflectance ratios? Or was it used in Figure 3,4,5? Please clarify.

The figure is meant to be a validation of the atmospheric correction algorithm. We will clarify this in Section 3.1 in the revised version.

Page 10 line 18. "This also gives confidence in the atmospheric correction method for other sensors, such as AVHRR, that do not carry channels with additional information on atmospheric water vapour." AVHRR channels do carry additional information. IR channels 4 and 5, 11 and 12µm can be used to determine the water vapor column. Even SEVIRI has these channels and are probably used operationally to compute TCWV. Here are some references.
Wu, A. ; Xiong, X. ; Angal, A., 2013, Derive a MODIS-Based Calibration for the AVHRR Reflective Solar Channels of the NOAA KLM Operational Satellites, Geoscience and Remote Sensing, IEEE Transactions on , Volume: 51 , Issue: 3 , Part: 1, DOI, 10.1109/TGRS.2012.2220780 (fig 2)
M. Schroedter-Homscheidt, A. Drews, S. Heise , 2008, Total water vapor column retrieval from MSG-SEVIRI split window measurements exploiting the daily cycle of land surface temperatures, Remote Sensing of Environment 112 (2008) 249–258

There are obviously other channels carrying information on atmospheric water vapour. However, these allow the water vapour column to be determined only in clear-sky conditions, and are thus not applicable for our inter-calibration approach, which relies on both clear and cloudy pixels. We will modify this sentence to: '… such as AVHRR, that do not carry the suitable channels for the determination of atmospheric water vapour columns in both clear and cloudy conditions.'

Page 11 MODIS band 2 saturation. Is there not a flag in the MODIS read code to filter the individual pixel level saturated radiances as not to contaminate the gridded level MODIS/SEVIRI reflectance pairs. This is a much cleaner approach then to have an arbitrary MODIS radiance threshold. The saturation of the land use channels is well known and the saturated pixels are simply filtered. I am concerned this might set a precedent to compare MODIS band 2 without removing the saturated pixels among novice researchers. I strongly suggest the authors to filter the MODIS saturated pixels and redo Fig 4,5, and Table 2.

We have used the MODIS 1-km L1b product (MYD021KM), which is aggregated from the 250-m product. If a 1-km pixel is completely filled by saturated 250-m pixels, a fill value is reported. These pixels are thus easily filtered. The problem occurs for the 1-km pixels that are partly filled by saturated 250-m pixels. In these cases, the non-saturated 250-m pixels are aggregated into the 1-km product, causing a bias in the 1-km reflectance.

In the full MYD021KM files, a pixel count is available, which could be used to filter 1-km pixels containing saturated 250-m pixels. However, to reduce the data volume we ordered post-processed

MYD021KM files, using the band subsetting option, and in these post-processed files the pixel count is not available. For complete filtering of saturated pixels we would have to re-order the full MYD021KM or MYD02QKM dataset, and redo the whole analysis.

However, filtering for $R_n < 0.6$ removes the vast majority of saturation-affected pixels, as shown in Fig. 4. Furthermore, applying an even lower threshold ($R_n < 0.5$) gives very similar inter-calibration statistics (e.g., a mean forced-fit slope of 0.937 instead of 0.940), as demonstrated in Table 2. Thus we concluded in Section 5 that our approach to deal with saturation works rather satisfactorily.

In conclusion, the (residual) impact of saturation on our inter-calibration results is very small, and re-doing the analysis along the lines sketched above is not required. Nevertheless, we will add the following sentences in Section 3.2:
'… land-surface applications. While individual saturated pixels could have been filtered out directly from the 250-m MODIS level-1b product, this is not possible for aggregated 1-km pixels in the post-processed 1-km level-1b files that we used, in case these pixels are partly filled by saturated 250-m pixels. Indeed, the impact of saturation is clearly visible in Fig. 4b, resulting in an overestimate of the slope. To take care …'

I am very surprised that land in the SEVIRI inter-calibration domain does not change the regression slope of the 0.86 channel inter-calibration especially between AVHRR or MODIS and SEVIRI. A typical vegetation spectra is available https://gsics.nesdis.noaa.gov/wiki/Development/20130304 under Fanfang Yu GSICS 2013 presentation on Thursday and shown here.

1) The gain difference between land/ocean and ocean-only is 1.2% in Table 2. Could it be the case the SEVIRI inter-calibration domain is dominated by ocean, so the difference is smaller? Or is land mostly dominated by bright clouds obscuring land? Please add another row in Table 2 that includes land only.

We have calculated statistics for land only (using an actual land-sea mask rather than a rectangular lat-lon box). The results will be included in Table 2. For channel 2 the land-only analysis results in a forced-fit slope of 0.932. This differs only 0.8% from the slope for all pixels and 2% from the slope for ocean-only. The number of pixels over land is roughly half that over ocean, so the domain is not strongly dominated by ocean.

2) The Terra overpass time also has greater probability of clear-sky land than Aqua-MODIS. Is the Fig4c MODIS/SEVIRI scatter plot similar for Terra and Aqua?

The scatter plots Figs. 4c and 4d for Terra are given in Figure A. Both are very similar to the Aqua versions. Note that the Terra plots have somewhat less pixels than the Aqua analogs, mainly because we gathered less Terra granules.

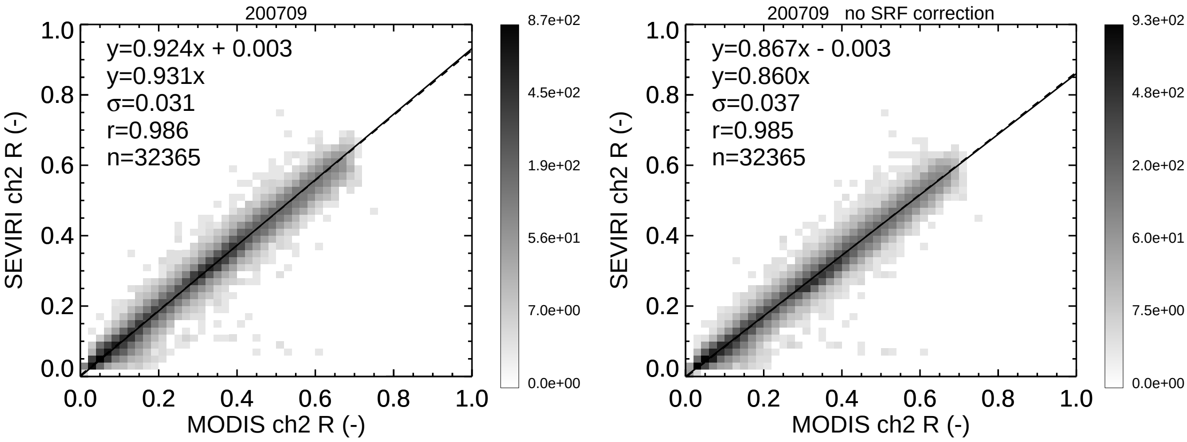

Figure A: Analogs of Figs. 4c and 4d from the manuscript, but now for MODIS-Terra instead of Aqua. Sigma is the standard deviation relative to the fit line.

3) There is a secondary concentration of reflectance pairs in Figure 4c that does not seem to be present in the other bands. Could the authors please comment where these points come from? It also seems there is a greater concentration of points to the left of the diagonal of Fig 4c, due to the MODIS saturation issue that would skew the slope. This paper almost makes the statement that the surface reflectance is irrelevant in inter-calibration, however the atmospheric correction can cause a considerable bias.

The secondary concentration of points for channel 2, with reflectance values around 0.3-0.4, is related to clear land surface (we verified this from a land-only scatter plot). A secondary concentration of points is actually also present for channel 3, but there it is located roughly between reflectance values of 0.2-0.3, and it cannot really be discerned from the ocean points with lower reflectance. We will add a short discussion of this in Section 3.2.

We did not intend to state that surface reflectance is in general irrelevant in inter-calibration. We really wanted to verify that our results are not significantly affected by the land pixels in the domain.

4) Can the authors please comment if this method could work to inter-calibration the 0.65µm and 0.86µm channels or broadband imagers such as Meteosat-7. Is it simply fortuitous that the 0.86µm bands inter-calibrated in this paper seem to be independent of surface spectra? See
Doelling, D. R., C. Lukashin, P. Minnis, B. Scarino, and D. Morstad, 2011: Spectral reflectance corrections for satellite intercalibrations using SCIAMACHY data. Geosci. Remote Sens. Lett., ***8****, doi:10.1109/LGRS.2011.2161751*

The method can be used for the 0.65 and 0.86 micron channels, as demonstrated in this paper. We did not apply it for broadband channels, as we focussed on SEVIRI. In principle it should be possible, but the uncertainty due to SRF correction may be larger. We will add a comment on this in the conclusions.

5) The AVHRR spectral response function begins at 0.70µm, whereas SEVIRI at 0.77µm. Page 16 line 21. In this case the AVHRR SRF is very broad. Could this explain the AVHRR/SEVIRI channel 2 Fig 7 result? Have you compared the ocean only and land only slopes for AVHRR and Meteosat? Is the gain difference between land/ocean and ocean different by only 1.2% as is the case with MODIS? I would like to see an equivalent Fig 4c for this pair for land only and also for ocean only.

The equivalent of Fig. 4c for SEVIRI vs. AVHRR (on NOAA-18) is shown in Figure B, as well as the ocean-only and land-only selections. The ocean-only and land-only forced fits differ by only 0.5%, i.e. even less than for MODIS vs SEVIRI.

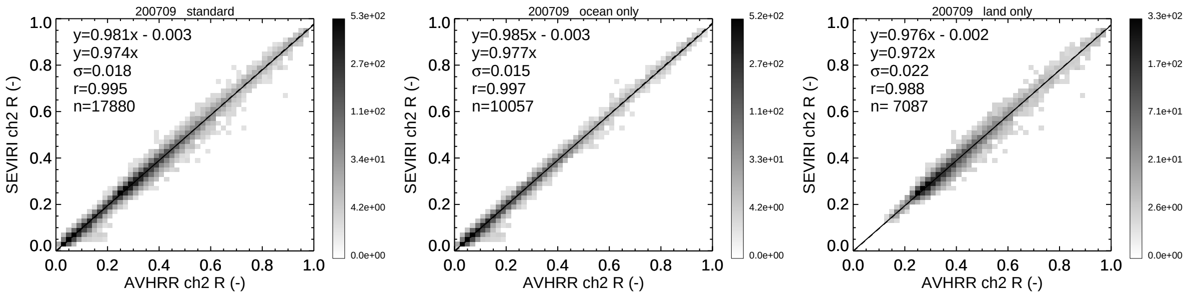

Figure B: Analogs of Fig. 4c from the manuscript, but now for AVHRR-NOAA18 instead of MODIS-Aqua and for land+ocean, ocean-only and land-only selections. Sigma is the standard deviation relative to the fit line.

Note that the correlations for these plots are higher than for SEVIRI vs. MODIS, which is related to the coarser spatial resolution (0.25 instead of 0.15 degrees), as was also apparent from the sensitivity experiments in Table 2. The number of points is much lower than for MODIS-Aqua, both because of the coarser spatial resolution and also because we gathered less granules.

To further address this review comment, we have compiled Fig. 7 for land and ocean separately (see the land-only version in Figure C). Differences in the mean slopes compared to Fig. 7 in the manuscript (which is based on land and ocean pixels) are smaller than 1%, trends are also similar, although a bit smaller. This confirms our finding that the impact of the spectrally varying reflectance over clear-sky land does not have much impact on the calibration results. However, the seasonal variation of monthly slopes in channel 2 is larger for the land-only selection, which may be related to seasonally varying spectral surface reflectance.

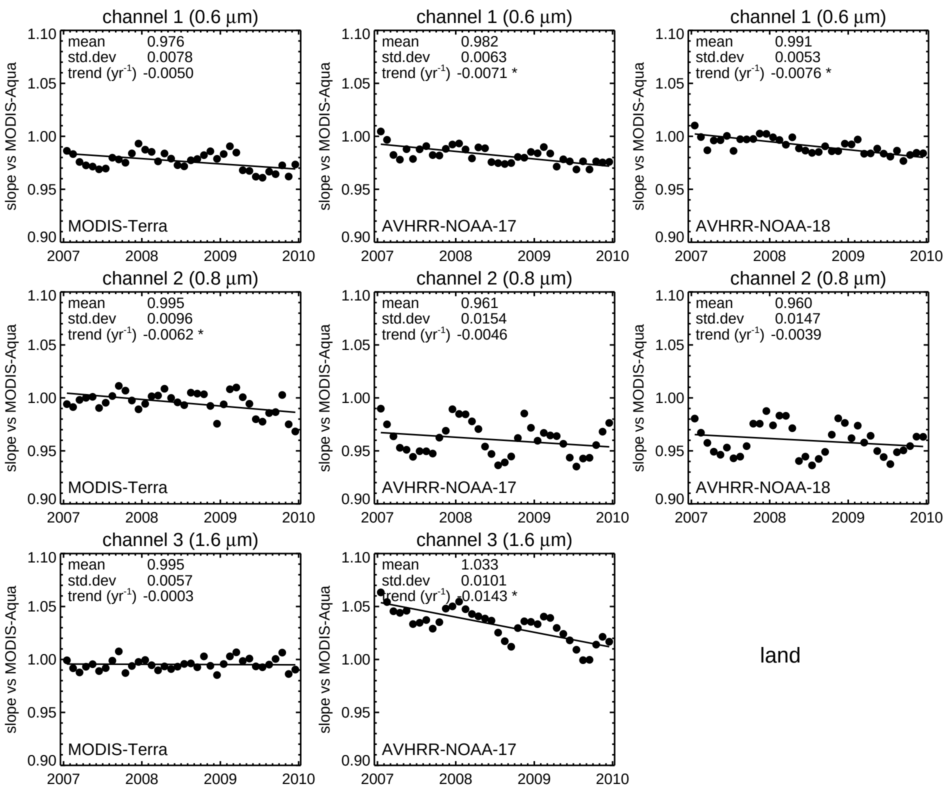

Figure C: Same as Fig. 7 in the manuscript but now for land-only.

A short summary of the ocean-only and land-only results for AVHRR vs. SEVIRI will be included in Section 6.

6) I would urge the authors to add in the conclusion that the spectral correction used is only an atmospheric correction and that caution must be used when intercalibrating clear-sky land.

We will include in the conclusion that we only perform atmospheric correction and no surface reflection correction. However, we also note that a strength of our method is that it does not only use clear but also cloudy pixels, so that the impact of uncertainties in land surface spectral reflectance is weakened.

Specific Comments:

Abstract line 20. The replace the word existing with Heidinger et al 2010 AVHRR calibration.
We would rather avoid including references in the abstract. With ‘an existing calibration’ we indicate that we looked at a particular calibration, but there are others that we did not consider. In the introduction we will replace the sentence ‘…, providing a means to verify existing calibrations based on direct SNOs.’ by ‘… In particular, the AVHRR calibration by Heidinger et al. (2010), which is based on direct SNOs, will be verified.’.

Abstract Page 2 line 10. Replace “off” by “offset”
OK.

Page 3 line3. Replace “using carefully selected targets” with using well-characterized targets”
OK.

Page 6 line 29. Replace “AVHRR observations” with “AVHRR nominal observations” (see general comments)
OK.

Page 7 line 8. Is there a reference for the official EUMETSAT calibration?
References for the calibration methodology are given in the paper. References for the actual calibration coefficients are not available to our knowledge. However, the calibration coefficients are contained in the SEVIRI level1b files. We will mention that in the revised version.

Page 8 line 15. How are the slope-only fits with fixed offsets performed? For example Figure 3.
These are performed as ordinary linear least-squares fits, but forced through the origin. Orthogonal fits are used just both for the free and forced fits. We will add this in Section 3.

Figure 3. I would suggest to title the reflectance pair scatter as all “without angle matching” and “angle matched”, to allow the user to easily identify the difference between the two plots.
Will be done. Thanks for the suggestion.

Figure 3. I suggest also providing standard error about the fit to qunatify the regression improvement in the plot statistics. All of the correlation coefficients are above 97% used in the paper.
The standard deviation with respect to the fit line will be included in the scatter plots in Figs. 3 and 4.

Page 13 line 14. Is the residual seasonal cycle in the monthly slopes perhaps due to a seasonal variation of the dynamic range, where some seasons do not have many bright

high clouds. The Meirink 2009 (fig 3) paper mentions that large SRF corrections are for low clouds, whereas smaller SRF corrections are needed for high clouds.
We did have a look at this, but found the seasonal variation of the dynamic range to be quite small. Thus, this was excluded as an explanation for the residual seasonal cycle in the monthly slopes. We will add this in Section 4.

Page 13 line 19 "The SEVIRI EUMETSAT reflectance is found .. "
This points to a more general issue. With statements like 'SEVIRI reflectance is too low' we refer to the EUMETSAT-calibrated SEVIRI reflectance. We will try to make this more clear in general.

Fig 5. Please clarify is the standard deviation the standard error about the fit? If the monthly slopes dropped rapidly due to instrument sensor degradation then the standard deviation would be large and not accurately assess the uncertainty of the trend as in the bottom panel of figure 6.
The standard deviation is indeed calculated relative to the fit line. This will be clarified in the text.

Table 2. It seems the results for geom. Criteria of channel 1, for $\Delta\theta < 20°$ and $\Delta\theta < 5°$ are reversed in the table. The greater angle limit should have more N than the restricted angle fit.
They are indeed reversed. This will be corrected. Thanks for pointing this out.

Table 2. add sunglint criterion, add rainbow criterion and add glory criterion, implies that the sunglint criterion is added to the previous row, etc. Yet the number of matches increases, indicating only one criterion is performed at a time. Please clarify.
In all rows only the mentioned setting is changed compared to the standard. Changes are not applied on top of each other. The word 'add' is confusing and will be changed to 'include'.

Table 2 The last column is the standard deviation of the monthly correlation coefficients. It is very difficult to quantify uncertainty with a correlation coefficient. I would rather have the standard error of the linear trend of the monthly slopes, similar to the standard deviations in Figure 5.
The standard deviation of the monthly slopes with respect to the fitted trend is already in Table 2 (column 6).

Page 14 line 21. It seems the scattering angle ranges listed are not to be avoided but are the valid domains, please recheck.
The scattering angle ranges are indeed the valid domains, and as a result the respective features are avoided. To make it (hopefully) clearer, we will rearrange as follows.
'We tested the following selection criteria:

- THETA < 135^o or THETA > 145^o: to avoid the cloud bow;
- etc. …'

Page 14 line 26. "the drawback is the number of grid points…" I do not know what this sentence means. With the standard criterion you get the same number of reflectance pairs monthly over the year? With the application of scattering angle the number of monthly grid points varies greatly?
With the standard criterion we indeed get a reasonably constant number of reflectance pairs per month over the year. Since the respective features (e.g., rainbow) occur only during certain months (because they require a particular sun-satellite geometry), these filters lead to a more variable number of reflectance pairs over the year. We will elaborate on this more in the manuscript.

Page 14 line 27. "For satellites with different overpass time" does this imply the AVHRR/SEVIRI inter-calibration pairs, since the NOAA satellite orbit is degrading in local time? Please clarify

This is a general statement, that with slightly different overpass times certain features (e.g. the rainbow) may be encountered much more frequently during a particular season, leaving hardly any pixels after filtering. Will be clarified.

Page 15 line 8. Why is this as expected? Is this a result of satellite navigation error or aggregation is now nearing a pixel or two per region? Please elaborate.
Aggregation leads to a decrease in standard deviation, and an increase in correlation. Therefore, increasing the resolution ('less aggregation') gives lower correlations, even irrespective of navigation errors. We will add an explanatory sentence in the text.

Page 15 line 16. The decreased dynamic range is only an issue with the Rn<0.5, since the average standard deviation of the Rn>0.1 is 0.0038 and is less than 0.0040 of the standard settings. (see comment above concerning the use of correlation coefficients)
These numbers are the standard deviation of the monthly slopes with respect to the trend line. The statement is referring to correlations (column 7), which are lower than in the standard selection for both Rn>0.1 and Rn<0.5 selections.

Page 15 line 22. The explanation of the how clear-sky is treated should be addressed in full in page 9 and then referred to here. (see general comments)
Will be done.

Page 16 line 10 and Page 18 line 25. It should be made clear here that the uncertainty is the inter-calibration methodology, based on sensitivity studies and not the overall intercalibration slope. The sentence that begins with "It needs to be emphasized" should mention specific systematic errors, such as the MODIS calibration uncertainty, the surface spectral signature, errors in the spectral response function, Met-9 solar constant estimate, etc. Could the authors please elaborate how the 1% and 1.5% uncertainties were derived?
The uncertainty estimates are indeed related to the inter-calibration methodology. They were derived from the maximum variation in slopes obtained in the sensitivity studies in combination with the standard deviation of monthly slopes, as outlined at the end of Section 5. We will add the specific systematic errors that are not reflected in these uncertainty estimates.

Page 18 line 1 Could this be a result of the degrading NOAA orbits as to alias the solar zenith angle dependencies into the inter-calibration results.
This is an interesting thought. However, there are two arguments against this:
(1) NOAA-17 and NOAA-18 have similar trends for channels 1 and 2, while they had very different orbital drift during the years 2007-2009. NOAA-17 was drifting substantially, while NOAA-18 was stable.
(2) NOAA-17 has a much stronger trend for channel 3 than for channels 1 and 2. If trends were related to changing solar angles as a result of orbital drift, they should be similar for all channels.
We will shortly discuss these arguments in the revised version.

Page 18 line 10. Could the authors elaborate if the inter-calibration method uncertainties mentioned above are applicable to the AVHRR/SEVIRI regression slopes?
We did a similar sensitivity study as in Table 2 for AVHRR vs. SEVIRI. This gave a very similar spread of inter-calibration slopes and standard deviations as for MODIS. Thus, the inter-calibration uncertainty is estimated to be the same. The AVHRR/SEVIRI sensitivity inter-calibration statistics will not be fully included in the paper, but the results will be summarized in Section 6.

Page 19 line 1. Replace defect with defective.
Will be done.