AGU PUBLICATIONS

Radio Science

RESEARCH ARTICLE
10.1002/2015RS005826

Key Points:
- Calibrations of microwave radiometers are accurate within ±0.5 K
- New quality criteria for calibrations were developed and tested
- The absolute temperature of a liquid nitrogen-cooled load was determined

Correspondence to:
N. Küchler,
nkuech@meteo.uni-koeln.de

Citation:
Küchler, N., D. D. Turner, U. Löhnert, and S. Crewell (2016), Calibrating ground-based microwave radiometers: Uncertainty and drifts, *Radio Sci.*, *51*, 311–327, doi:10.1002/2015RS005826.

Received 9 OCT 2015
Accepted 26 MAR 2016
Accepted article online 31 MAR 2016
Published online 21 APR 2016

Calibrating ground-based microwave radiometers: Uncertainty and drifts

N. Küchler[1], D. D. Turner[2], U. Löhnert[1], and S. Crewell[1]

[1]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany, [2]NOAA National Severe Storms Laboratory, Forecast Research and Development Division, Oklahoma, USA

Abstract The quality of microwave radiometer (MWR) calibrations, including both the absolute radiometric accuracy and the spectral consistency, determines the accuracy of geophysical retrievals. The Microwave Radiometer Calibration Experiment (MiRaCalE) was conducted to evaluate the performance of MWR calibration techniques, especially of the so-called Tipping Curve Calibrations (TCC) and Liquid Nitrogen Calibrations (LN2cal), by repeatedly calibrating a fourth-generation Humidity and Temperature Profiler (HATPRO-G4) that measures downwelling radiance between 20 GHz and 60 GHz. MiRaCalE revealed two major points to improve MWR calibrations: (i) the necessary repetition frequency for MWR calibration techniques to correct drifts, which ensures stable long-term measurements; and (ii) the spectral consistency of control measurements of a well known reference is useful to estimate calibration accuracy. Besides, we determined the accuracy of the HATPRO's liquid nitrogen-cooled blackbody's temperature. TCCs and LN2cals were found to agree within 0.5 K when observing the liquid nitrogen-cooled blackbody with a physical temperature of 77 K. This agreement of two different calibration techniques suggests that the brightness temperature of the LN2 cooled blackbody is accurate within at least 0.5 K, which is a significant reduction of the uncertainties that have been assumed to vary between 0.6 K and 1.5 K when calibrating the HATPRO-G4. The error propagation of both techniques was found to behave almost linearly, leading to maximum uncertainties of 0.7 K when observing a scene that is associated with a brightness temperature of 15 K.

1. Introduction

Passive microwave radiometry has been used for temperature profiling, integrated humidity, and liquid measurements [e.g., *Crewell and Löhnert*, 2007; *Löhnert and Crewell*, 2003], and, in synergy with additional remote sensing instruments, for deriving microphysical and macrophysical structures of clouds and precipitation [e.g., *Löhnert et al.*, 2015; *Turner*, 2007]. Data from satellite microwave radiometers (MWRs) have been assimilated into numerical weather prediction (NWP) models [*Bauer et al.*, 2010] and used to validate radiative transfer models [*Clough et al.*, 2005].

Furthermore, ground-based MWRs have produced long-term data sets with high temporal resolution [*Turner et al.*, 2007; *Forkman et al.*, 2012; *Löhnert and Maier*, 2012] and can potentially observe short-term fluctuations of atmospheric constituents, e.g., water vapor [*Wulfmeyer et al.*, 2015; *Steinke et al.*, 2015], as well as provide observations on the vertical structure and temporal development of the boundary layer. Moreover, *Ebell et al.* [2013] found that the synergy of ground-based and satellite measurements improved the quality of atmospheric temperature profiles and humidity information.

All MWR applications rely on the calibration accuracy of MWRs that have been associated with large uncertainties [*Maschwitz et al.*, 2013]. Therefore, further studies of the accuracy of MWR calibration are necessary. The quality and reliability of data collected with a MWR depends on the instrument's thermal stability, noise level, and the calibration accuracy [*Solheim et al.*, 1998]. The calibration is needed to convert measured voltages/counts into brightness temperatures (T_B).

The long-term, unattended operation of MWRs provides important retrieval data products for NWP simulations [*Illingworth et al.*, 2015]. This places increasing demands on improving the calibration accuracy of these radiometers to ensure long-term consistent measurements. Furthermore, improvements in instrument technologies must be verified in order to avoid involvement of new uncertainties sources.

Therefore, the Microwave Radiometer Calibration Experiment (MiRaCalE) was conducted where a state-of-the-art ground-based microwave radiometer was calibrated daily using different calibration techniques. The goal of the campaign was to investigate the accuracy of different calibration techniques and to provide accuracy estimates for MWR T_B measurements.

MWRs are generally calibrated by so-called hot-cold calibrations, ideally, using two reference points that span the full atmospheric measurement range, assuming the detector behaves linearly. Ground-based MWRs use a built-in ambient target as "hot" reference, whereas the "cold" calibration point is realized either with a liquid nitrogen (LN2) cooled blackbody (hereafter LN2cal) [*Ulaby et al.*, 1981] or with a clear-sky zenith measurement which is a so-called Tipping Curve Calibration (hereafter TCC) [*Han and Westwater*, 2000] or a cavity cooled by Peltier elements [*Fernandez et al.*, 2015].

Han and Westwater [2000] and *Maschwitz et al.* [2013] found uncertainties of 0.5 K for TCCs. Uncertainties up to 1.5 K were determined for LN2cals by *Hewison and Gaffard* [2003] and *Maschwitz et al.* [2013]. *Pospichal et al.* [2012] found a resonant effect causing a standing wave at the detector, which contributes strongly to the total uncertainty of the LN2 approach [*Maschwitz*, 2012]. Moreover, *Maschwitz et al.* [2013] found differences up to 1.5 K between LN2cals and TCCs and emphasized the need to further investigate the reproducibility of LN2cals and the absolute accuracies of LN2cals and TCCs. In addition to the resonant effect, *Paine et al.* [2014] found that LN2cals are affected by the entrainment of oxygen into the liquid nitrogen bath inducing an increase of the cold reference's temperature up to 0.4 K during a LN2cal, depending on wind conditions and the length of the calibration activity. *Kaisti et al.* [2014], who combined models and observation, estimated the accuracy of the cryogenic target to be $\pm$1.25 K.

Although the publications presented above provide general information on radiometric calibration accuracy, their findings depend on the designs of the radiometers and calibration targets that were used. For the radiometer (HATPRO-G4) used in this work, estimates of the absolute accuracy of LN2cals and TCCs have not been found. At the moment about 90 HATPROs are continuously measuring in several climatological regions on the globe, such as subpolar or tropical climates [*Shupe et al.*, 2013; *Stevens et al.*, 2015], and therefore provide great potential for climatological studies and data assimilation [*Illingworth et al.*, 2015]. Furthermore, an important aspect that has been neglected generally in the past is the spectral consistency, which means that any systematic calibration error equally affects all radiometer channels.

The absolute radiometric accuracy and the spectral consistency of MWR calibrations were objectives of MiRaCalE and will be presented in the following. Section 2 gives an overview about MiRaCalE's experimental setup and instrumentation. In section 3, we briefly introduce the investigated calibration techniques. Sections 4 and 5 contain the results of MiRaCalE where we discuss calibration uncertainties of MWR calibration techniques and instrument drifts and obtain the temperature accuracy of the cryogenic load, respectively. In section 6 the impact of spectral inconsistencies on the multifrequency retrieval method of integrated water vapor of *Turner et al.* [2007] is assessed. The findings of this work are summarized and discussed in the last section.

2. Microwave Radiometer Calibration Experiment

MiRaCalE took place between 22 September and 14 November 2014 at the National Weather Center in Norman, OK, USA, located at 35°10′54.3″N, 97°26′24.4″W, 345 m mean sea level. The climatological average of cloud fraction over central Oklahoma is less than 10% during the fall season [*Xi et al.*, 2010] which is an optimal precondition for collecting TCCs with high quality. During MiRaCalE, 10 LN2cals were conducted and 2841 TCCs were collected with a fourth-generation ground-based humidity and temperature profiler (HATPRO-G4).

The HATPRO-G4 detects atmospheric radiation in 14 channels simultaneously using direct detection with a temporal resolution of about 1 s [*Rose et al.*, 2005]. The frequency channels are distributed along the pressure broadened water vapor absorption line at 22.235 GHz (K band) and along the complex of superimposed rotational transitions of oxygen at 60 GHz (V band). Furthermore, a window channel at 31.40 GHz (included in K band) is used to increase the sensitivity to cloud liquid water.

The half-power beam widths of the antenna systems are 3.5° and 1.8° for the K and V bands, respectively. The half-power band widths are 230 MHz between 22.24 GHz and 54.94 GHz, 600 MHz at 56.66 GHz, 1000 MHz at 57.30 GHz, and 2000 MHz at 58 GHz giving a lower noise level for the three most opaque channels where the

spectral gradient is low [*Maschwitz et al.*, 2013]. The noise level of HATPRO-G4, given by the manufacturer, is 0.1 K and 0.2 K root-mean-square error at 1 s integration time in the K and V bands, respectively.

The K band receiver is made of Gallium Arsenide technology, whereas in the V band Indium Phosphide is used. Therefore, the detector response in the V band is less linear than in the K band and the gain (detector sensitivity in V/(W m^{-2} Hz^{-1} sr^{-1}) fluctuates on time scales of seconds in the V band [*Maschwitz*, 2012]. In the K band the gain is assumed to be stable over several minutes. The instrument uses an internal noise sources (one for each band) for several calibration procedures. A white noise signal (which has a constant equivalent noise temperature) is produced by noise diodes that are stable over several months [*Radiometer Physics GmbH*, 2013].

Moreover, HATPRO-G4 is equipped with a pyramidal-shaped, internal target that is placed in a Styrofoam box. The box is transparent to radiation for frequencies less than 100 GHz, is hermetically insulated, and is located directly beneath the collecting mirror [*Rose et al.*, 2005]. It is used for most of the radiometer's calibration procedures. The air inside this box is continuously stirred with fans to reduce thermal gradients within the absorber material. However, thermal gradients across the entire blackbody are not monitored and therefore remain an unknown uncertainty source. *McGrath and Hewison* [2001] found gradients much less than 0.2 K for an ambient target of similar design having no venting system.

The temperature of the absorber is determined by measuring the temperature of the inside air with a sensor having an accuracy of ± 0.1 K. A second temperature sensor is placed a few millimeters beside the first to verify the reference measurement. If both sensors deviate more than 0.1 K, then the measurement will be flagged and the calibration will be discarded. The maximum difference between the temperature sensors inside the internal blackbody during MiRaCalE was 0.1 K, which is hereafter assumed to represent the accuracy of the internal target. A detailed information on the radiometer design is given in *Radiometer Physics GmbH* [2013].

3. Calibration Techniques

The spectral radiance (SR) of the observed scene B_{sc} in W m^{-2} Hz^{-1} sr^{-1} observed at the frequency ν within the antenna beam of the HATPRO-G4 is related to the output voltage U_{sc} of the radiometer via

$$U_{sc} = g(B_{sc} + B_R)^{\alpha} \tag{1}$$

where g is the gain, B_R is the receiver noise SR, and α is the nonlinearity parameter [*Crewell et al.*, 2001].

For convenience when comparing measurements spectrally, it is common to express observed atmospheric SR in terms of Planck equivalent brightness temperatures T_B solving Planck's law

$$B(\nu) = \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{h\nu}{k_B T_B}\right) - 1} \tag{2}$$

for T_B, where k_B is the Boltzmann constant, h the Planck constant, and c the speed of light. If the observed scene is a blackbody, its T_B will be equal to its physical temperature. Thus, the SRs of calibrations loads can be directly obtained from equation (2).

The unknown parameters in equation (1), i.e., g, B_R, and α can be determined using different calibration techniques. A LN2cal provides information on all unknowns using a cryogenic load emitting the SR B_C, the internal ambient target (B_H), and the built-in noise diodes, added to B_C and B_H, respectively, to produce a third and fourth calibration reference. Furthermore, the noise diodes' equivalent SRs B_N are derived during the LN2cal procedure. It also is possible to obtain B_N from TCCs in the K band; however, this does not work in the V band channels, as the radiative transfer is no longer linear with air mass due to high gaseous absorption.

The HATPRO-G4 uses TCCs to update the system noise B_R and the gain. Thereby, it is assumed that the system's nonlinearity is constant. *Maschwitz et al.* [2013] showed that this is assumption valid for several months. TCCs are realized using the internal target as hot reference and the clear-sky zenith SR as cold calibration point. The latter is determined in an iterative procedure assuming a horizontally stratified atmosphere [*Han and Westwater*, 2000].

The built-in noise diodes are used as calibration standard once their equivalent SRs B_N were determined during a LN2cal. This, so-called noise diode calibration (NDC), updates the gain in the V band by periodically

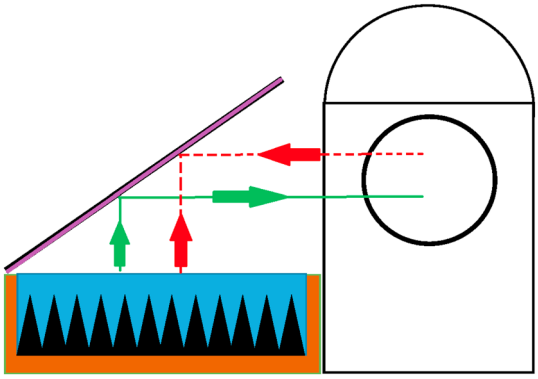

Figure 1. Pyramidal blackbody (black) immersed in liquid nitrogen (blue) in a nonabsorbing high-density styrofoam container. The blackbody's signal (green) is redirected by a reflector (purple) toward the parabolic mirror (black circle). Red dashed line: the thermal emission of the radiometer is reflected at the liquid's surface which is superimposed with the target's signal.

adding noise (with a frequency of 10 Hz) to the observed scene's SR. In the K band, the NDC can be used to update the system noise B_R and the gain while observing the internal target with and without additional noise signal. Although the NDC updates B_R and g as the TCC in the K band, TCC are preferable as their reference points span the full measurement range.

The fourth calibration procedure, used by the HATPRO-G4, is the so-called hot load calibration. While observing the internal hot load, the gain in the K band and the system noise in the V band are updated, respectively. A more detailed description of the calibration techniques can be found in the Appendix A.

4. Performance of Calibration Techniques

The goal of a calibration is to provide accurate brightness temperatures over the full measurement range. Hence, it is crucial to determine the uncertainty of radiometric calibrations and how frequent they must be repeated (hereafter "repetition frequency") to account for drifts in the instrument and therefore to ensure accurate MWR measurements. In section 4.1 we evaluate the spectral consistency and uncertainty estimates of LN2cals. The quality of hot load calibrations and NDCs are discussed in section 4.2. Spectral consistency and uncertainty estimates of TCCs and instrument drift rates are investigated in section 4.3 including the introduction of an improved quality criterion for TCCs.

4.1. Quality and Uncertainty Estimates of Liquid Nitrogen Calibrations

LN2cals of the HATPRO-G4 are mainly affected by three error sources: (i) the uncertainty of the refractive index of liquid nitrogen; (ii) nonzero reflectance of the radiometer's horn antenna; and (iii) the entrainment of oxygen into the bath [*Maschwitz et al.*, 2013; *Pospichal et al.*, 2012; *Paine et al.*, 2014]. *Kaisti et al.* [2014] showed that the increase of the LN2 boiling point due to hydrostatic pressure is smaller than 0.05 K for filling heights of the cryogenic load of a few centimeters (maximum 7 cm above the target) as used in our measurement setup.

Point (i) leads to an uncertainty due to the unknown contribution of the receiver's noise signal reflected back into the field of view (Figure 1) and point (ii) causes multiple reflections forming a standing wave signature in the detector's output voltages. *Maschwitz et al.* [2013] suggested that these two effects generate uncertainties of the cryogenic load of up to 1.5 K and 0.9 K in the K and V bands, respectively. However, they assessed neither the averaging method of the manufacturer that was included at a later point nor the entrainment of oxygen found by *Paine et al.* [2014]. As the latter yields an unknown absolute temperature of the cryogenic bath, a third reference [e.g., *Kaisti et al.*, 2014] is needed to determine the absolute accuracy, which has not yet been investigated for measurement setups in which liquid nitrogen has a reflecting interface. Therefore, the total uncertainty observing the cold blackbody is assumed to be between the uncertainty with and without resonant effect, which is 0.6–1.5 K and 0.6–0.9 K in the K and V bands, respectively [*Maschwitz et al.*, 2013].

4.1.1. Uncertainty Estimates Observing Calibration References

To verify the findings of *Maschwitz et al.* [2013], but including the effects of entrainment of oxygen and the adjusted calibration algorithm provided in the HATPRO manufacturer's software, the uncertainty of LN2cals

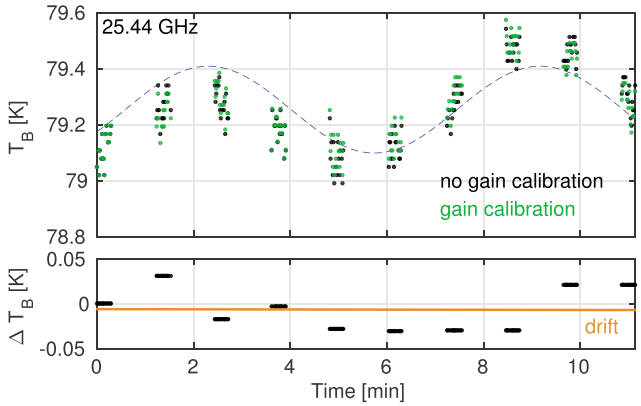

Figure 2. (top) Experimental time series of the RVSLN2 converted into T_Bs without gain calibrations (black) and with gain calibrations every minute (green). (bottom) Difference between T_Bs derived with and without gain calibrations and its 10 min trend (orange). The time resolution is 1 s, whereas the time series is not continuous due to the measurement scan pattern (15 s gaps). The blue line indicates the sinusoidal oscillation applying a simple sine fit. Note the sine fit does not include a linear increase that is expected due to the entrainment if oxygen into the cryogenic bath.

was assessed by performing control measurements on the calibration targets after the calibration. When observing the LN2 cooled target, it is necessary to integrate over a full period of the standing wave in order retrieve its mean value (see Figure 2).

Note that the signals that were recorded after the LN2cals and were used for this analysis were not continuous, which was due to the scan pattern that was performed in this case (shown in Figure 2). First, the LN2 cooled blackbody was observed, then a noise signal was added to the latter, followed by the observation of the internal blackbody both with and without additional noise signal (gaps in Figure 2). In total, every cycle consists of four 15 s observations which were repeated 10 times. The illustrated signal (Figure 2) was recorded on 15 October 2014 and will be later referred to as the reference voltage signal of the observation of the LN2 cooled blackbody (RVSLN2). The temperature of the LN2 was 79.04 K determined after *Maschwitz et al.* [2013] using the mean surface pressure at the experimental site, the receiver's temperature, and the refractive index of LN2 determined by *Benson et al.* [1983] of 1.2. As clearly seen by the sinusoidal shape in Figure 2, the time series covers more than one period of the standing wave. The first 15 s of the internal blackbody observation were used to determine the hot target's brightness temperature after each LN2cal.

From MiRaCalE, control measurements are available for 6 out of 10 LN2cals containing the same scan pattern as shown in Figure 2. To have a quantitative measure for the uncertainty of LN2cals, the variability of all six control measurements (in each channel) are considered in terms of the standard deviation (SD 6; knowing that this is statistically not representative). Thus, a perfect calibration has a SD 6 equal to zero. Figure 3 illustrates the differences between T_Bs retrieved with the manufacturer's software and the brightness temperature of the cold calibration target that was expected to be 79.04 K. The measurements were made immediately after the LN2cals. The largest values for SD 6 found in the K and V bands are 0.22 K and 0.26 K at 22.24 GHz and 52.28 GHz, respectively (Table 1).

Errors in T_B that originate from the cold calibration target decrease toward the hot calibration point and increase toward T_B smaller than T_C. Therefore, LN2cals at V band frequencies can be considered as sufficiently precise with respect to instrument's noise level in the V band of 0.2 K (assuming an unbiased hot target), since the typical atmospheric scene brightness temperatures observed are at or above 80 K in the V band. However, K band channels can measure T_Bs lower than 15 K in very transparent conditions (i.e., low opacities) and are strongly affected by biased cold calibration points. Offsets in T_C up to 0.5 K were found for an individual calibration (see Figure 3), which translate into uncertainties of 0.7 K at 15 K scene T_B assuming linear error propagation and an unbiased hot target observation [*Maschwitz et al.*, 2013].

In general, the hot reference is also biased as was found by observing the hot target for 15 s after the LN2cals. The maximum values of SD 6 of the hot target are 0.12 K and 0.28 K in the K and V bands, respectively, again at 22.24 GHz and 52.28 GHz. The values agree within the instrument's noise level including the accuracy of the internal target of 0.1 K. Here the impact on atmospheric observations is opposite from above, since

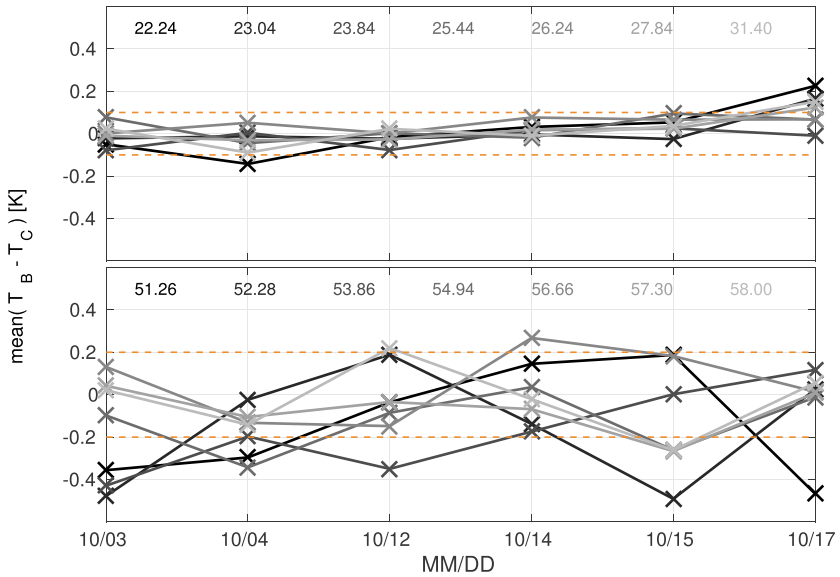

Figure 3. Difference between expected value and mean observed cold target temperatures (crosses) directly after LN2cals. The manufacturer's software accounts for changes of g in the K and V bands automatically. Since the observation was recorded directly after the calibration, B_R was assumed to stay constant over the observation time. Each point is a 10 min integration of a view of the cold target in order average the standing wave to cover a full period in all channels, hence, to minimize its influence. The orange dashed lines mark the radiometer's noise level in the K and V bands (i.e., 0.1 K and 0.2 K, respectively) provided by the manufacturer.

the atmospheric observations in the K band, rarely reaching 100 K, are hardly affected by biased hot load temperatures. The V band channels are affected by about 0.2 K depending on the frequency and opacity of the observed scene (again assuming linear error propagation). A more detailed assessment of the error characteristics over the full temperature range will be discussed later in this section.

The SD 6(T_C) and SD 6(T_H) values at 22.24 GHz are significantly larger than those from the other K and V band channels. This might be due to the geometry of the HATPRO-G4's horn antenna that is adjusted to the center

Table 1. Standard Deviations (SD 6) of Cold and Hot Target Mean Brightness Temperatures of Six Observations Recorded Directly After Six Different LN2cals[a]

ν (GHz)	SD 6(T_C) (K)	SD 6(T_H) (K)
22.24	0.22	0.12
23.04	0.04	0.07
23.84	0.09	0.04
25.44	0.10	0.06
26.24	0.07	0.03
27.84	0.08	0.06
31.40	0.11	0.08
51.26	0.22	0.27
52.28	0.26	0.28
53.86	0.16	0.21
54.94	0.25	0.15
56.66	0.10	0.17
57.30	0.18	0.11
58.00	0.10	0.17

[a]Every cold target temperature, contributing to SD 6, is a mean value of about 10 min in order to cover a full period of the standing wave. Every hot target temperature is a mean of 15 s observation, equal to the integration time that was used during the LN2cals.

frequency of the entire K band to perform well over the full band width of channels of about 10 GHz. Thus, its performance decreases toward the outermost channels, which is also indicated by a slight increase of SD 6(T_C) and SD 6(T_H) at 31.40 GHz (Figure 3). In the V band, the four lowest frequency channels exhibit similar values, so that there is no clear evidence for a geometric effect of the horn antenna.

4.1.2. Evaluating Spectral Consistency Observing Calibration References

We found that LN2cals produce consistent results for individual channels; however, differences among the channels indicate biased calibrations, which are not considered by the manufacturer's software. Figure 3 suggests opportunities to define quantitative thresholds for the quality of LN2cals. First, one could define a threshold for the maximum difference that is accepted between the measured and the perfect reference temperature, which corresponds to the noise level of the radiometer (Figure 3, orange lines) and must capture all values of the channels of the considered receiver unit (i.e., K band or V band). However, this one-measurement-series criteria would reject all calibrations in both bands considering only the cold target observations (see Figure 3).

A second approach is to define a "spectral consistency" value in terms of a standard deviation among the channels located in a receiving unit, i.e. SD_K and SD_V. If this spectral consistency value has to be smaller than the instrument's noise level, five calibrations in the V band and two calibrations in the K band will be accepted in Figure 3.

The rejection of the K band channels is mainly due to the cold load observation, where the resonant effect causes frequency dependent offsets in addition to the noise. This is due to the mean T_Bs (one for each channel) of 10 min that contain more than one full cycle of the observed wave. Since the period of the wave decreases with frequency, the number of averaged wave numbers during the 10 min integration period also changes.

Moreover, the standard deviation itself does not reflect any systematic offsets from the expected temperatures; hence, the mean values should be taken into account as well. How spectral inconsistent calibrations influence the accuracy of geophysical retrieval will be discussed in section 6.

4.1.3. Scene Dependent Uncertainty Estimates

To complete the analysis of LN2cals, we determined the uncertainty for the full measurement range of the HATPRO-G4:

1. Calibration parameters were obtained from equation (A1) to equation (A4) using the SRs $B_C = B(T_C)$ and $B_H = B(T_H)$, which are computed from the physical temperatures of the calibration targets. This parameter set is used as reference.
2. Calibration parameters were obtained from equation (A1) to equation (A4) using the SRs $B'_C = B(T_C + \Delta T_C)$ and $B'_H = B(T_H - \Delta T_H)$. The targets' SRs were perturbed according to their assumed uncertainties ΔT_C, given by *Maschwitz et al.* [2013], and $\Delta T_H = 0.1$ K. The perturbations of opposite sign were used to derive the maximum possible uncertainty range.
3. Both calibration parameter sets were used to calculated SRs from voltages that correspond to scene brightness temperatures between 3 and 320 K.
4. The resulting SR differences ΔB at scene temperatures between 3 and 320 K are then converted into brightness temperatures to provide uncertainty estimates ΔT_B in the temperature regime.

Figure 4 shows absolute values of ΔT_Bs over the full regime. The error propagation is almost linear except at very small T_Bs where the nonlinearities of equation (2) and equation (1) become significant. In the K band, errors up to 2 K are evident at low T_Bs. All V band channels have an uncertainty which is smaller than 0.6 K at typical scene T_Bs ($\geq$ 80 K).

4.2. Quality and Uncertainty Estimates of Hot Load and Noise Diode Calibrations

As noise diode and hot load calibrations are mostly instrument specific, we will discuss in the following those aspects which either are necessary to support our analysis or contain general information on MWR calibration.

On short time scales (5–10 min), the stability of the radiometer in the K band is preserved by frequent gain calibrations using the internal target. The stability of the gain in the K band was investigated using the RVSLN2 (i.e., the voltage time series of the cryogenic load). Figure 2 shows the RVSLN2 at 26.24 GHz converted into T_Bs (using equations (1) and (2)) with and without frequent gain calibrations. The difference is smaller than 0.05 K for the full time series (more than 10 min), and the drift is therefore negligible. Disturbing the internal target's temperature by 0.1 K has no effect on the retrieved T_Bs in Figure 2.

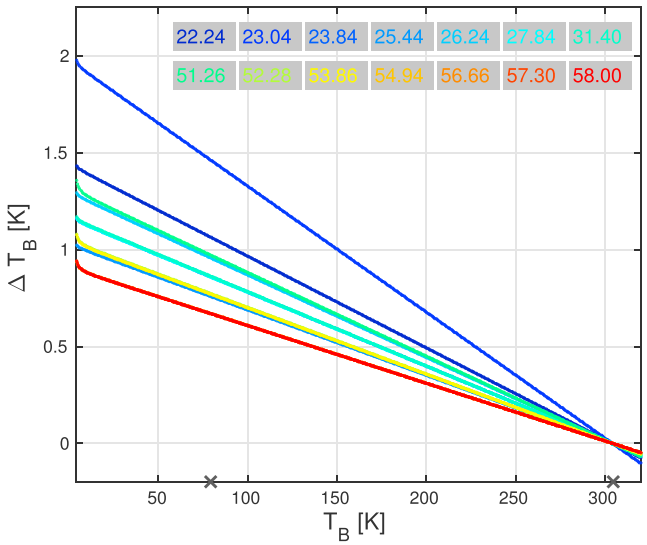

Figure 4. Absolute value of the uncertainty of LN2cals depending on T_B of the observed scene. The uncertainty that results using the values of *Maschwitz et al.* [2013] are colored. The reference targets' physical temperatures are marked on the *x* axis.

Noise diode calibrations can serve as secondary standard to update the system noise B_R and the gain in the K- and V bands, respectively. Therefore, the SR of the noise diode (B_N) has to be stable between two LN2cals. To describe the latter with a statistically representative value, e.g., a time series of B_N over at least 6 months with about two LN2cals per week, would be desirable; however, this data set does not yet exist.

Uncertainty estimates for NDCs in the K band were obtained by performing NDCs consecutively every 8 s. In between the calibrations 2 s zenith measurements were recorded. The zenith T_Bs obtained with NDCs were normalized by subtracting zenith T_Bs, which were obtained from gain calibrations, to account for changing scene temperatures. In total 99 NDCs were conducted. The standard deviation of the differences (for any number of data points n: $2 \leq n \leq 99$) is smaller than 0.2 K in all K band channels. An uncertainty of 0.1 K of the ambient target affects all K band channels by less than 0.5 K at clear-sky zenith scenes between 15 K and 30 K (depending on frequency channel and integrated water vapor (IWV)). If the noise diodes are stable over several months, the uncertainty of NDCs will be about 0.5 K in the K band. For the V band channels an uncertainty assessment for NDCs could not be performed since the voltages from high frequent noise switching were not available.

4.3. Quality and Uncertainty Estimates of Tipping Curve Calibrations

The quality of TCCs is typically determined by applying statistical thresholds to the linear regression of opacity-air mass pairs. For a perfect TCC, the linear interpolation (see Figure 5) has no residual variance ($\chi^2 = 0$), intercepts the origin of the coordinate system ($b = 0$), and has a linear correlation coefficient which is equal to one ($\mathrm{corr}(a, \tau) = 1$) (we will refer to these three criteria hereafter as the Chi-Tau(0)-Correlation (CTC) criterion). An additional restriction is that all K band channels need to pass the quality criteria simultaneously. The thresholds that were used in this work were determined by a statistical analysis of 2841 TCC that were collected during MiRaCalE. The thresholds are $\chi^2 < 2 \cdot 10^{-4}$, $|\tau(0)| < 10^{-3}$ and $\mathrm{corr}(a, \tau) > 0.9991$ and were selected so that the number of accepted TCCs is statistically representative. We admit that these thresholds are subjective.

4.3.1. Drift of the System Noise

To investigate the necessary repetition frequency to account for drifts of B_R, only TCCs were used to refresh the system noise. Figure 6 illustrates the evolution of the receiver noise brightness temperature of TCCs that were filtered with the CTC criterion. Shown are only K band channels of the HATPRO-G4 over 26 days, since the atmosphere was too opaque during MiRaCalE to collect successful TCCs with V band channels. The drifts in T_R vary between −0.02 K/d and −0.03 K/d within the K band channels, which would lead to a drift of ≈2 K in 100 days if the drifts stayed constant over this time period.

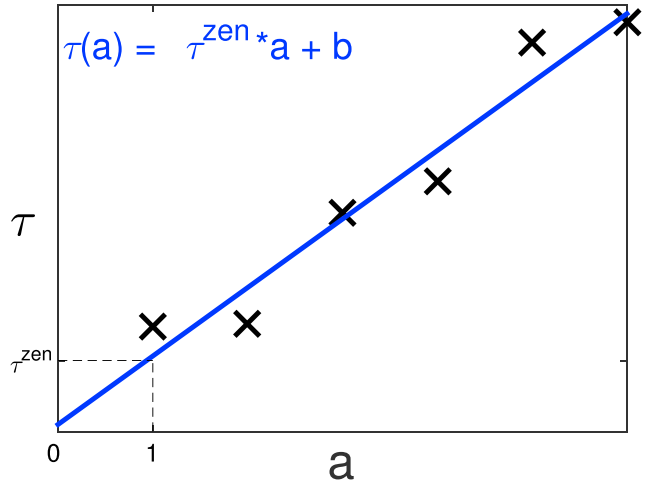

Figure 5. The opacities τ derived at each angle scale linearly with the corresponding air masses a. b is the offset that occurs at the $a = 0$ when the calibration is not perfect.

Note that the evolution of T_R (B_R) illustrated in Figure 6 does not necessarily represent the actual change of the receiver's noise figure, as the calibration parameters B_R and g are adjusted assuming α to be constant and B_R and g might compensate each other, i.e., are correlated. Nevertheless, although the apparent drifts of B_R do not represent the physical truth, they represent the mathematical solution of TCCs providing ideally accurate measurements.

To illustrate the effect of drifts of B_R (T_R) on brightness temperatures, RVSLN2 was converted into SRs (T_Bs) using the calibration parameter sets that correspond to B_Rs (T_Rs) shown in Figure 6. Remember that the RVSLN2 was recorded on 15 October (Figure 7, red line); thus, all g-B_R pairs of the TCCs are different from instrument's state at this day because of drifts in the calibration between the time when the TCC was performed and 15 October. The gain drifts were corrected by using an observation of the internal blackbody directly prior to the RVSLN2. The resulting drifts over time due to changing B_R (T_R) are shown in Figure 8. Again, linear

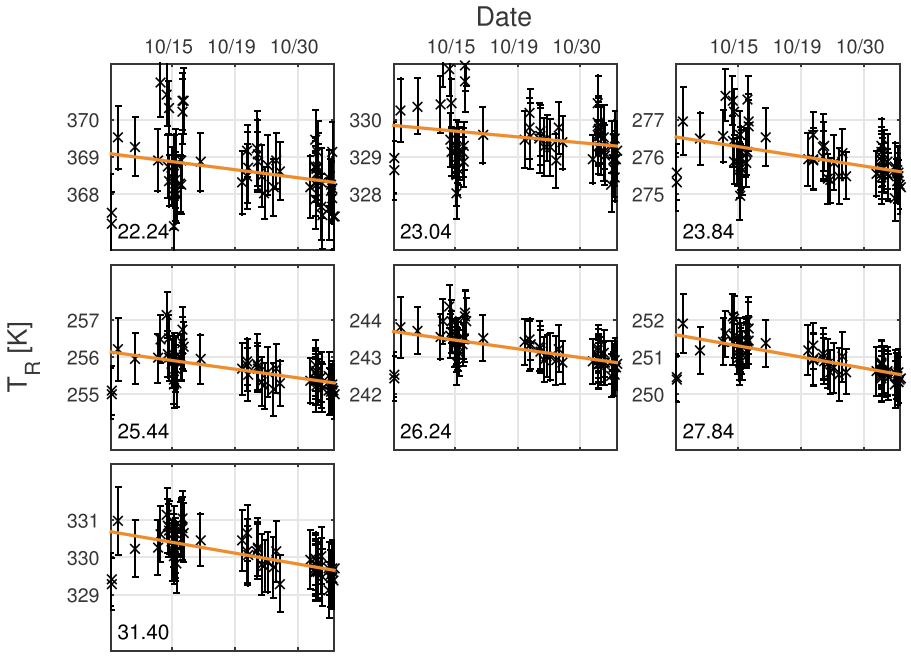

Figure 6. A 26 day time series of T_R of successive TCCs filtered by the CTC criterion. The error bars of T_R were obtained assuming Gaussian error propagation and were derived from the cold and hot reference and the measurement uncertainties.

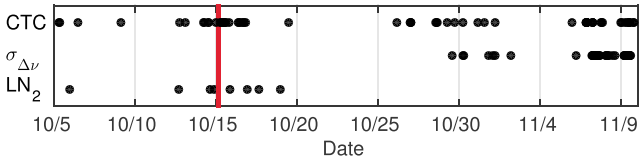

Figure 7. Time line of TCCs that passed (top) the CTC criterion or (middle) the σ_Δ criterion. (bottom) Also illustrated are LN2cals that were conducted during MiRaCalE. The date of the RVSLN2 is marked with the red line.

regressions were determined for all channels (Table 2), which provide a mean drift rate at 79.04 K (T_C) of 9.4 ± 2.3 mK/d that indicates that updating B_R (T_R) once a day, i.e., correcting for a drift of ≈ 0.01 K, ensures consistent measurements.

4.3.2. Spectral Consistency as Quality Check

The experiment offered a second opportunity to investigate the goodness of the TCCs by using a filter criterion. Since TCCs use clear-sky zenith SR as cold calibration point ($T_C \approx 15-30$ K for the K band channels depending on water vapor amount), the RVSLN2 was used as third reference to validate the quality of TCCs. T_Bs of the RVSLN2 were derived from the calibration parameters provided by the TCC ($T_{B,tip}$) and then were normalized with T_Bs calculated with a reference calibration parameter set from a spectrally consistent LN2cal (T_{B,LN_2}). This was necessary to account for spectrally deviating mean T_Bs of the RVSLN2 due to the channel-differing periods of the observed wave in RVSLN2 [*Pospichal et al.*, 2012]. The difference

$$\Delta T_{B,\mathrm{cal}}(\nu) = T_{B,\mathrm{tip}}(\nu) - T_{B,LN_2}(\nu) \tag{3}$$

will be the same in each K band channel if the TCC is spectrally consistent. This criterion is defined as the spectral standard deviation $\sigma_{\Delta\nu}$ of the channel differences $\Delta T_{B,\mathrm{cal}}(\nu_i)$ ($i = 1, \ldots, 7$ with seven channels) monitoring the cold target (henceforth $\sigma_{\Delta\nu}$ criterion, where $\Delta\nu$ represents the frequency band from 22.24 GHz to 31.04 GHz):

$$\sigma_{\Delta\nu} = \sqrt{\frac{1}{6}\sum_{i=1}^{7}\left(\Delta T_{B,\mathrm{cal}}(\nu_i) - \overline{\Delta T_{B,\mathrm{cal}}}\right)^2}. \tag{4}$$

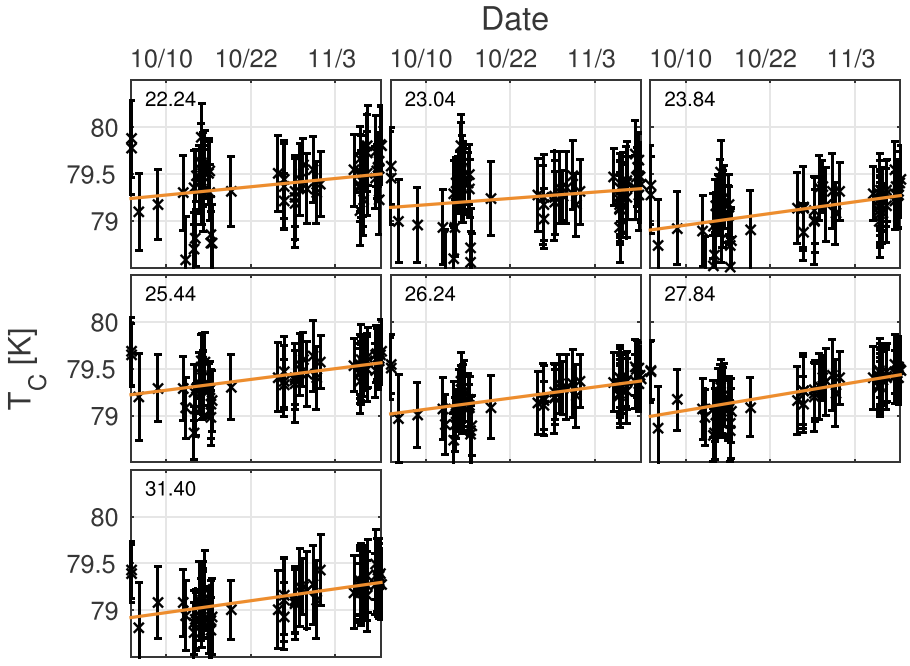

Figure 8. Cold target temperatures calculated from the RVSLN2 using TCCs that passed the CTC criterion. The error bars correspond to the uncertainty of the cold calibration point interpolated to 79.04 K.

Table 2. Drifts of T_R at 79.04 K and Cold Target's Temperatures Interpolated to 15 October Derived Using Data Selected by the CTC Criterion and the $\sigma_{\Delta\nu}$ Criterion

	CTC Criterion		$\sigma_{\Delta\nu}$ Criterion	
ν (GHz)	Drift (mK/d)	T_C (K)	Drift (mK/d)	T_C (K)
22.24	7.4	79.33	9.2	79.23
23.04	5.6	79.29	11.6	79.17
23.84	10.4	79.20	11.3	79.08
25.44	9.6	79.45	8.0	79.36
26.24	9.9	79.29	9.8	79.18
27.84	12.5	79.36	6.7	79.28
31.40	10.6	79.27	5.4	79.21

This approach does not consider the absolute temperature of the third reference which ideally results from applying calibration parameter sets to the recorded voltage signal but instead tests the spectral consistency. Hence, if the calibration deviates in all channels by the same amount from the third reference's T_B, the $\sigma_{\Delta\nu}$ criterion will not identify the calibration as invalid.

The critical value of $\sigma_{\Delta\nu}$ to identify a spectrally consistent (and let us assume an accurate) calibration was set to 0.1 K that corresponds to the radiometer's noise level in the K band. This makes the $\sigma_{\Delta\nu}$ criterion an instrument specific threshold that can be determined for every instrument individually and is not chosen by experience as the CTC criterion introduced before. Furthermore, the $\sigma_{\Delta\nu}$ criterion is based entirely on the spectral consistency and does not necessarily depend on statistical parameters of the linear regression.

Although $\sigma_{\Delta\nu}$ is not an absolute criterion, it is used as the uncertainty estimate of the cold calibration point, which is a reasonable assumption as will be shown in the following. Figure 9 illustrates the evolution of T_C that was derived from TCCs that passed the $\sigma_{\Delta\nu}$ criterion (similar to Figure 8). The fitted curves and the error bars (i.e., $\sigma_{\Delta\nu}$ at the cold calibration point) overlap for more than 66% of the accepted TCC in six K band channels. At 22.04 GHz overlaps occur in 58% of the cases. This agreement between the fits and the accepted TCC within the error bars underlines that it is reasonable to consider $\sigma_{\Delta\nu}$ as the uncertainty of the cold calibration point.

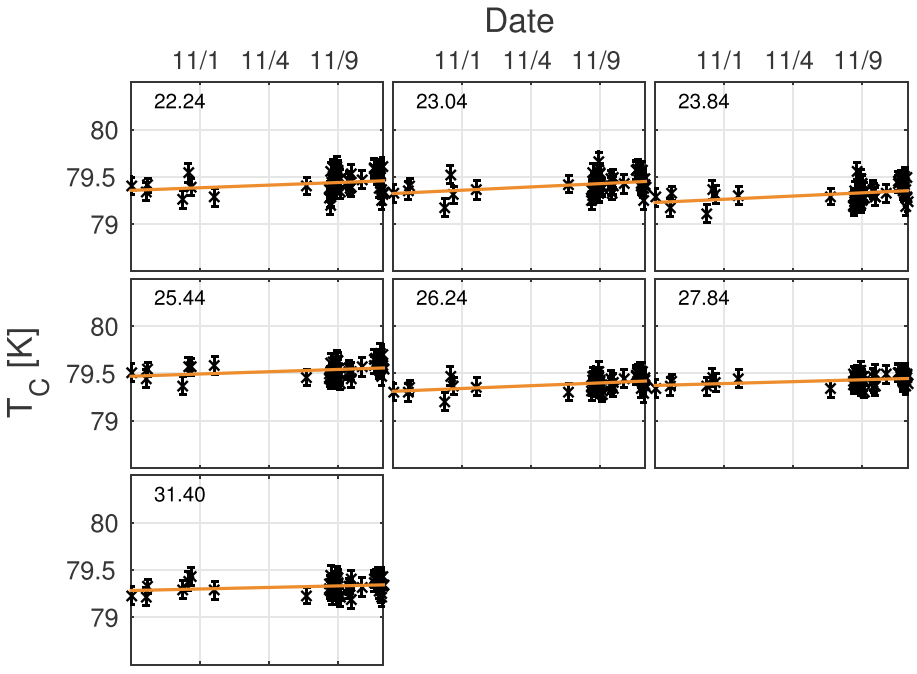

Figure 9. Cold target temperatures calculated from the RVSLN2 using TCCs that passed the $\sigma_{\Delta\nu}$ criterion. The error bars correspond to the uncertainty of the cold calibration point being $\sigma_{\Delta\nu}$ interpolated to 79.04 K.

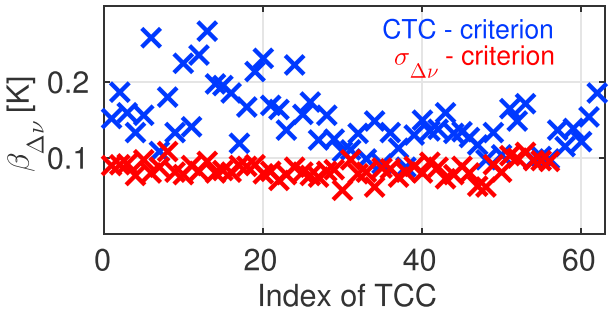

Figure 10. Spectral consistency $\beta_{\Delta\nu}$ using the RVSLN2. The TCCs are indexed in chronological order.

The mean drift rate of the regressions in Figure 9 (Table 2) is 8.9 ± 2.3 mK/d, which is very similar to the results of the CTC criterion. Hence, control measurements on a third blackbody close to a TCC (e.g., 1 day prior or afterward) would allow us to restrict the $\sigma_{\Delta\nu}$ criterion to an absolute value, i.e., $T_C \pm \sigma_{\Delta\nu}$, since the drift of T_R was negligible. Moreover, if the drift of T_R is known from a time series of TCCs as is shown in Figure 6 the $\sigma_{\Delta\nu}$ criterion will be a general applicable quality check.

The spectral consistency of TCCs was investigated in an analog way to the LN2cals. We defined the spectral standard deviation $\beta_{\Delta\nu}$ of the RVSLN2 for one single calibration, which is similar to the quality threshold $\sigma_{\Delta\nu}$, but uses an absolute voltage unlike $\sigma_{\Delta\nu}$ that is defined using a difference with respect to a reference signal. Figure 10 shows $\beta_{\Delta\nu}$ for TCCs that were accepted by the CTC and the $\sigma_{\Delta\nu}$ criteria. The use of the CTC criterion gives values up to 0.3 K. As expected, the $\sigma_{\Delta\nu}$ filter provides values distributed around 0.1 K which is equal to the $\sigma_{\Delta\nu}$ criterion threshold of 0.1 K. This shows that the CTC criterion does not necessarily ensure spectral consistent measurements that are a prerequisite for accurate multifrequency retrievals.

4.3.3. Scene Dependent Uncertainty Estimates

To complete the analysis of TCCs uncertainty, estimates will be derived. The error of the zenith SR B^{zen} (obtained from equation (A6)),

$$\Delta B^{zen} = \frac{\partial B^{zen}}{\partial \tau^{zen}} \Delta \tau^{zen}, \tag{5}$$

depends on the quality of the slope of the linear regression (calculated after *Fahrmeir et al.* [2010]) being represented by $\Delta\tau^{zen}$. Since B^{zen} is the cold reference during TCCs, ΔB^{zen} represents the uncertainty of the cold calibration point. When deriving equation (5), it was assumed that the mean atmospheric SR influences B^{zen} negligibly [*Maschwitz et al.*, 2013].

Since it was found for LN2cals that the error propagation is almost linear (see Figure 4), it is sufficient to obtain the uncertainty range for TCCs from a linear interpolation between the inaccuracies of the calibration references, i.e., $\pm\Delta T_C$ and $\mp\Delta T_H$. The maximum values of ΔT_C, which were found after applying the CTC criterion, were smaller than 0.6 K in all K band channels. Since the $\sigma_{\Delta\nu}$ criterion is independent of the linear regression, we consider the value of $\sigma_{\Delta\nu}$ itself as the uncertainty of the cold calibration point; hence, $\pm\Delta T_C$ is always smaller than 0.1 K.

5. Absolute Accuracy of the Cryogenic Load

The linear equations of the drifts of T_B at 79.04 K, due to changes of T_R (see Figures 8 and 9), were used to determine the cold target's temperature on 15 October. The two independent filter criteria of TCCs, i.e., the CTC and $\sigma_{\Delta\nu}$ criteria, yielded (fitted) temperatures between 79.20 K and 79.45 K. The RVSLN2 was recorded after an absolute calibration with LN2; hence, entrainment of oxygen into the bath was likely. *Paine et al.* [2014] found an increase of approximately 0.2–0.4 K after refilling the cryogenic load (the initial temperature was 79.04 K) that would correspond to a cold target's temperature of approximately 79.24–79.44 K while recording. This agrees with the values that were obtained from TCCs within 0.2 K. Since the bath's temperature was not measured, i.e., was not verified, the accuracy of the cold target's temperature at all channels is assumed to be ±0.5 K. This is a significant reduction of the uncertainty estimates found by *Maschwitz et al.* [2013] by up to 70% depending on the frequency channel. Moreover, the fact that the CTC criterion and the $\sigma_{\Delta\nu}$ criterion both "overestimate" the brightness temperature of the cold reference support the assumption that oxygen was entrained into the bath.

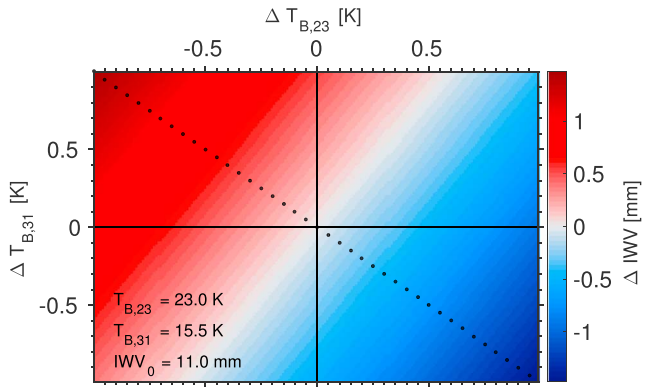

Figure 11. Error of the IWV due to biased input brightness temperatures. The IWV was retrieved after *Turner et al.* [2007] under clear-sky conditions.

Note that the cold reference's temperature determined with TCCs results from linear regressions of several TCCs which represents an average over MiRaCalE. Although the residual variances of the linear regressions in Figure 8 are smaller than 0.1 K in all K band channels, differences up to 0.9 K between the fit and TCCs that passed the CTC criterion were found which increases the total uncertainty. In contrast, the deviations between the fit and TCCs in Figure 9 are smaller than 0.4 K at all channels indicating a better performance of the $\sigma_{\Delta\nu}$ criterion. In total, these findings suggest that the CTC criterion, the $\sigma_{\Delta\nu}$ criterion and LN2cals reproduce the cold target's temperature reliably.

6. Spectral Consistency in Geophysical Retrievals

To determine the influence of spectral inconsistencies on geophysical retrievals, the spectral discrepancies at the calibration times must be determined and then converted into the uncertainties that correspond to the observed T_Bs. Although small spectral standard deviations were found across the seven frequency channels, differences up to 0.8 K between two V band channels were found at 79 K during LN2cals (Figure 3; e.g., between 51.26 and 52.28 GHz on 15/10), which increase toward smaller T_Bs.

A detailed analysis of several multifrequency retrievals is beyond the scope of this paper; nevertheless, the following example will sketch the main problem. The influence of spectral inconsistencies on the integrated water vapor (IWV) obtained by the retrieval of *Turner et al.* [2007] (hereafter IWVret) was investigated. IWVret requires T_Bs at 23.8 (T_{B23}) and 31.4 GHz (T_{B31}), environmental temperature, pressure, and relative humidity at the surface. The HATPRO-G4 provides T_Bs at 23.84 GHz and 31.40 GHz and is equipped with a meteorological station measuring environmental temperature, humidity, and wind. Introducing offsets ΔT_B to the input brightness temperatures T_B ($T'_B = T_B + \Delta T_B$) yields the error of IWV

$$\Delta\mathrm{IWV} = \mathrm{IWV}(T_{B23}, T_{B31}) - \mathrm{IWV}(T'_{B23}, T'_{B31}). \tag{6}$$

Figure 11 shows ΔIWV with surface values of 985 hPa, 287 K, and 43% relative humidity during clear-sky conditions. The homogeneous stratification of ΔIWV arises from the retrieval algorithm using a linear combination of T_{B23} and T_{B31} with different multipliers. The bisecting line contains all combinations of spectral differences allowing relative offsets between −2 K and +2 K.

A spectral difference of 1 K (e.g., $\Delta T_{B23} = -0.5$ K and $\Delta T_{B31} = +0.5$ K) induces an error of 0.73 mm IWV (0.73 kg m^{-2}, ≈ 6%). For example, the observation of the diurnal cycle of water vapor, which is on the order of 1 mm [*Steinke et al.*, 2015], would be affected by this spectral inconsistency by about 70% depending on the region and weather conditions. Moreover, this systematic bias does not include the uncertainties arising from the retrieval algorithm but increases the total uncertainty.

7. Summary and Discussion

Different calibration techniques for a state-of-the-art microwave radiometer (MWR) have been investigated. The main focus was on the analysis of calibrations with liquid nitrogen (LN2cal) and tipping curve calibrations

(TCCs) to determine their error sources, stability, and accuracy. Therefore, the calibration techniques have been examined both individually and comparatively. Additionally, new quality criteria for both techniques have been presented and tested and the drift rate of the radiometer was assessed.

Liquid nitrogen calibrations are influenced by several error sources, e.g., the entrainment of oxygen, the resonant effect, and the inaccuracy of the refractive index of liquid nitrogen. In the V band, deviations up to 0.5 K from the perfect cold calibration target temperature (T_C) were found during control measurements that were recorded directly after the calibrations. However, the V band channels are only slightly influenced by uncertainties of the cold calibration target, since uncertainties originating from the cold target decrease toward the hot calibration point where the atmospheric measurement range of the V band is located generally. The K band channels show smaller variations when observing a liquid nitrogen-cooled blackbody; however, uncertainties arising at T_C increase toward smaller scene brightness temperatures (T_Bs) and therefore increase the measurement uncertainty up to $\pm$ 0.7 K at low brightness temperatures.

It was found that the spectral comparison of control measurements after LN2cals is beneficial to identify biased calibrations. However, after the identification of a biased calibration, there will still be the need to repeat the LN2cal. This can be laborious since oxygen may have entrained into the bath during the first calibration, which will change the initial bath temperature if the container is simply refilled. Thus, either a second container is needed or the physical temperature of the blackbody must be measured with a thermistor that then introduces new error sources.

Two different quality criteria (filter techniques) were used to validate the quality of TCCs. The first criterion, which is typically used by the MWR community, is based on the statistical analysis of opacity-air mass pairs (CTC criterion), whereas the critical values that discriminate between a "good" and a "bad" calibration had to be determined subjectively. The second criterion ($\sigma_{\Delta\nu}$ criterion) that we have presented here is a more objectively based approach that uses a spectral comparison of brightness temperatures when observing a liquid nitrogen-cooled blackbody at seven frequencies and therefore preserves spectral consistency.

The dates on which TCCs passed the CTC criterion ($N = 62$) or the $\sigma_{\Delta\nu}$ criterion ($N = 53$) are marked in Figure 7. Although the number of TCCs that passed the two quality criteria is similar, the distribution in time is different and there are only two samples that passed both tests. An explanation has not yet been found.

For the first time, the accuracy of TCCs was assessed using the LN2 cooled blackbody as a third reference. Indeed, the accuracy estimates are strongly affected by the uncertainty of the third reference itself. The average T_B of the LN2-cooled blackbody, which was calculated with TCCs that passed both the CTC and the $\sigma_{\Delta\nu}$ criteria, agrees in both cases within 0.5 K with T_C. Maximum deviations from T_C considering a single TCC up to 0.9 K and 0.4 K were found using the CTC and $\sigma_{\Delta\nu}$ criteria, respectively. Thus, the $\sigma_{\Delta\nu}$ criterion, which is instrument specific and objectively chosen, provides accurate results and varies less than the CTC criterion from TCC to TCC. However, the CTC and the $\sigma_{\Delta\nu}$ criteria do not necessarily indicate that similar TCC samples are good. On time scales of several weeks, the $\sigma_{\Delta\nu}$ criterion will be influenced by channel-dependent drifts of the receiver noise brightness temperature (T_R), which must be taken into account.

The repetition frequency to recalibrate the MWR's characteristic parameters, and therefore to account for instrument drifts, was assessed during MiRaCalE. The instrument's equivalent noise temperature T_R should be updated at least every 50 days to keep the bias due to drifts smaller than 0.5 K in the K band. This update can be done (also more frequently) either with a tipping curve calibration or with a noise diode calibration. In the V band, T_R is updated during hot load calibrations, hence, with a much higher temporal frequency. Updating the gain every 2 min provides sufficient accuracy in the K band. The repetition frequency for LN2cals depends on the stability of the nonlinearity and the noise diodes. Neither could be estimated due to MiRaCalE's short running time and the small number of LN2cals that were conducted. Regarding the drifts of α and T_N found by *Maschwitz et al.* [2013], it seems sufficient to calibrate MWRs every 3 to 6 months to ensure accurate measurements.

The two different calibration techniques (i.e., TCCs and LN2cals) produced very similar T_Bs when observing the cold blackbody. The absolute accuracy of the blackbody's brightness temperature at the boiling point of LN2 was found to be $\pm$ 0.5 K at all frequency channels. This reduces the uncertainty of the cold reference during LN2cals, which has been assumed by *Maschwitz et al.* [2013], by up to 70% depending on the frequency channel. Furthermore, this suggests that the refractive index of LN2 should be reinvestigated to gain clarity, since current uncertainty estimates are based on a publication from 1983 [*Benson et al.*, 1983].

To further improve model evaluations and retrievals, the error sources as the entrainment of oxygen and the standing wave must be minimized. The former could be mitigated by a closed (covered) calibration target. However, the placement of a surface on which condensation is likely due to low temperatures adds a further uncertainty source. The latter could be realized by signal processing methods creating a reliable averaging method for the standing wave or by using an in situ probe to measure the bath temperature directly. Additionally, the influence of spectral inconsistencies on retrievals and models should be evaluated in more detail in order to provide reasonably accurate estimates of the retrieved quantities. This evolution would be one further step in ensuring that MWR data are reliable in data assimilation and model evaluations.

The approaches used here, and in particular the $\sigma_{\Delta\nu}$ criterion, should be generally applicable. However, it is important to note that the drift rates of the parameters are instrument specific (type/generation) and therefore not necessarily transferable to any radiometer type. To investigate the calibration accuracies and parameters drifts for other radiometer types, the second phase of the TOPROF (Toward Operational ground-based Profiling with ceilometers, Doppler lidar, and microwave radiometers for improving weather forecasts (European Cooperation in Science and Technology (COST Action ES1303))) Joint Calibration Experiment was conducted in fall 2015 in Meckenheim, Germany, and is currently under evaluation.

Appendix A: Calibration Techniques

This section gives a detailed description of the calibration techniques of the HATPRO-G4. Although the equations are adjusted for the used radiometer system, the principle of the calibration techniques is transferable to other radiometers.

A1. Liquid Nitrogen Calibration

To obtain the three unknowns (g, α, B_R), three calibration points are necessary. The HATPRO-G4 realizes one point with its built-in ambient temperature blackbody. The second calibration point is a pyramidal foam absorber blackbody immersed in LN2 (see Figure 1). The third reference is realized with built-in noise diode, which is used to inject additional signal into the radiometer's detector when viewing one of the other calibration targets. Hence, a new unknown, namely, the effective spectral radiance (SR) B_N of this noise source, is introduced. However, two additional calibration points close the equation system by looking at both of the calibration targets with the noise source on. Equation (1) can be rewritten accordingly

$$U_C = g(B_R + B_C)^\alpha \tag{A1}$$

$$U_H = g(B_R + B_H)^\alpha \tag{A2}$$

$$U_{CN} = g(B_R + B_C + B_N)^\alpha \tag{A3}$$

$$U_{HN} = g(B_R + B_H + B_N)^\alpha \tag{A4}$$

where the index C stands for cold (LN2) and the index H for hot (ambient target). The index N refers to the added noise signal. The equation system is solved numerically to obtain g, B_R, B_N, and α.

A2. Tipping Curve Calibration

The principle of TCCs is to replace the LN2 cooled blackbody by the cold clear sky. The technique uses observations at different air masses (a), i.e., at different elevation angles. For horizontally homogeneous conditions, the air mass scales linearly with the opacity as long as the opacity is small; hence, we are in the linear regime:

$$a(\theta) = \frac{1}{\sin(\theta)} = \frac{\tau(\theta)}{\tau^{\text{zen}}} \tag{A5}$$

where τ^{zen} is the zenith optical depth and $\tau(\theta)$ is the atmospheric opacity at elevation angle θ. Opacity-air mass pairs are obtained performing bilateral elevation scans at a fixed azimuth position. τ is derived from first-guess calibration values providing SRs $B(\theta)$

$$\tau(\theta) = ln\left(\frac{B_{\text{mr}}(\theta) - B_{\text{cb}}}{B_{\text{mr}}(\theta) - B(\theta)}\right). \tag{A6}$$

with the SR of the cosmic background B_{cb} and the mean atmospheric SR B_{mr} [*Han and Westwater*, 2000].

Equation (A6) follows from the simplified radiative transfer equation without scattering, whereas B_{mr} represents the opacity weighted atmospheric radiation. B_{mr} must be determined independently and was obtained prior to the calculations using the Monochromatic Radiative Transfer Model of *Clough et al.* [2005] that was driven using colocated radiosonde data. Because B_{mr} appears in both the numerator and denominator, equation (A6) has only a small sensitivity to any errors in B_{mr}. *Maschwitz et al.* [2013] found that a root-mean-square error of T_{mr} (B_{mr}) of 3 K has no significant impact (<0.1 K) on the calibration of the K band channels.

Linear interpolation of the opacity-air mass pairs yields a linear equation where the slope corresponds to the updated value of $\tau^{zen} = \tau(90°)$ that is used to recalculate the zenith SR solving equation (A6) for $B(\theta = 90°) = B^{zen}$. The latter serves as the cold calibration reference. Assuming that α has remained constant since the latest LN2cal, the equation system reduces to a two-point calibration (i.e., determining g and B_R) with the internal blackbody as the hot target. The process is repeated iteratively, until g and B_R converge to the optimal solution.

A TCC works only at frequencies with sufficiently low opacity, hence, when the cosmic background radiation is not completely attenuated at all angles which are used during the calibration procedure. During MiRaCalE, this was only the case for K band channels. Moreover, the moisture conditions only allowed TCCs down to three air masses in 0.5 air mass steps, as the opacities at larger air mass values were too high.

A3. Noise Diode Calibration

In contrast to the K band, the gain of the V band receiver fluctuates on time scales of seconds due to different receiver technologies. This fluctuation is compensated by periodically adding noise (B_N with 10 Hz) to the input signal, so that the scene SR can be obtained independent of the gain by taking the ratio of equation (1) with and without additional noise signal. Therein, α and B_R are assumed to be constant.

A 170 h burn-in process was performed by the manufacturer which provide a stable noise signal that is expected to change less than 0.1 K per month [*Radiometer Physics GmbH*, 2013]. In the K band, the noise injection is utilized to determine the system noise SR when the radiometer is viewing the internal blackbody by taking the ratio between equations (A4) and (A2) and then solving for B_R.

A4. Hot Load Calibration

Short-term fluctuations of the system's gain in the K band (over a period of a few minutes) are corrected using views of the internal target. The observation of a stable and known reference provides the opportunity to update the gain by solving equation (A2) for g, assuming that B_R and α are stable over longer time periods (weeks and months, respectively). In the V band, the gain is adjusted by permanent noise switching. The hot load calibration is used to update the system noise by solving equation (A2) for B_R.

Acknowledgments

MiRaCalE was hosted by the National Severe Storms Laboratory and the School of Meteorology of the University of Oklahoma in cooperation of the University of Cologne. Radio sounding data from the Southern Great Plains was provided by the ARM research facility. We would like to thank Rainer Haseneder-Lind, Doug Kennedy, and Birger Bohn for their assistance in the field. We would like to thank also Radiometer Physics GmbH for technical support. Comments from three anonymous reviewers were greatly appreciated and improved the quality of this paper. This work was supported in part by grant DE-SC0008830 from the U.S. Department of Energy as part of the Atmospheric System Research program. The data used are not available on the Internet, but free access is possible contacting nkuech@meteo.uni-koeln.de.

References

Bauer, P., A. Geer, P. Lopez, and D. Salmond (2010), Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation, *Q. J. R. Meteorol. Soc.*, *136*(652), 1868–1885, doi:10.1002/qj.659.

Benson, J., J. Fischer, and D. Boyd (1983), Submillimeter and millimeter optical constants of liquid nitrogen, *Int. J. Infrared Mill*, *4*, 145–152, doi:10.1007/BF01008973.

Clough, S., M. Shephard, E. Mlawer, J. Delamere, M. Iacono, K. Cady-Pereira, S. Boukabara, and P. Brown (2005), Atmospheric radiative transfer modeling: A summary of the AER codes, *J. Quant. Spectrosc. Radiat. Transfer*, *91*, 233–244, doi:10.1016/j.jqsrt.2004.05.058.

Crewell, S., and U. Löhnert (2007), Accuracy of boundary layer temperature profiles retrieved with multifrequency multiangle microwave radiometry, *IEEE T. Geosci. Remote Sens.*, *45*(7), 2195–2201, doi:10.1109/TGRS.2006.888434.

Crewell, S., H. Czekala, U. Löhnert, and C. Simmer (2001), Microwave radiometer for cloud cartography: A 22-channel ground-based microwave radiometer for atmospheric research, *Radio Sci.*, *36*(4), 621–638, doi:10.1029/2000RS002396.

Ebell, K., E. Orlandi, A. Hünerbein, U. Löhnert, and S. Crewell (2013), Combining ground-based with satellite-based measurements in the atmospheric state retrieval: Assessment of the information content, *J. Geophys. Res. Atmos.*, *118*(13), 6940–6956, doi:10.1002/jgrd.50548.

Fahrmeir, L., R. Künstler, I. Pigeot, and G. Tutz (2010), *Statistik: Der Weg zur Datenanalyse*, Springer, Berlin.

Fernandez, S., A. Murk and N. Kämpfer (2015), GROMOS-C, an novel ground-based microwave radiometers for ozone measurement campaigns, *Atmos. Meas. Tech.*, *8*, 2694–2662, doi:10.5194/amt-8-2649-2015.

Forkman, P., O. M. Christensen, P. Eriksson, J. Urban, and B. Funke (2012), Six years of mesospheric CO estimated from ground-based frequency-switched microwave radiometry at 57°N compared with satellite instruments, *Atm. Meas. Tech.*, *5*, 2827–2841, doi:10.5194/amt-5-2827-2012.

Han, Y., and E. Westwater (2000), Analysis and improvement of tipping calibration for ground-based microwave radiometers, *IEEE T. Geosci. Remote Sens.*, *38*(3), 1260–1276.

Hewison, T., and C. Gaffard (2003), Radiometrics MP3000 microwave radiometer performance assessment, *Tech. Rep.*, Met Office Observations Development, U. K.

Illingworth, A., D. Cimini, C. Gaffard, M. Haeffelin, V. Lehmann, U. Löhnert, E. J. O'Connor, and D. Ruffieux (2015), Exploiting existing ground-based remote sensing networks to improve high resolution weather forecasts, *Bull. Am. Meteorol. Soc.*, *96*, 2107–2125, doi:10.1175/BAMS-D-13-00283.1.

Kaisti, M., M. Altti, and T. Poutanen (2014), Uncertainty of radiometer calibration loads and its impact on radiometric measurements, *IEEE Trans. Microw. Theory Tech.*, *62*(10), 2435–2446, doi:10.1109/TMTT.2014.2349873.

Löhnert, U., and S. Crewell (2003), Accuracy of cloud liquid water path from ground-based microwave radiometry: 1. Dependency on cloud model statistics, *Radio Sci.*, *38*(3), 8041, doi:10.1029/2002RS002654.

Löhnert, U., and O. Maier (2012), Operational profiling of temperature using ground-based microwave radiometry at payerne: Prospects and challenges, *Atmos. Meas. Tech.*, *5*(5), 1121–1134, doi:10.5194/amt-5-1121-2012.

Löhnert, U., et al. (2015), JOYCE: Jülich observatory for cloud evolution, *Am. Meteor. Soc.*, *96*, 1157–1174, doi:10.1175/BAMS-D-14-00105.1.

Maschwitz, G. (2012), Assessment of ground-based microwave radiometer calibration to enable investigation of gas absorption models, PhD thesis, Univ. of Cologne. [Available at http://kups.ub.uni-koeln.de/5390.]

Maschwitz, G., U. Löhnert, S. Crewell, T. Rose, and D. Turner (2013), Investigation of ground-based microwave radiometer calibration techniques at 530 hPa, *Atm. Meas. Tech.*, *6*, 2641–2658, doi:10.5194/amt-6-2641-2013.

McGrath, A., and T. Hewison (2001), Measuring the accuracy of MARSS—An airborne microwave radiometer, *J. Atmos. Ocean. Tech.*, *18*, 2003–2012, doi:10.1175/1520-0426(2001)018<2003:MTAOMA>2.0.CO;2.

Paine, S., D. Turner, and N. Küchler (2014), Understanding thermal drift in liquid nitrogen loads used for radiometric calibration in the field, *J. Atmos. Ocean. Tech.*, *31*, 647–655, doi:10.1175/JTECH-D-13-00171.1.

Pospichal, B., G. Maschwitz, N. Küchler, and T. Rose (2012), Standing wave patterns at liquid nitrogen calibration of microwave radiometers, in *Proceedings of the 9th International Symposium on Tropospheric Profiling*, L'Aquila, Italy, doi:10.12898/ISTP9prc.

Radiometer Physics GmbH (2013), Instrument Operation and Software Guide, *Radiometer Physics*. [Available at http://www.radiometer-physics.de/.]

Rose, T., S. Crewell, U. Löhnert, and C. Simmer (2005), A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere, *Atmos. Res.*, *75*, 183–200, doi:10.1016/j.atmosres.2004.12.005.

Solheim, F., J. Godwin, and R. Ware (1998), Passive ground-based remote sensing of atmospheric temperature, water vapor, and cloud liquid water profiles by a frequency synthesized microwave radiometer, *Meteorol. Z.*, *7*, 370–376. [Available at http://www.schweizerbart.de//publications/detail/artno/025010704/Meteorolog_Zeitschrift_N_f_7_Jhg}.]

Shupe, M., et al. (2013), High and dry: New observations of tropospheric and cloud properties above the greenland ice sheet, *Bull. Am. Meteorol. Soc.*, *94*, 169–186, doi:10.1175/BAMS-D-11-00249.1.

Steinke, S., S. Eikenberg, U. Löhnert, G. Dick, D. Klocke, P. Di Girolamo, and S. Crewell (2015), Assessment of small-scale integrated water vapour variability during HOPE, *Atmos. Chem. Phys.*, *15*, 2675–2692, doi:10.5194/acp-15-2675-2015.

Stevens, B., et al. (2015), The Barbados cloud observatory—Anchoring investigations of clouds and circulation on the edge of the ITCZ, *Bull. Am. Meteorol. Soc.*, doi:10.1175/BAMS-D-14-00247.1.

Turner, D. (2007), Improved ground-based liquid water path retrievals using a combined infrared and microwave approach, *J. Geophys. Res.*, *112*, D15204, doi:10.1029/2007JD008530.

Turner, D., S. Clough, J. Liljegren, E. Clothiaux, K. Cady-Pareira, and K. Gaustad (2007), Retrieving liquid water path and precipitable water vapor from the atmospheric radiation measurement (ARM) Microwave radiometers, *IEEE T. Geosci. Remote Sens.*, *45*(11), 3608–3609, doi:10.1109/TGRS.2007.903703.

Ulaby, F., R. Moore, and K. Fung (1981), *Microwave Remote Sensing—Volume I*, Addison-Wesley, Reading, Mass.

Wulfmeyer, V., R. Hardesty, D. Turner, A. Behrendt, M. Cadeddu, P. Di Girolamo, P. Schlüssel, J. Van Baelen, and F. Zus (2015), A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles, *Rev. Geophys.*, 1–77, doi:10.1002/2014RG000476.

Xi, B., X. Dong, P. Minnis, and M. Khaiyer (2010), A 10 year climatology of cloud fraction and vertical distribution derived from both surface and GOES observations over the DOE ARM SPG site, *J. Geophys. Res.*, *115*, D1212, doi:10.1029/2009JD012800.