AMT Final Response **13.06.13**

Prior to the point-to-point response we give some information on the repeatability and validity of the liquid nitrogen calibration, because all three reviewers wish some clarification with this regard. The following repeatability analysis is also included in the new Section 4.1.5 “Repeatability and Validity” in the updated version of our paper.

As “repeatability” we understand the capability of a calibration to reproduce the calibration parameters on time scales with negligible instrument drift. The repeatability is addressed by 11 calibrations that were performed within about two hours, using the same radiometer which was deployed at RHUBC-II. The measurements were conducted at the Juelich Research Center on November 10, 2011.

It is assumed that the repeatability is characterized by the stability of the noise diode temperature TN, which is determined with every calibration. Using TN from the above mentioned 11 calibrations allows to map the uncertainty within TN into the range of measured brightness temperatures. For these simulations the detector voltages, the target temperatures, and the other calibration parameters (system noise temperature TR, detector gain g and the detector non-linearity α) are kept constant while only TN is varied. The impact on measured TB is demonstrated for the HATPRO channels at 51.26 GHz in Figure resp1 (Figure 6 a) in the updated paper). For this channel the standard deviation of TN is σ(TN)=2.0 K, the effect on typical TB values under RHUBC-II conditions (TB=38.4 K is the mean temperature measured on August 16, 2009) is σ(TB)=0.4K.

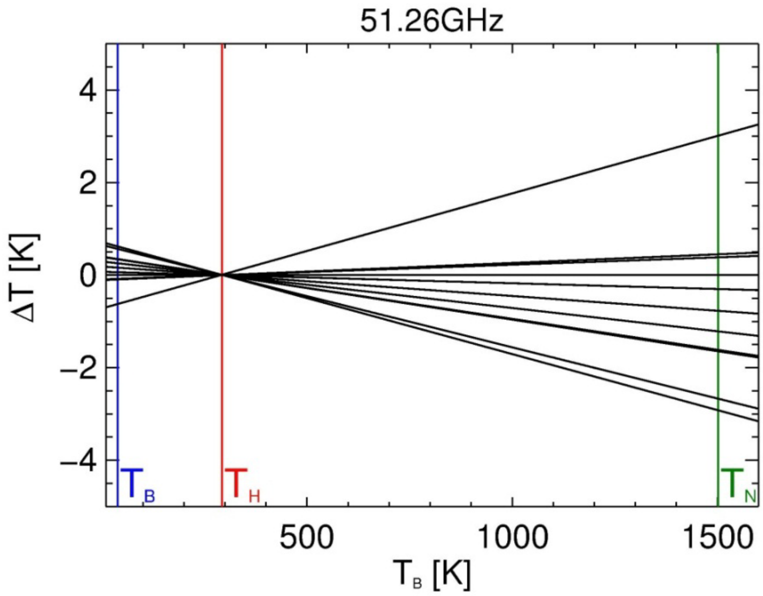

Figure resp1: The impact of the noise diode temperature on TB measurements. Mean brightness temperature TB = 38.4 K, ambient target temperature TH = 293.1K, average noise diode temperature TN = 1501.3 K.

The results are summarized for all channels in Figure resp2 (Figure 6 b) in the updated paper). It shows that σ(TB) ranges between 0.2 K and 0.4 K for the non-opaque channels. For the opaque channels above 54 GHz, the impact of TN on TB is negligibly small, because the calibration is dominated by the ambient target temperature. The Figure is added in the new Section 4.1.5 in the updated version of our paper.

We can conclude that the repeatability of the liquid nitrogen calibration is smaller than the systematic calibration uncertainty we determined to ±0.9 - ±1.6 K in the K-band and ±0.2 - ±1.0 K in the V-band (l. 331ff in the paper). It is assumed that the uncertainty is mainly caused by noise and small variations from calibration to calibration in the level of LN2 in the cold calibration load, which changes the resonance condition for standing waves between the receiver and the LN2 surface (l. 445ff in the paper).

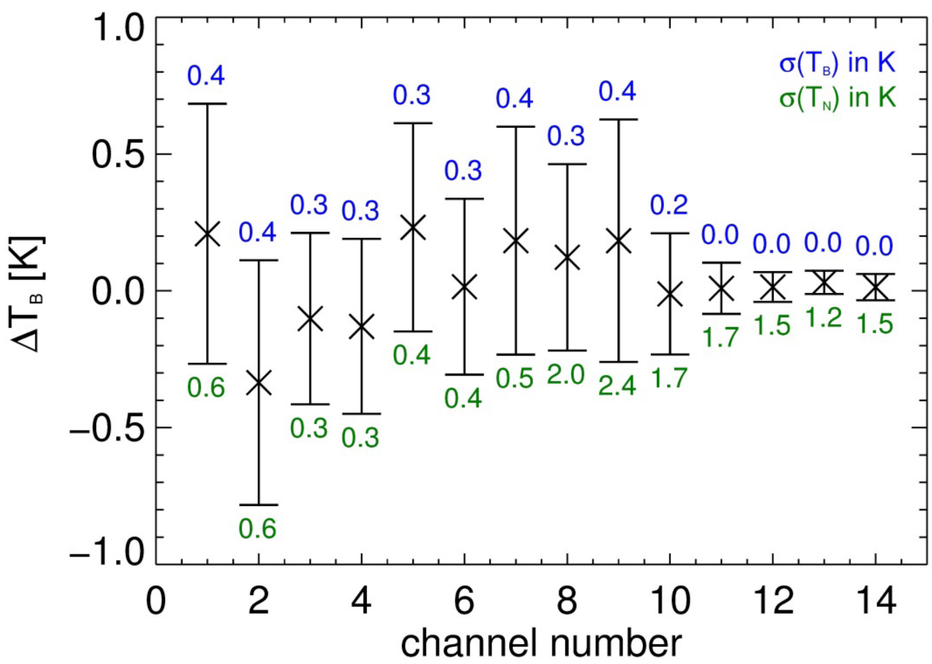

Figure resp2: Repeatability of the liquid nitrogen calibration. The mean difference (X) from the reference calibration (first calibration on Nov. 10) and standard deviation σ(TB) (error bars) for the analyzed 11 liquid nitrogen calibrations is given for all HATPRO-G2 channels. The colored numbers give the standard deviations for TN and TB.

The second point we would like to address here is the validity period of the LN2 calibration. This aspect is rather important for the V-band channels that cannot be calibrated by the tipping curve calibration. Measurements of these channels depend on the noise diode temperature derived from a LN2 calibration.

Our paper includes a comparison of the LN2 calibration and the tipping curve calibration (l. 556ff). Results of the liquid nitrogen calibration, performed on August

11, are compared to tipping curve results from August 16. This comparison is only possible when the instrument's drift in between the analyzed days is negligible.

The drift is recorded by the variation of TN over nearly 2 years. We analyzed 28 liquid nitrogen calibrations performed with HATPRO-G2 in at the Juelich Research Center between June 2010 and May 2012. A trend analysis of TN for these calibrations reveals (Table 1) that the TN increase by -0.012 to +0.007 Kelvin per day in the K-band and by 0.046-0.061 Kelvin per day in the V-band. This means within the 5 days between August 11 and 16, TN changes by less than 0.1 K in the K-.band and 0.3 K in the V-band. The impact on TB can be derived similarly to the repeatability analysis discussed above. When the reference TN (first calibration on Nov 10) is modified by the 5 day drift, TB is affect by less than 0.05 K for all channels (Table resp1, Table 3 in the updated paper) and therefore negligible for our analysis in the paper.

K-band							
Frequency [GHz]	22.24	23.04	23.84	25.44	26.24	27.84	31.40
Mean TN [K]	401.1	399.0	353.4	342.5	370.5	363.3	340.1
TN trend [K/day]	+0.007	+0.006	+0.005	+0.005	+0.005	+0.006	-0.012
TB drift [K] (5 days)	-0.03	-0.02	-0.02	-0.02	-0.02	-0.02	+0.02
TB drift [K] (74 days)	-0.41	-0.23	-0.30	-0.30	-0.29	-0.29	+0.23
V-band							
Frequency [GHz]	51.26	52.28	53.86	54.94	56.66	57.30	58.00
Mean TN [K]	1500.1	1341.8	1227.3	1181.3	1079.6	1083.8	1134.8
TN trend [K/day]	+0.046	+0.059	+0.054	+0.046	+0.049	+0.048	+0.061
TB drift [K] (5 days)	-0.04	-0.05	-0.03	-0.01	-0.01	-0.01	-0.01
TB drift [K] (74 days)	-0.52	-0.69	-0.49	-0.19	-0.12	-0.11	-0.12

Table resp1: Results of 28 liquid nitrogen calibrations performed with HATPRO-G2 at the research center Jülich between June 2012 and May 2012. 5days represent the time period between the two days analyzed in the paper, 74 days is the duration of the measurement campaign RHUBC-II.

Another important aspect that is of interest for all reviewers is that we recently found an error in the post-processing of the brightness temperature data. The uncertainty estimates of the calibration results remain unchanged. However, re-processing the data leads to a better agreement between the LN2 and the tipping curve calibration. For six of the nine calibrated channels calibrated both methods agree within the assumed uncertainties. For the other three channels the unexplained discrepancy is below 0.5 K. Figures 9 and 10 (Figure resp3) in the original paper are updated (Figures 10 and 11 in the new version). The abstract and Section 5.3 are adapted accordingly.

Furthermore, the radiative transfer calculations for non-opaque V-band channels (Figure resp3) now include the effect of the band pass filter's shape. The radiative transfer calculations with a resolution of 10 MHz are convolved with the filter shapes to simulate HATPRO-G2 measurements. The caption is adapted accordingly.

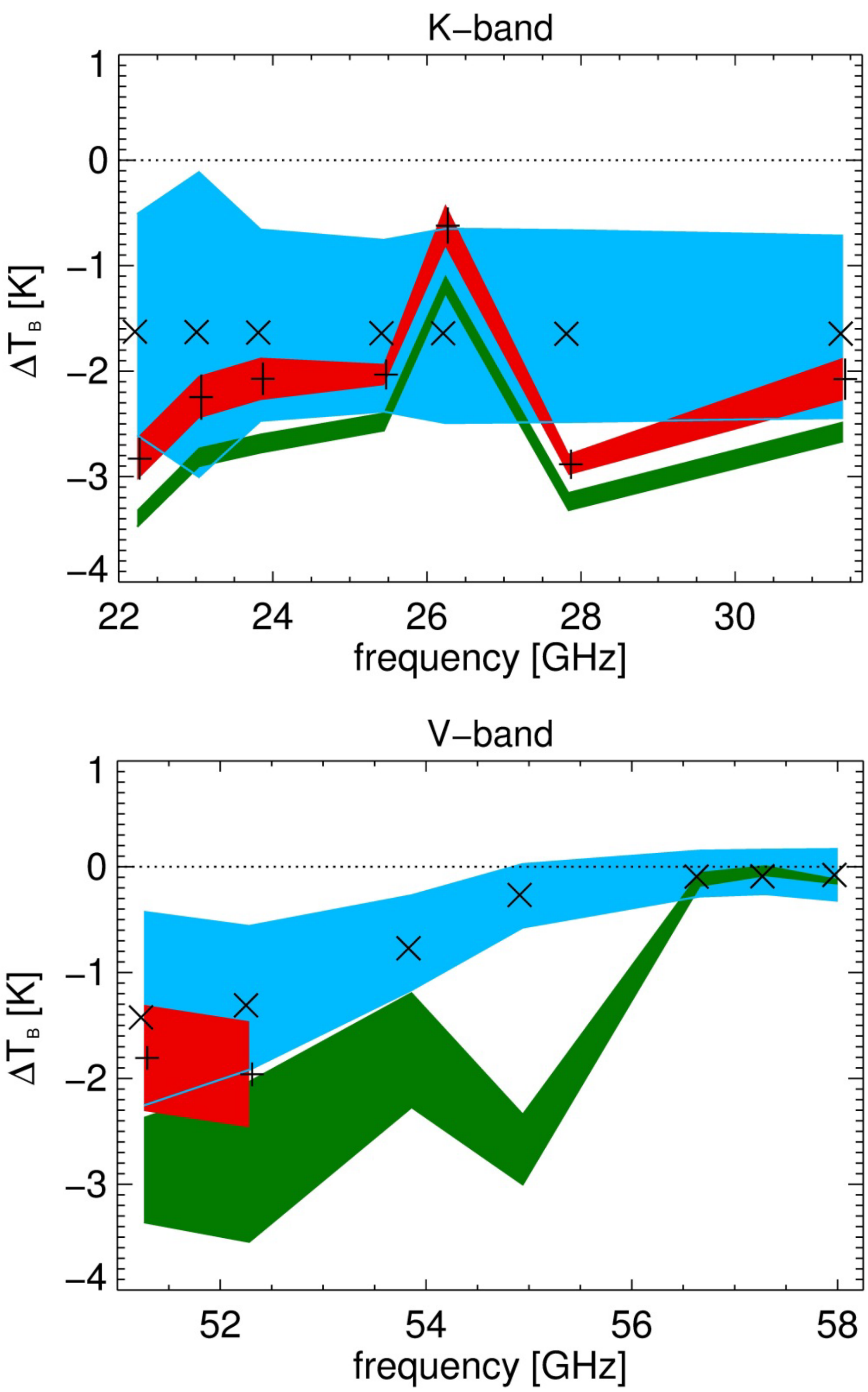

Figure resp3: Comparison and assessed uncertainty ranges of the liquid nitrogen and tipping curve calibration for HATPRO-G2 channels. Red: tipping curve calibration, blue, LN2 calibration, green: radiative transfer calculations. V-band calculations < 56 GHz have been convolved with the real band pass filter shapes measured with a resolution of 10 MHz.

In the point-to-point response, the reviewer's original comments are marked with bullets.

Ref#1

- The manuscript presents a comprehensive error analysis of two commonly used calibration techniques for microwave radiometers, that is the calibration against liquid Nitrogen and the tipping curve calibration. Theoretical consideration are then applied to measurements that have been obtained by the HATPRO-G2 radiometer as part of a campaign in Atacama between August and October 2009. While the error assessment is rather extensive and definitely appropriate for the measurements it is not novel by itself so that it would justify a publication. The interesting part therefore would be to apply this analysis on a number of measurements under different conditions. The authors present one liquid Nitrogen calibration measurement of one day and compare it to a tipping curve calibration of another single day. This is a pity since much more calibrations have been performed during the campaign under a three months' period, covering a wider range of tropospheric conditions that would affect the tipping curve calibrations. This is even described in the outlook paragraph and leaves me to wonder why the authors did not use more, if not all, of the calibration measurements from the campaign. The authors neither considered the Jülich LN2 measurements which could at least contribute some more evidence of repeatability of their calibration study, especially when looking at the effects on the calibration caused by uncertainties in the LN2 refractive index, the resonance effect due to LN2 levels, and instrumental non-linearity.

The authors refer to the given aspect in the beginning of this document. The paper is updated accordingly.

- Under paragraph 2 probably a numbering is missing for the subsection 'Microwave Radiometer'.

This paragraph belongs to 'Measurement Campaign'. As it is the only a subsection it is not numbered. To avoid misunderstandings we remove the section title.

- Here the authors mention observations at elevation angles of 9.6 and even 4.8 degrees. At least 4.8 degrees seems to be pretty low with respect to the antenna opening angle (2 HPBW) of 7 degrees. For good reasons they performed their tipping curve calibration measurements at elevation angles 30 degrees and higher up. An opening angle of 7 degrees would strongly affect low elevation measurements because of the increasing variability of the atmosphere.

The authors agree. However, in ‘Microwave Radiometer’ we only describe the actual measurement mode performed during the campaign.

- Different elevation angles for tipping curve calibration are given in section 3.2 and 4.2.

In Section 3.2 the general setup of the campaign is described while in Section 4.2 only the elevation angles used within our analysis are mentioned.

- I suggest the authors describe the 4-point calibration using the internal noise diode a bit more in detail (or present a reference).

The software manual of the HATPRO-G2 includes a comprehensive description of the 4-point calibration scheme. It has been added as a reference in Section 3.1.

- Unfortunately, they present only one such LN2 calibration for this study and leave it to the reader to interpret this as either ’onefits-all’ or ’no further LN2 calibrations could be performed’. I understand that the calibration measurements during the campaign have been performed with the internal noise diode after the initial LN2 calibration instead of using LN2 calibrations continuously, which would have given a much larger base for this study.

Unfortunately, no further LN2 calibration was performed during the campaign (However, it is common practice to use a single calibration for extended measurements periods). Therefore, we add after l. 235:

“We address this issue by analyzing 11 LN2 calibrations that were performed within about two hours, using the same radiometer which was deployed at RHUBC-II.
The measurements were conducted at the Juelich Research Center on November 10, 2011. In the new subsection “4.1.5 Repeatability and Validity”, we integrate the results from the beginning of this document and give an estimate for the calibration repeatability. It is shown that the repeatability lies within the assumed uncertainty range of the calibration procedure. It is assumed that the repeatability is mainly limited by the impact of standing waves on a single calibration. This effect is already included in our analysis.

- The authors refer to Table 4 frequently, however, I cannot find the numbers in the table as stated in the text. This needs clarification. When giving a number in the text with a reference to a table then I would expect the number to show up in the table. See for instance line 664 ff where neither of the mentioned numbers 0.2 or 0.5 shows up in Table 4.

The numbers given in l. 664ff just give the correction which is applied to the measurements. At this we shall be more accurate: “At 51 GHz and 52 GHz, the correction is 0.1 K and 0.2 K at two air masses, and +0.5K and -0.5K at three air masses, respectively”. In contrast Table 4 (Table 5 in the updated version) gives the uncertainties of the tipping curve results. There is no error assumed for the air mass correction. The reference to Table 4 is wrong and is taken out.

- Furthermore the caption of Table 4 mentions a variable 'air' which does not show up in the table. The same seems to be the case with the variable TN mentioned in the caption but missing in the table in Table 2.

The variable 'air' is removed from the caption.

- In Figure 8 the authors argue that the larger spread of the opacity-air mass correlations for frequencies closer to the water vapor line center is due to the stronger inhomogeneities of atmospheric water vapor closer to the line center. I would agree with that. However, this is not reflected in the V-band cases being far away from any line center. If one would expect a difference there according to the argument above, then the 51.26 GHz channel should have less spread than the 52.28 GHz channel. What could be the reason for the larger spread in the 51.26 GHz channel?

Under the dry RHUBC-II conditions, water vapor measurements away from the line center lose their sensitivity to water vapor. This reduces also the observed variability. For the V-band channels the situation is different: The two channels at 51.26 and 52.28 GHz have much larger signal-to-noise ratios. However, their different weighting functions could also lead to different variabilities.

- In the final comparative assessment the authors compare TB measurements deduced from tipping curve measurements with those TB values deduced from basically the noise diode measurements. They assume a stable behavior of the noise diodes over the time between the LN2 calibration on August 11 and the tipping curve measurements on August 16. Without presenting the Jülich LN2 calibration measurements regarding the stability of this method I would have my doubts on whether this comparison of calibration methods is justified.

The used noise diodes have completed a burn-in phase of 170 hours by the diode manufacturer before they are integrated. Still, the radiometer's manufacturer recommends to re-calibrate the noise diode by a LN2 calibration within the first two years of a radiometer's lifetime in order to eliminate residual drifts of the noise. Between July 2010 and May 2012 28 LN2 calibrations were performed with HATPRO-G2 at the Juelich Research center. The noise diode temperature TN was recalibrated for all channels with every calibration. The variation of TN documents the instrument's drift during this time period. In the beginning of this document and the new Section 4.1.5 in the update version of the paper we show that the drift of the noise diode temperature between August 11 and August has negligible impact on the measurements. In Section 5.3 we add in l. 861ff:

"However, in Section 4.1.5 we show that the comparison is still possible, because the instrument drift within this period is negligible."

- Since the authors mention one principle problem for radiometers, standing waves, I would like to know how they deal with it. Using discontinuous channels in a filter bank always contains the risk of undiscovered standing waves that affect the measurements. I don't understand their comment in this context ' : : : the assessed uncertainty within this study is too small'. Standing waves might also contribute to the observed resonance effect discussed in section 4.1.2 (if not a numbering for this subsection is forgotten here). For me this would be more convincing than that the horn antennas and amplifiers are optimized for the center frequencies in the respective bands and therefore produce some kind of higher amplitude there.

The resonance effect we discuss in Section 4.1.2 is caused by a standing wave. The comment ' … the assessed uncertainty within this study is too small' refers to the amplitudes of the resonance. As mentioned in Section 4.1.2, the standing waves can be treated by repeated LN2 calibrations. Note that by capturing the signal modulation standing waves are not undiscovered any more, but can be corrected. Anyway, with reprocessing of the raw data (confer the beginning of this document) some speculations that were to explain the large discrepancy between the two calibration techniques in the K-band are obsolete. Therefore, we delete l. 895ff:

"Moreover, it is unlikely that the systematic offsets result from a frequency dependent LN2 refractive index n_LN2, because this would contradict laboratory measurements (Section 4.1.2). An effect that can in principle explain the frequency dependence of Delta TB (Fig.10) is the impact of standing waves. However, the assessed uncertainty within this study is still too small."

Ref#2

- The paper "Investigation of Ground-Based Microwave Radiometer Calibration Techniques at 530hPa" is a detailed discussion of errors in different calibration methods for microwave radiometers. The paper is relevant for the microwave remote sensing community and fits well to the scope of AMT. For these reasons I can recommend to publish it after some corrections. There is one major error in Section 3, Equation (1): The temperature in the Planck equation is not the "black body equivalent brightness temperature Tb", but the thermodynamic temperature T. See e.g. Eq. (3) in [Han+Westwater 2000]. The statement "Note that in case of a black body, Tb is equal to the physical temperature of the object" is also incorrect. Even for an ideal black body this is just an approximation, and its validity depends on temperature and frequency. The paper of Han+Westwater gives an in depth discussion of different definitions of the brightness temperature, i.e. Rayleigh-Jeans or thermodynamic Tb. The authors should mention which approximation they will follow in the rest of the paper, and double check that they use it consistently.

- There is one major error in Section 3, Equation (1): The temperature in the Planck equation is not the "black body equivalent brightness temperature Tb", but the thermodynamic temperature T. See e.g. Eq. (3) in [Han+Westwater 2000]. The statement "Note that in case of a black body, Tb is equal to the physical temperature of the object" is also incorrect. Even for an ideal black body this is just an approximation, and its validity depends on temperature and frequency. The paper of Han+Westwater gives an in depth discussion of different definitions of the brightness temperature, i.e. Rayleigh- Jeans or thermodynamic Tb. The authors should mention which approximation they will follow in the rest of the paper, and double check that they use it consistently

We agree that Equation 1 is formally not correct: we replace “TB” by “Tphys”. Furthermore, we add some more information on the definition of brightness temperature and the calibration equations. Line 186ff:

“Note that in case of a black body, TB isequal to the physical temperature of the object. Measurements at reference targets are needed to derive a calibration characteristic (TB=f(Udet)) that can be applied to detected voltages. For HATPRO-G2 it is expressed by the following set of calibration parameters:” is replaced by:

“Calibration procedures based on Equation 1 provide measurements expressed in Planck equivalent brightness temperatures

TB^{PL} = B^{-1}(I),

with the received spectral radiance I. In the following, given temperature values are always Planck equivalent. The relation between detected voltages and measured brightness temperatures TB is expressed by the following set of calibration parameters:”

Furthermore we insert in l. 223: “Instead of using Planck spectral radiances B(T) within these calibrations equations, Equation 1 is expanded in terms of (h*nu/k_b T) and truncated after the first term. For HATPRO-G2's frequency range of 22 GHz< f < 58 GHz and temperature range of T_C< Tphys <T_H the truncation error of this approximation is negligible.”

- The abstract states estimated uncertainties of 0.2-0.4K for the tipping calibration and 0.5-0.9K for the LN2 calibration, and that "Systematic offsets, which may cause the disagreement of both methods within their estimated uncertainties, are discussed". However, one of the main findings of this study seems to be that the two calibration schemes do not agree within their uncertainties. This disagreement should be more clearly in the abstract, i.e. by giving the remaining bias after the obvious corrections of LN2 boiling point and pointing.

As stated in the beginning of this document, the measured brightness temperatures had to be re-processed. Qualitatively, the calibration results remain unchanged. However, re-processing the data leads to a better agreement between the LN2 and the tipping curve calibration. For six of the nine channels which can be calibrated with both methods agree within the assessed uncertainties. For the other three channels the unexplained discrepancy is below 0.5 K. Figure 10 (Figure 11 in the updated version, Figure resp3 here), the abstract and Section 5.3 is adapted accordingly.

- Section 3.1: The LN2 target is described as "...mounted on the radiometer for calibration. The load is filled with egg carton shaped styrofoam to improve the target's black body properties." Since this mounting can have some effect on the observed calibration bias it should be described in more detail, e.g. whether the LN2 target is placed alongside of HATPRO and viewed via a 45deg reflector as described in the HATPRO manual. The load is presumably not filled with "egg carton shaped styrofoam", but with an microwave absorber (e.g. carbon loaded PU foam)

We change the quoted sentence to: '… mounted alongside the radiometer for calibration. The calibration target is observed from above using a reflector which is tilted by 45°. The load is filled with a microwave absorber to guarantee the target's black body properties.' As long as the mirror is dry, it has no significant effect on the observed brightness temperature.

- Section 3.2: According to section 3.2 tipping curve calibrations have been made every 6h with the manufacturer's default angles (30, 33.3, 38.4, 45.6 56.4 and 90deg) towards 50deg N azimuth. In section 4.2, however, the continuous tipping observations at 90, 45, 30, 15 deg elevation were used for this study at 70 and 250deg azimuth were used for calibration. Which statement is correct, and if both scanning schemes were used, how well do they agree with each other?

Section 3.2 describes the measurement mode during the campaign. Performing the tipping curve calibration with the manufacturer's method gave us the possibility to compare the results with the results we derived from the continuous elevation scans. Both methods agree very well. In Section 4.2 we use the continuous scans in order to capture the variability of the tipping curve calibration throughout the day. The results cannot be used to determine the detector non-linearity alpha and the noise diode temperature TN, because no noise is injected during the scans. Nevertheless, the calibrated zenith TB values from each single scan do not depend on these parameters. The information given above is added to Section 4.2 in the updated version of our paper.

- Section 3.3: It is stated that "TN is stable enough to be used as a secondary calibration standard for several months". It might be useful to provide some numbers for the TN stability from repeated calibrations over extended periods (even if this was outside of the RHUBC campaign).

Between July 2010 and May 2012 28 LN2 calibrations were performed with HATPRO-G2 at the Research center Jülich. The noise diode temperature TN was recalibrated for all channels with every calibration. The variation of TN documents the instrument's drift during this time period. In the beginning of this document and in Section 4.1.5 of the updated paper we show that the drift within during RHUBC-II is in below 0.5 K in the K-band and below 1 K in the V-band.

- Is the sentence "Using TR from a previous calibration gives a corrected detector gain g" correct, or should it read TN? It is not made very clear in the text that g is determined from switching the noise diode on and off.

The sentence is correct. The noise diode is observed nearly continuously and it is used to determine the detector gain of V-band channels with a frequency of 10 Hz. For these channels TR is determined with every hot load calibration. In order to make this clearer we modify the text to: ‘Using TR from a previous hot load calibration (Section 3.4) gives a corrected detector gain g.

- Section 4.1 it should be made clear that Eq. 10 refers only to the normal incidence reflections at the LN2 interface of the cold load. The current wording of this paragraph is more general ("when pointing to a calibration target"), but e.g. for the convoluted foam absorber Eq. 10 cannot be used.

Equation 10 can be used for both calibration targets. However, we agree that for the ambient temperature target the reflectivity r is assumed to be zero. In this case we agree and modify the sentence to: ‘… when pointing to the LN2 target’.

- Section 4.1.1 Misplaced semicolon after Eq. 12. At the end of the section is an empty pair of brackets (), to be either removed or replaced by (NIST).

We fill the empty brackets with ‘NIST’.

- Section 4.1.2 Erroneous citations in the sentence "The results are in agreement with several other experiments at frequencies between 0.5GHz (Shitov et al., 2011) and 516GHz (Vinogradov et al., 1967)". The Vinogradov provides measurement data at lambda=2.3mm. Shitov’s experiment operates at 800GHz, but he did not determine n of LN2 and only cites [Hosking et al 1993] with the 0.5-10GHz data.

The authors agree we replace the quoting of (Shitov et al., 2011) by (Hosking et al. ,1993). The 516 GHz originate from an experiment by (Altshuler et al. 1971). We replace the frequency by 130 GHz (2.3 mm) to fit with (Vinogradov et al., 1967).

- The theory behind the standing wave errors is not explained very well, and the meaning of the equation " res(s)(res(s(t))=0)" before Eq. 15 is not clear to me. I’d recommend to read and cite e.g. J. Randa et al, “Errors resulting from the reflectivity of calibration targets,” IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 1, pp. 50–58, Jan. 2005.

We modify the sentence to ‘… a small perturbation res(s). Averaged over time res(s) is zero (overline(rest(t))=0).’ A reference of the suggested article is added to our paper.

- The time varying standing wave error is estimated with 0.6K at 23.04GHz, and less for other channels. However, the authors state that they cannot integrate over one or more oscillation periods of the resonance because the load was not observed long enough. They should clarify whether this estimate is based on further measurements outside of the RHUBC campaign, or whether the LN2 target was observed long enough during the campaign to see the variability. It is a pity that the authors do not show more results from LN2 calibrations over ex- tended time periods (e.g. during the tests at Juelich). This would help to give a better understanding of the errors which occurred in the single short-term LN2 calibration of 11.8.2009.

The long-term observations of the LN2 target were carried out in Jülich. Therefore, we modify the beginning of section ‘Resonance Effect’ (Line 445) to:

‘Continuous observations of the LN2 evaporating from the cold load were carried out at the Research Center Jülich. The observations were conducted with the same radiometer that was previously deployed at RHUBC-II‘.

- Section 4.1.3, apparently, an uncertainty of alpha affects the error on Tb. It should be made clear whether this error depends on the type of calibration, i.e. whether it is the same for LN2, tipping curve or noise diode calibrated data. Section 4.2.2: Probably the first sentence "The tipping curve procedure uses Tb observations to derive opacities at different air mass values" is worded misleadingly. I understood that only the detector voltages are measured at different air masses, and that this information is used to derive opacities and Tb.

The tipping curve procedure iteratively determines the new calibration. That means TB measurements are used as a starting point. In order to prevent misunderstandings. We change the sentence to: ’The tipping curve procedure uses scene observations to derive opacities at different air mass values’.

- Section 5.1 Line 816: "In general, for different channels the overall uncertainty is between 0.6K and 2.7K". These values do not correspond to the ones in Tab. 4, and also Fig. 4 is inconsistent with them and Tab. 4. There is also missing ")" in that line.

The numbers given in l. 816 are wrong. It should be +-0.3 K and +-1.6 K. Generally, we decided to avoid “uncertainty ranges” (between min and max value). Instead we give uncertainties, with +- and a value being half of the original uncertainty range. Table 4 is modified accordingly. In the original version of Table 4 the values for the uncertainty at the “hot” calibration point were by mistake already ‘uncertainties’ instead of “uncertainty ranges”: When the updated total uncertainties are doubled, they differ by up to 0.2K from the original “uncertainty ranges”. The same mistake happened for the tipping curve uncertainties in the column “Tmr”. This leads to uncertainty ranges that differ by up to 0.2K at 51 and 52 GHz. The affected Sections are updated accordingly.

- References: Currently the references appear in random order, and not alphabetically.

The references are ordered by their first appearance in the paper. We agree that an alphabetical order is preferable.

- Fig 1: The V-Band spectra at 530hPa show spectral line features which are not resolved by the broadband filters at the higher frequencies. At normal pressure where the atmosphere becomes opaque this will not be an issue, but at 530hPa it should be necessary to convolve the precise channel response function with the spectrum. How has this been achieved in the present analysis?

As mentioned in the beginning of this document, the band pass filter shapes of the non-opaque V-band channels have been included in Figure resp3 (update of Figure 10 in the paper). These channels are affected by the single absorption lines, which start to separate under low pressure conditions. Especially on the slope of the oxygen absorption complex, the filter shape has an impact of more than 1 K compared to the mid-frequency calculations. In the K-band the effect is negligible. For these channels the mid-frequency results are sufficient.

- Table 2: The caption refers also to the noise diode temperature TN, which is not shown in the table.

The noise diode temperature is removed from the caption.

- Table 3: The columns with yes/yes and no/no are a bit confusing, I assume they indicate whether beam width and air mass correction was applied to the data. It is not clear, however, why this correction affects the number of samples which pass the quality check. Also missing ")" in the caption after "Sec. 4.2.5".

We replace 'yes/yes' and 'no/no' in Table 3 by corr.' and 'uncorr.' . The line 'beam width corr. / air mass corr.' is taken out. The caption extended by: Results from calibrations including a beam-width correction and an exact air mass calculation are given in column 'corr'. The uncorrected results are given in column 'uncorr.'

Ref#3

- This manuscript provides a very detailed, mostly theoretical error analysis for the HATPRO-G2 instrument while deployed in Atacama. The measurements were conducted as part of a campaign conducted between August and October 2009. The results and conclusions of the entire paper hinge upon a single calibration performed at Atacama on 11 August 2009 and a comparison to a series of tipping curves from a single day on 16 August 2009. Why are the tipping curves only shown for a single day? Repeatability under different conditions (presumably there is some variation tau over the 3 month campaign) is a key here as the authors admit, yet they fail to address this issue. Was there a problem with the instrument on all other days? While I understand that there are probably difficulties with returning to Atacama to perform additional calibrations, I do not understand why the authors do not make full use of 28 calibrations which were performed at Julich and which would give some indication of the repeatability of their calibration technique. Again, without some indication of repeatability it is not clear what to make of this study. The error analysis on its own is reasonable and certainly appropriate for a measurement study. If there were more measurements shown this could be an interesting publication. But, given the small number of measurements shown, the error analysis is not sufficiently novel to warrant publication. Almost all of these points have already been addressed in previous studies such as Han and Westwater (which the authors doappropriately repeatedly reference).Without some evidence of repeatability, the results will either be ignored or they mightbe used to form incorrect conclusions.

The authors refer to these aspects in the beginning of this document. The paper is updated accordingly.

- 3.3 “TN is stable enough to be used as a secondary calibration standard for several months.” Is there a plot that shows this? Or a reference?

The used noise diodes have completed a burn-in phase by the diode manufacturer before they are integrated. Still, the radiometer’s manufacturer recommends to re-calibrate the noise diode by a LN2 calibration within the within the first two years of a radiometer’s lifetime in order to eliminate residual drifts of the noise. We add a reference to the ‘Instrument Operation and Software Guide’

- 4.1 There is a lengthy theoretical discussion of possible problems with the cold calibration. Reflections from the surface of the liquid nitrogen may cause large errors, but there are other issue not discussed such as the condensation which can typically occur over a cold load.

HATPRO-G2 uses a dew blower/heater system to prevent condensation on the calibration reflector and the radome during the calibration procedure. This information is added to Section 4.1.

- While the discussion is perhaps not unreasonable, all of the results (with the exception of some discussion on non-linearities) seem to be based upon a single calibration. Without any evidence of repeatability this whole section is deeply flawed.

The authors agree. In the beginning of this document and in the new Section 4.1.5 of the updated paper you find a discussion on the repeatability of the LN2 calibration.

- The authors themselves do present a very nice idea for a study of calibrations as the LN2 evaporates, but instead of actually doing the study they merely state: “Therefore, a more practical solution is to determine the calibration parameters from repeated calibrations while the LN2 evaporates.” This is a very nice idea, and if the results from this were shown in this paper I would not hesitate to recommend publication.

The benefit from repeated calibrations has not been analyzed statistically. However, the method is now implemented in the operational software of HATPRO radiometers. It is expected that the standing waves effect is negligible, when the new calibration procedure is applied. In our paper the standing wave effect is included, because it is inherent in the uncertainty of the LN2 calibration during RHUBC-II.

- Surprisingly, in 4.1.4, the authors state “The variability in alpha is investigated by 28 LN2 calibrations that were performed with HATPRO-G2 at the Research Center Jüllich, Germany between July 2010 and November 2011.” These calibrations need to be studied to address the issues brought up in 4.1.1, 4.1.2., and 4.1.3? There is no need to do those parts of the study at 530hPa.

The authors agree in this point. We used liquid nitrogen calibrations performed at the research center Jülich to analyze the repeatability and temporal validity of this method. You find the results at the beginning of this document and in the new Section 4.1.5 of the updated paper. However, Sections 4.1.1 is not affected, because it describes a systematic effect. The Section 4.1.2 also not is affected, because the effect of repeated calibrations is included in the amplitudes of the detected resonances. Section 4.1.3 already includes the results from the 28 LN2 calibrations in Jülich.

- “For both receiver bands, amplitudes in the band’s center show higher amplitude, because the horn antennas and amplifiers are optimized to the central frequency.” – It seems unlikely that the either the antenna or the amplifier optimized over such a narrow range as to cause this effect.

We agree in this point. Most likely it is the isolator, which causes this effect. The isolator is designed to cover the whole band width of each receiver. However, towards the band edges the performance might slightly decrease. Furthermore, the isolator is the first component in the receiver chain after the horn.

- 4.2 The problems listed in this section are problems which any MWR study has had to address, and almost every subsection begins with a reference to Han and Westwater. It might be justifiable to publish these sections if it were presented in such a way as to be of some use to others. However, the errors calculated in Section 4.2 and the various subsections (4.2.1, 4.2.2, 4.2.3, 4.2.4) are all given as Tb for a very specific case, which makes it impossible to extrapolate the results to anything other than measurements at precisely these frequencies for precisely this atmospheric state. Can’t all of these errors just be expressed as x*T_mr(1-exp(-tau)), where x is some calculated error. This would also help to eliminate the repeated statement “has no effect on K-band and XXK effect on V-band”, which, I think, is really just a statement of the relative optical depths in those 2 bands.

We agree that the question how “universal” the estimated calibration uncertainties is of large interest and could be a topic for a follow-up paper. However, they are not solely a function of the channel opacity. They also depend on the instrument’s noise levels, the channels frequency, the band-pass filters, the antenna beam width, and the calibration type (four point calibration or two point calibration), and targets.

Even though our paper focuses on the analysis of a specific case, it contains the effect of different opacities. The effect of uncertainties at the calibration points (Figures 3 and 4), and the non-linearity parameter (Figure 5) allow to estimate the uncertainty for any measured brightness temperature.

- Figure 6 – Isn't this slope just given by (1-exp(-tau))? I don't think that there is any need to plot this.

We agree that the slope given in Figure 6 is a function of channel opacity.
Nevertheless, we included the plot (Figure 7 in the updated paper), because it visualizes the sensitivity with the RMSE of the regression analysis.