Final reply to anonymous referees

AMTD contribution by Löhner and Maier:
Assessing the potential of passive microwave radiometers for continuous temperature profile retrieval using a three year data set from Payerne

Original Referee comments are in italic
Replies to reviewer queries begin with a bold R
Added text is in courier style

Referee1 (C2563):
General
The contents of the article do not match the title and the abstract. In particular, the article deals with the comparison of the data set acquired by the HATPRO instrument with temperature profiles taken by sondes. In order to ‘assess the potential’ I would expect to see an estimation of the improvement of modern numerical weather forecast (NWP) when those data are used, especially in the light of the limitations (altitude resolution, altitude range) of the measurements taken with the aid of HATPRO. The authors state in the introduction and in the summary that the NWP would benefit from observation taking place every few minutes but this is not elaborated on.

R1a The reviewer is right, the title of the paper is misleading – by “assessing the potential” we mean to show how accurate microwave radiometer (MWR) retrieved temperature profiles are, where uncertainties have their origin and what needs to be improved to make such measurements potentially interesting for NWP. In order to avoid this misunderstanding, we have changed the title of the paper to: Operational profiling of temperature using ground-based microwave radiometry at Payerne: prospects and challenges.

Calpini et.al. (2011) deal with wind measurements and not with temperature measurements. They explicitly state, that humidity and temperature measurements are not yet assimilated into the model because of their preliminary state.

R1b Very true; the idea of this paper is to exactly show how preliminary the state of microwave radiometers is and what steps are necessary to make them indeed suitable for assimilation. To make this more clear, we therefore now state in the second paragraph of the Section 1: A European network of Wind Profilers has already been established within the CWINDE project and the ability of wind profilers to improve numerical weather prediction model performance has been demonstrated i.e. by Calpini et al. (2011). However, NWP models also require upper air temperature and humidity data for improved predictions, especially in the lower troposphere where severe weather is frequently triggered and satellite remote sensing capability is limited. Calpini et al. (2011) explicitly note that MWR are still within a validation phase and have therefore not been operationally assimilated into the MeteoSwiss weather forecast model. This paper intends to contribute to this validation phase with respect to continuous temperature profiling and make clear what current MWR are capable of in terms of accuracy, where error sources can be found and how systematic errors can be minimized during operational measurements.
I would therefore suggest to either append the article with a study to actually ‘assess the potential’ of the HATPRO measurements or to move the article to another journal which is specifically devoted to the description of data sets (e.g. AMT’s sister journal Earth System Science Data, ESSD). In my view, the first would be very beneficial for designing observation strategies and to provide information on were to improve the instrumentation further. It would increase the scientific value of the manuscript considerably.

R1c We very much agree with the reviewer, that showing the impact of MWR measurements on the NWP model forecast would be significant addition of scientific value. However, this is, as stated above, not part of the original intent of this paper. Estimating the improvement of NWP models when using MWR data would be going towards data assimilation experiments, i.e. either an OSE (Observation System Experiment) or an OSSE (Observation System Simulation Experiment). Both require an enormous amount of scientific and financial resources. The authors are both involved in MWRnet (http://cetemps.aquila.infn.it/mwrnet/), where international efforts are currently being pursued to organize such experiments within the next year. However, we disagree with the reviewer that the current paper exclusively deals with the description of a data set. From our point of view, this paper shows how to handle operational MWR measurements properly in order to come from brightness temperature measurements to temperature profiles. In order to characterize the quality of brightness temperature measurements and retrieval methods, we need to compare to radiosonde data. This comparison makes it possible to find out under which conditions MWR are suited for retrieving temperature profiles, what errors may occur and how these can be eliminated. In this sense we see a strong link to the objectives of AMT.

Please relate your findings to other publications dealing with the same or a similar topic, in particular Cimini et.al. (2006) and Crewell and Löhnert (2007).

R1d Cimini et.al. (2006) and Crewell and Löhnert (2007) are referred to in the third paragraph of Section 1, also in the first paragraph of Section 6.2.

Specific
It would be very helpful if a time series of temperatures at a particular altitude level would be presented. Examples of temperature profiles recorded by the HATPRO instrument and the sonde would help in assessing the statements too.
Figure 5: HATPRO-retrieved time-height temperature contours between 0 and 4 km on 09 September, 2008.

Figure 6: Temperature profile at 0 (left) and 12 (right) UTC on 09 September 2008 between 0 and 4 km. The bold line shows the radiosonde profile, the dotted line the HATPRO retrieval and the dashed line the radiosonde profile smoothed with the averaging kernel.

R1e Yes, we fully agree with the reviewer. Instead of a temperature time series at one level we have added a 24 hour temperature time-height contour that shows how the Payerne MWR can capture the diurnal cycle of the boundary layer (BL) on a late summer day. Additionally, we have included a figure showing the direct MWR-radiosonde comparison at 0 and 12 UTC on that day. We have added the following to the manuscript text in the second paragraph of Section 6: As an example of a continuous MWR temperature profile retrieval over 24 hours, Fig. 5 shows a time-height contour of a late summer day with a typical development of the BL starting from a night-time inversion towards a well mixed day-time BL. After sunset, IR cooling induces again an inversion close to the surface. A warming of ~4 K in the altitude range between 2 and 4 km can also
be observed throughout the course of the day. Both at 0 and 12 UTC the MWR retrieval is able to accurately reproduce the radiosonde profile in the region below 1 km (Fig. 6). Especially the low-level inversion at night is nicely captured. Above 1 km, the limited vertical resolution of the MWR measurements become obvious, i.e. during both cases the lifted inversions are not retrieved.

Please mention ‘section 6’ in the end of the introduction.

R1f Done

The authors pose, among others, the question of how good HATPRO performs during extreme conditions. I gather this question is dealt with in section 6.4, which is named ‘significant weather’. I wonder if the passing of a front is the only extreme condition which affects the measurements of the HATPRO. What is about very low or very high temperatures?

R1g: We have now replaced the old Fig. 8 by the following figure (Fig. 12 in revised manuscript), which additionally shows temperature retrieval performance during inversions, cold extreme and warm extreme cases:

![Temperature profile differences (STDEV) during all-sky (1816), frontal (252), warm extreme (330), cold extreme (332) and temperature inversion (165) cases (see text for details) from August 2006 – December 2009 between radiosonde and HATPRO measurements.](image)

We have modified the text in Section 6.4 related to this figure to:

The so far shown results are valid for an average over all cases. However, it is also necessary to show how the MWR temperature retrieval performs in frontal or extreme atmospheric conditions, where it is especially important for models to receive accurate measurement input. The following subsets of cases are extracted from the radiosonde data set: frontal conditions, warm extremes, cold extremes and extreme temperature inversions.

A situation is classified as frontal when available records report a front or occlusion crossing Zürich (170 km NE of Payerne) up to 12 hours before or after the launch of a Payerne radiosonde. The warm extremes are classified as cases when the surface temperature is outside of the 1-sigma range as a positive deviation, whereas cold extremes are classified as cases when the surface temperature is outside the 1-sigma range as a negative deviation. If the radiosonde
profile shows a continuous temperature rise with height below 4 km throughout a 500 m height interval, the profile is classified as an extreme temperature inversion case. As can be seen in Fig. 12 the STDEV can vary quite considerably with height depending on the subset of cases evaluated. Throughout the lowest 4 km the STDEV during frontal passages are even up to 0.5 K smaller than in the all-sky cases. This is due to missing inversions during frontal passages, in contrast to fair weather situations, i.e. during winter anti-cyclonic situations or when residual layers occur at night after a sunny day. This becomes clear when comparing to other extreme cases. In the 165 evaluated extreme inversion cases MWR temperature retrieval STDEV is smaller than 1 K below 400 m but is then large than 2 K already at 2 km. Inversions are detected, however their amplitude and sharpness are smoothed out, leading to large differences when comparing level-to-level with the vertically highly resolved soundings. The STDEV for the cold curve resembles the inversion curve because cold extremes (i.e. night-time inversion due to strong radiative cooling in winter) because 45 % of the analysed inversion cases are also classified as a cold event. Warm extremes, which can be associated with well-mixed boundary layers, show up to 0.5 K better STDEV than the all-sky cases. For these cases, only less than 1 % are also classified as an inversion. Similarly, only less than 5 % of the frontal cases are classified as an extreme inversion event. This makes clear that the performance of the MWR temperature retrieval can be considered as reliable during frontal conditions. However, quality-controlled data availability decreases down to 60% in these cases, which is a result of the activated internal HATPRO precipitation flag. Hence, temperature profiles during frontal passages can only be characterized well at times when precipitation is not reaching the surface and thus obscuring the MWR measurement. During all of the subsets analysed above, the BIAS between MWR and radiosonde varies between -0.5 and +0.5 K as a function of height.

The authors cite a study by Löhnert and Crewell (2003) stating that the HATPRO(?) measurements contain profile information on the temperature throughout the atmosphere. Later, however, in page 7452 line 5 the authors state that 95 % of the information are from the lowermost 4 km. At first sight, this sounds contradictory. Please clarify.

R1i: Indeed, the formulation in the first paragraph of Section 3.1 information on the temperature throughout the atmosphere is not precise. Instead we now state: ...the seven channels of the second HATPRO band (51.26, 52.28, 53.86, 54.94, 56.66, 57.30, 58.00 GHz) contain information on the vertical profile of temperature within the lower and middle troposphere due to the homogeneous mixing of O₂ (Crewell and Löhnert 2007).

The HATPRO section

Either omit details and cite a publication (I assume Rose et.al., 2005) or keep details but make sure they are correct and describe the instrument to a necessary degree.

R1i: We agree. We have shortened the calibration section to the necessary degree and refer to Rose et al. (2005) for further details in the last sentence of Section 3.1

Two examples:

page 7442 line 16-17
Polystyrene is transparent for millimeter waves, so what is the actual black body? My suggestion: only state that black body measurements are made and refer for any details to the relevant publication.

R1j: The black body approximation used as calibration target consists of a piece of carbon loaded foam. With polystyrene we were referring to the material of the box surrounding the black body, which contains the liquid nitrogen in the case of the cooled calibration target. We have now omitted this formulation to avoid any misunderstanding.

page 7442 line 19
This is not quite true and depends on the exact definition of the Brightness temperature used. Refer to Janssen (1993) for more details. Again, keep only the relevant statements and refer to the instrument publication (Rose et.al., 2005 ?) for anything else.

R1k: True, this depends on the exact definition of brightness temperature (TB). For the Rayleigh-Jeans equivalent TB, TB and physical temperature of a black body have approximately a constant offset in the microwave range, however for the Planck-equivalent TB this holds true by definition. To make this clear, we now formulate in Section 3.1: Note that when looking at a perfect black body the physical temperature and the Planck-equivalent brightness temperature are identical. Of course in praxis, when looking at the cold load during the calibration, a correction due to the non-perfect black body characteristic (reflectivity of LN2 surface > 0) must be applied.

section 6.2
How many observations are actually done during conditions which are not considered clear sky?

R1l: With all-sky conditions we mean clear sky and cloudy. In Section 6.1 we analyse the 487 clear sky conditions, these are part of the 1816 all-sky cases analysed in Section 6.2. We now state more clearly in the beginning of Section 6.2: Next, the retrieval performance is evaluated using 1816 simultaneous radiosonde / MWR measurements during all-sky conditions. Additionally to the clear-sky cases, this data set also contains cloudy cases. The retrieval algorithm coefficients, which are applied to this data set, are derived from all 1992-2009 radiosondes as seen in Figure 2 and described in Section 5.

How do measurements perform if the measurements are not during clear sky conditions but pass the quality checks? Is there a bias between measurements during clear sky and overcast conditions?

R1m: The STDEV between MWR and sonde is only slightly worse when regarding only quality controlled cloudy conditions. For reasons of clarity and better overview, we have decided not to show this line additionally in the new Fig. 8 (see below). Both in clear sky and all sky conditions the bias is mostly below 0.1 K – this implies that also during cloudy sky conditions the bias is within this range.

page 7542 line 2pp
Please provide an example for the AVK. The authors stated that 4 pieces of independent information can be retrieved from the HATPRO measurements. In the following and in the figures (e.g. fig 5) the impression is created, that the temperature profile is rather highly resolved in altitude.
R1n: We thank the reviewer for this good idea. We have now included a new figure (Fig. 7) showing the averaging kernels for the mean Payerne temperature profile for six distinct heights, as well as the cumulative distribution of the degrees of freedom of signal with height.

![Figure 7: Temperature averaging kernels (δT_{retrieved,i}/δT_{true,j}) for the mean temperature profile from the complete Payerne data set (1992-2009) as a function of height shown for different heights of perturbation (left). Cumulative Degrees of Freedom (DOF) for temperature with height (right). The DOF at each height corresponds to the diagonal value of the averaging kernel. The four different curves correspond to different seasons (summer/winter) and different radiosonde ascent times (0/12 UTC) derived from complete Payerne data set (1992-2009).](image)

In the third paragraph of Section 6 we have now modified the text to: Using the average (1992-2009) Payerne profile of temperature, humidity and pressure, the number of independent pieces of information contained in the TB for retrieving the temperature profile (degrees of freedom for signal, DOF) can be determined following Rodgers (2000). The calculated averaging kernels for 6 distinct heights are shown in Fig. 7, which are defined by the sensitivity of the retrieved value at height with index i to the true value at height j (δT_{retrieved,i}/δT_{true,j}). Note, a perfect vertical resolution would give rise to a delta function with a value of 1 at height i=j and 0 at all other heights j≠i. Thus, the broadness of the averaging kernels give information on the vertical resolution. The diagonal components of the averaging kernel matrix correspond to the DOF for each level and the trace of the averaging kernel matrix yields the total DOF, which are ~4 in case of the temperature retrieval. If the cumulative distribution of the degrees of freedom with height is regarded (Fig. 7), one can conclude that about 85% (95%) of the temperature information originates from the lowest 2 km (4 km).

Have the radiosonde data been smoothed using the AVK matrices? What is the correlation between temperature obtained from HATPRO measurements at different altitude levels?
R1o: No, up to now we only show level-to-level comparisons. Note, that separate sets of linear regression coefficients were derived for every height level using all V-band TBs to retrieve only the temperature at exactly the levels mentioned in Section 5.2. However, we fully agree with reviewer that it is important to show comparisons without the retrieval smoothing error. In the last paragraph before Section 6.1 we now state: The radiosonde profile can be brought onto the vertical resolution of the MWR temperature retrieval by the following multiplication (averaging kernel smoothing):

\[ T_{\text{smooth}} = T_{\text{retrieved}} + A(T_{\text{true}} - T_{\text{retrieved}}) \]  

(1)

This accounts for the limited vertical resolution of the MWR temperature retrieval and the resulting differences can now be analysed more precisely towards measurement, forward modelling or statistical representativeness error. In Fig. 6, the smoothed radiosonde profile is very close to the MWR retrieval in the lowest 2 km, however it becomes clear that the MWR retrieval cannot resolve the exact details shown within the radiosonde profile, i.e. the night-time isothermal layer from 1 to 1.3 km. During night above 2 km, a persistent underestimation (1-2 K) of the MWR retrieval occurs between radiosonde and retrieved profile as well between smoothed and retrieved profile. This points to the fact, that this underestimation is not due to vertical resolution effects but is more likely due to one of the error sources mentioned above.

Also in the second paragraph of Section 6.2 we state: Additionally, Fig. 8 shows the BIAS and STDEV values using the smoothed radiosonde profiles (Eq. 1). Whereas the BIAS using the smoothed and non-smoothed radiosonde values do not differ, the STDEV values are much lower and range from 0.3 to 0.7 K in the lowest 4 km. These numbers are the errors to expect from then random uncertainty in TB measurement (0.2 - 0.5 K) as well as from the radiosonde measurement. Hereby, the latter is due not only to the sensor uncertainty of 0.2 K (Tab. 2), but also due to temporal delay and spatial drifts of the radiosonde with respect to the quasi-instantaneous MWR measurement.

We have also added an additional figure (Fig. 10) showing the level-to-level error correlation of temperature for six distinct reference heights.
Figure 10: Temperature error covariances calculated from 1816 all-sky conditions from August 2006 - December 2009 between HATPRO and radiosonde measurements. The error covariances are shown for six specific reference height levels. Asterisks show the reference heights.

With respect to this Fig. 10 we now state in Section 6.2:
In order to assess the retrieval uncertainty in a complete way, it is also necessary to calculate the level-to-level error covariances. These give information on the correlation of the temperature error in height $i$ with the temperature error in height $j$ as shown in Fig. 10 for six distinct heights. The relatively sharp and low maximum peak values below 1 km underline the high potential for BL temperature profiling, whereas increasing broadness and maximum peak values of the error covariance curves at heights above 2 km again underline decreasing vertical resolution and accuracy with increasing height. Generally the covariance information is necessary when assimilating MWR temperature data into NWP models, i.e. within a 3D or 4DVAR assimilation system.

page 7455 line 12
Is non-crucial the proper expression? Do the authors mean 'insensitive' or 'little affected'?  
R1p: We now state in Section 6.4: ... the MWR temperature retrieval can be considered as reliable during frontal conditions.

Figure 5 to 8
What is actually shown in these figures? From the text I understand, shown are the bias, i.e. the difference of MWR and sonde temperatures and the standard deviation of those differences. If this is true the labels are wrong and should be corrected.  
R1q: The figures have been replaced (Fig. 8 – Fig. 12) and the labels have been corrected and now state STDEV.
Figure 7
Labels and tickmarks are too small.

R1q: Due to only small differences in the comparison of the day/night STDEV lines, Fig. 7 (now Fig. 11) has been modified to show only the BIAS comparisons. In the paper labels and tickmarks are now larger.

Figure 11: Temperature profile differences (BIAS) during all-sky conditions from August 2006 - December 2009 between MWR and radiosonde measurements differentiated by day and night; red lines show the retrieval results only at 12 UTC, whereas black lines only at 0 UTC. 920 12 UTC, respectively 896 0 UTC MWR measurements passed the quality control and are evaluated in the plot. Dashed lines show retrieval results using 2m-temperature (Ts) as an additional predictor in the linear regression.

Referee2 (C2718):

...  
In my point of view modifications are needed in your manuscript to substantiate the general validity of the conclusions. What I mean is as follows: The authors advertise in the title to assess “the potential of passive microwave radiometers for continuous temperature profile retrieval”. Therefore observations of the microwave profiler HATPRO are applied and the system is described in Section 3.1. So far so good; but in the following the notation HATPRO is used as additional term for everything (HATPRO measurement, HATPRO brightness temperature, HATPRO temperature retrieval, HATPRO precipitation sensor, HATPRO data set, HATPRO quality control, HATPRO retrieval coefficients, HATPRO/radiosonde cases, etc.). The authors should either distinguish between HATPRO relevant statements (i.e. offset, bandpass and beamwidth studies) and conclusions which are generally valid for microwave profiling, or modify the title.

R2a: We have now distinguished between HATPRO relevant issues and conclusions, which are generally valid for microwave temperature profiling throughout the text. In the latter case we have exchanged HATPRO with MWR.

After corrections and clarifications the manuscript should be accepted for publication.
SPECIFIC COMMENTS
Page 7439 line 18pp
The authors apologize to the reader for analysing a period in the past and be afraid possibly not to describe the “state-of-the-art MWR anymore”. Readers become discouraged by this statement and I propose to delete the phrase for this reason. As a kind of compensation authors offer (line 20) insider information “when relevant to the analysis”. Partly they aren’t relevant because they have no influence on the conclusions of the paper.
R2b: After carefully considering this comment, we have deleted the four sentences after the questions addressed bullets in Section 1.

Example: page 7449 line 2, “... the random blower and heater are now operated with maximum power in the latest RPG HATPRO software version.”.

R2c: We have reformulated this sentence in the third paragraph of Section 5.3 to a general operations requirement: In order to prevent this, MWR radome blowers and heaters should be operated with maximum power and continuous checks of the radome state should be carried out during calibration.
Example: page 7441 line 20, “Just recently the manufacturer has acquired the possibility of measuring the channels bandwidths precisely...”.

R2d: We have replaced this sentence with in the first paragraph of Section 3.1: Typical HATPRO channel bandwidths are illustrated in Figure 1.

R2e: Yes, we clearly see the argument of the reviewer and have omitted terms such as now and just recently. At the end of Section 4 we have reformulated the sentence to a general recommendation: This clearly demands for sophisticated automatic RFI filters, as well as for quality controls concerning the sanity of the receiver system.

We have left the remaining two passages in Center-frequency offset and Bandpass effect in Section 5.3 unchanged, because we argument on the basis of changes made between HATPRO G1 and HATPRO G2 instruments.

Page 7447 line 21-22
The authors should provide some more details about the operating schedule.
- What is the reason to perform elevation scans every 15 min?
- What is done between?
- Are temperature profiles available calculated from zenith observations only?
- If yes, how do they agree with elevation scan retrievals?

R2f: In the first paragraph of Section 5.2 we have now added the following text:
Elevation scans were performed every 12-13 min, so that high-quality temperature profiles are available approximately four times an hour. In between, zenith TB observations are performed to allow the measurements of atmospheric water vapor and liquid water path.

We also mentioned in the very beginning of Section 5.1:
At the opaque center of the O₂ absorption complex at 60 GHz, most of the temperature information originates from near the surface, whereas further away from the line, the atmosphere becomes less and less opaque so that more and more information also originates from higher atmospheric layers. Using the frequency dependent information
alone at zenith results in the order of ~2.5 independent pieces of vertical temperature information throughout the troposphere (Löhnert et al. 2009).

In the then following second paragraph of Section 5.1 we build up an argumentation which shows why we use additionally elevation scanning for temperature profile retrieval.

Further down (page 7453 line 17) is stated “The high accuracies below 1 km are due to the information contained in the elevation scans”. That is not proven in the manuscript. Nevertheless, this is an interesting and important point for an evaluation of the potential of continuous monitoring. It would be valuable if authors could illustrate the effect with observed data.

R2g: This has been shown by Löhnert et al. (2009), which we now refer to in the first paragraph of section 5.1, but also by Crewell and Löhnert (2007). Hence, we do not think it is necessary to repeat this analysis here. In this respect we have added a new Fig. 10 (see above reply to reviewer R10), which shows the temperature error covariances and can be used to theoretically underline the high vertical resolution in the BL. See text added in Section 6.2 with respect to Fig. 10 (R10).

Page 7451 line 14
Fig.3 shows systematic differences for selected frequencies and 90 elevation. An additional table or text extensions containing numerical values of TB differences for zenith and other elevation angles would make it easier for the reader to be aware of this important fact.

R2h: Yes, we agree. We have added a new table (Tab. 3), which shows the TB offsets of the of the first four V-band channels for all elevation angles used.

Table 3: Clear sky offsets (MWR – simulated from sonde) and standard deviations (both in K) for the MWR elevation angles and optically thin frequency channels during the period between the fourth and fifth LN2 calibration (268 cases).

<table>
<thead>
<tr>
<th>Elevation angle / channel in GHz</th>
<th>51.26</th>
<th>52.28</th>
<th>53.86</th>
<th>54.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>90°</td>
<td>3.96/1.70</td>
<td>0.23/1.32</td>
<td>1.50/0.51</td>
<td>-0.35/0.39</td>
</tr>
<tr>
<td>42°</td>
<td>5.70/1.73</td>
<td>0.71/1.22</td>
<td>0.96/0.43</td>
<td>-0.03/0.47</td>
</tr>
<tr>
<td>30°</td>
<td>5.51/1.99</td>
<td>-0.01/1.26</td>
<td>0.59/0.38</td>
<td>-0.04/0.42</td>
</tr>
<tr>
<td>19.2°</td>
<td>5.04/2.33</td>
<td>-0.49/1.29</td>
<td>0.35/0.38</td>
<td>-0.08/0.44</td>
</tr>
<tr>
<td>10.2°</td>
<td>2.22/1.48</td>
<td>-0.89/0.62</td>
<td>0.14/0.40</td>
<td>-0.15/0.47</td>
</tr>
<tr>
<td>5.4°</td>
<td>0.52/0.95</td>
<td>-0.44/0.42</td>
<td>-0.10/0.41</td>
<td>-0.20/0.46</td>
</tr>
</tbody>
</table>

We now state at the end of the second paragraph of Section 5.3:
TB offsets are also present in the off-zenith observing directions as shown in Tab. 3, where the necessity of considering the offset correction for the optically thinner frequency channels even down to low elevation angles becomes apparent.
Page 7453 line 24pp
The summary “… we can say HATPRO can deliver reliable and accurate temperature profiles in 88% of all-sky cases … and an uncertainty from … to …” sounds too strictly rather as a natural law than a result of your Payerne study. Probably other availability rates and uncertainty ranges could be detected at other sites (e.g. tropics or Arctic).

R2i: At the end of Section 6.2 we have reformulated to: In summary the Payerne MWR was able to deliver reliable and accurate temperature profiles in 86% of all-sky cases, with an uncertainty ranging from 0.5 K in the lower boundary layer and rising up to 1.7 K at 4 km height.

Page 7455, Section 6.4 Significant weather
I’m not sure if the findings of this section can be really assigned to a category ‘Significant weather’. Authors admit that only 60% of the cases, classified as frontal, passed the quality-control. The question remains, if precisely those 40% of cases which had been rejected aren’t more typical for those weather situations defined as ‘significant’ by the authors.

R2j: These 40% of the frontal cases are the times when precipitation reaches the ground and obscures the MWR measurement by droplets and smaller puddles on the radome. To make this clear we have added in Section 6.4: However, quality-controlled data availability decreases down to 60% in these cases, which is a result of the activated internal HATPRO precipitation flag. Hence, temperature profiles during frontal passages can only be characterized well at times when precipitation is not reaching the surface and thus obscuring the MWR measurement.

Further, I don’t understand (line 16pp), why are the regression coefficients not derived for the subset ‘significant weather’? In the manuscript is declared that MWR radiances were calculated for approximately 10000 radiosondes (page 7438).

R2k: True, this would be a possibility to consider; however as shown in Fig. 12, the linear regression deals with the frontal passages in a satisfactory way. One could also consider deriving special retrievals for inversion conditions, different seasons, weather regimes and so forth. The difficulty then is to objectively characterize the actual situation between two radiosonde ascents, which are mostly 12 hours apart or maybe not even available at the MWR site. Also, changing from one set of regression coefficients to another between two measurements may lead to discontinuities in the results, which would require re-processing of the retrieval results. We have made the experience, that using one all-weather all-season set of regression coefficients leads to an easy and robust operational application as well as an error that is relatively easy to characterize.

Moreover, why are nevertheless the rms errors smaller than those for all cases although the regression coefficients are not associated with this subset? Please, clarify!

R2l: This is due to the smaller amount of inversions (< 5 %) in the frontal data subset in comparison to the whole data set (~10%). We have now explained the effect that leads to this accuracy reduction in section 6.4. Please refer to reply R1g.

Fig.5-8
In the text exclusively the expression ‘STDEV’ is used and in the legend the notation ‘rms’. The captions refer only to ‘Temperature differences’. If you mean always the same, please, say so; otherwise clarify!
R2m: The new figures (Fig. 8 and Fig. 9) have replaced the old figures (Fig. 5 and Fig. 6) where we have corrected the legends as well as the caption text. This holds true also for the new Fig. 10 and Fig. 11 (formerly Fig. 7 and Fig. 8), see R1q and R1g.

Figure 8: Temperature profile differences (BIAS and STDEV) during clear-sky conditions from August 2006 – December 2009 between MWR and radiosonde measurements. The MWR retrieval coefficients and the TB-offset correction were derived using exactly the same clear-sky measurements as considered in the evaluation. Black lines show the retrieval results without using the systematic TB offset correction while green lines show the results applying the systematic TB offset correction (OC). Additionally, in red the STDEV is shown after smoothing the radiosonde profile with the averaging kernel (AVK). A total number of 486 matching MWR/radiosonde cases are evaluated in the plot.

Figure 9: Temperature profile differences (BIAS and STDEV) during all-sky conditions from August 2006 – December 2009 between MWR and radiosonde measurements. The MWR retrieval coefficients were derived from the whole 1992-2009 radiosonde data set, whereas the TB correction was derived only from clear-sky measurements as in Figure 8. Black lines show the retrieval results without using the systematic TB offset correction while green lines show the results applying the systematic TB offset correction (OC). Additionally, in red the STDEV is shown after smoothing the radiosonde profile with the averaging kernel (AVK). A total number of 2107 matching MWR/radiosonde cases were considered; in 1816 cases the MWR measurements passed the quality control and are evaluated in the plot.
Fig. 7
Legend and labels are too small. I recognize quite similar plots in the left and right panel. Differences are apparent only for the levels from 0 to 300m. I think the authors should accentuate these levels more clearly by fitting the y-axis.

R2n: Fig. 7 is now the new Fig. 11 (see R1q) with better visibility.

MINOR COMMENTS
Page 7436, in the abstract HATPRO should be introduced as ‘microwave’ profiler
R2o: Done.

Page 7449 line 4, “third” should be “fourth”
R2p: Changed.

Page 7452 line 23, “m.s.l.” should be “a.g.l.”
R2q: m.s.l. has been omitted.

Page 7468 Fig. 7, “A total number of 2088 HATPRO/radiosondes were available” is pasted from Fig. 6 into the caption of Fig. 7. This part of sentence is dispensable because no results are shown regarding this data set
R2r: A total number of 2088 HATPRO/radiosondes were available has been omitted in the caption of Fig. 11 (see R1q).

Referee3 (C2630):
...
From this data set the authors have shown that this radiometer is very stable over very long period of time (year) which is a very important point and have produced some part of the error characterisation of the retrieved profile. What is missing to my point of view is the vertical correlation of the error and some comments on the operational feasibility of such calibration method. So with a better characterisation of the error, (see comments), I think this paper is worse to be published, because it shows that at least the radiometer (for the V band) considered in this study is stable over long period of time. It shows to the community that LN2 calibrations are not so easy to perform, can introduce significant bias and need to be checked. Work still needs to be done to get reliable calibration.

Specific Comments
1) With such a data set and method, the brightness temperatures are extremely well calibrated to the data set and the type of retrieval used. Effects due to radiatif model, centre frequency shift, band pass beam width effect are fully absorbed in the calibration and are completely coherent with the statistical method which use the centre frequency, pencil beam approximation and a larger data set of radiosondes from Payerne. Because this radiometer is very stable this calibration will remain valid out side of the data set. In the clear sky condition comparison, it would be interesting to know what is the standard deviation of the bias computed for the period used for figure 5 for each of brightness temperature, this will show some how the link between the noise on Tb and the accuracy of a profile, for a perfectly well calibrated radiometer.
R3a: The new Tab. 3 (see R2h) now also gives information on the standard deviation of the calculated TB offsets as a function of angle and frequency channel for the longest sequence between two calibration (fourth and fifth), excluding the optically thick channels. While the
standard deviation is around or less than 0.5 K at 53.86 and 54.94 GHz at all used elevation angles, the 51.26 GHz channel shows a maximum of above 2 K at 19.2° elevation, whereby the 52.28 GHz value is approx. constant at 1.3 K down to 19.2 and then decreases to below 0.5 K at 5.4°. This noise of the bias is clearly related to the optical depth, which in turn is related to the absorption coefficient. So next to the remaining MWR noise, these values may also reflect a random uncertainty of the \( \text{O}_2 \) absorption model. A further contribution to this noise is certainly also the spatial drift and the temporal delay of the radiosonde measurements during ascent with respect to the instantaneous MWR measurement. We now state in the second paragraph of Section 5.3:

The temporal variation of the TB offsets (Tab. 3) is less than 0.5 K for the channels that are more optically thick. In contrast to this, the two optically thinner channels (51.26 and 52.28 GHz) show variations up to 2.5 K. Since these channels are more transparent, the effects of the time delay and the spatial drift of the radiosonde with respect to the instantaneous measurement MWR are more evident. Additionally, these variations may arise from radiometric noise as well as from random uncertainties in the oxygen absorption model.

2) For these results to be reproducible we need a perfectly well calibrated radiometer. What the authors should address is how many radiosondes are required to calibrate the radiometer. The radiometer is quite stable but on the lower channel one can see some time evolution in the bias. Would the calibration done using the data in April May 2008 gives equivalent results on the retrieved profile that the data used in June July 2008. Is this calibration method something that can be reasonably done if the radiometer is not located at a radiosonde site. Or could the same quality of calibration be obtained using NWP model profile. It would be interesting to see if some of the offset are not brightness dependant in particular in the lowest channel.

R3b: The data shown in the manuscript is actually already using a TB offset correction as a linear function of the current TB measurement itself. The following figure demonstrates the TB offset (y-axis) dependence on the TB value itself (x-axis) for the 51.26 GHz channel during the different calibration sections:
Note that this channel shows the strongest dependence on TB. We tested this TB correction scheme in comparison to using the constant values between two subsequent calibrations and found impacts far significantly less than 0.1 K on the retrieval BIAS and STDEV. We have now made this explicitly clear in the text (last paragraph of Section 5.3): For each measurement period between two subsequent LN\textsubscript{2} calibrations, a separate offset correction procedure is developed and applied to every TB measurement. This correction is carried out via a linear relation depending on the measured TB itself and is also a function of frequency and elevation angle. The TBs with and without the offset correction are both applied to MWR measurements and the results are discussed in Section 6 below.

We concur with the reviewer that it is necessary to think of how to apply the TB offset correction at a site with no operational radiosonde measurement. Currently we are installing a procedure at JOYCE (Jülich Observatory for Cloud Evolution: www.geomet.unicoeln.de/joyce) which will use COSMO analyses to correct for the systematic TB offsets. We expect that approximately 10 – 15 clear sky cases should be sufficient to come up with a valid TB correction. We plan to update the TB correction for each detected clear sky case. Post processing between two subsequent radiosondes will then deliver the final quality controlled data set. We plan to publish results of this procedure once we have collected enough data and have come up with an optimal implementation.

*If the first 50 m of the radiosonde measurements are faulty because of lack of ventilation around the sensor, it might be better to bottom the 50 m of the radiosonde by nearby tower measurement, this might decrease the random error seen in the brightness inter comparison for the highest channel and reduce the rms in the retrieved profile if compared using this bottom profile.*

**R3c:** The retrieval algorithm has been developed for height levels 0 and 50 m. The highest-level temperature measurement of the tower is at 30 m as stated in Section 3.3. This height is located in middle of the two retrieval levels, so that either an extrapolation to the 50 m height of the tower measurement or an interpolation of the temperature retrieval to the 30 m height level would be necessary. Due to the strong temperature gradients and frequent
inversions, we think that the accuracy of such a procedure would not be better than the 0.53 K difference between sonde and tower mentioned in Section 3.3.

3) For assimilation purpose, when the retrieved profile are assimilated rather than the brightness temperature, it is very important to compute the error correlation of the retrieved profile. The authors should show the error correlation of the retrieve profile. They could also give a profile with a variable resolution which reflects the real resolution. It could be also useful to check if the noise at each altitude is Gaussian. This wide data set gives a good opportunity to evaluate this.

R3x: The error covariance of the retrieval is now included in the new Fig. 10. We have also added some corresponding text to this figure. Please refer to R1o. In order to consider the retrieval resolution when comparing to the radiosonde, we have included profiles in our analysis that have been smoothed with the averaging kernel. We elaborate on this in the last paragraph before Section 6.1 and in the second paragraph of Section 6.2. Please refer again to R1o. And yes indeed, as shown in the plot below, we can see that the derived error statistics resemble an idealized Gaussian distribution for two distinct evaluated heights (250 and 2000 m).

4) On the retrieval aspect, the authors make us think that the compensating bias in the day and night plots, are due to the use of radiosonde at 0 and 12h only. It is not fully clear to me, if this come from the fact that the data set used to compute the retrieval coefficient is not well distributed, or if the method will show no bias in average only on data set who shows the same distribution of profile that the training data set. Could the authors clarify this point. The addition of surface temperature indeed improve the first few hundred metre, but the compensating bias still exist aloft.
The two plots above show the day/night retrieval bias behavior using all radiosonde training data (N=12524) to derive the regression coefficients (as noted in the paper). The coefficients are then applied to simulated TBs of exactly these 12524 cases and evaluated as a function of day/night. The number of day and night cases are very similar. A similar bias behavior can be seen in Fig. 7 of the original manuscript when applying the retrieval to measurement. So we see that the day/night sub-samples results in the observed bias characteristic. However, as shown above, the bias profile is very dependent of the noise we add to the TBs when deriving the regression coefficients. An increase in TB noise of 0.3 K (from 0.2 K, left to 0.5 K, right) leads to an increase of the maximum bias of 0.2 K.

We also agree with the reviewer that the retrieval improvement effects of include the surface temperature are only small. This is why we have excluded this minor aspect from the abstract.

To clarify these two points we have also modified Section 6.4 to: When analyzing TB offset corrected temperature retrievals for 0 and 12 UTC radiosondes separately, the STDEV values are within 0.15 K of each other throughout the lowest 4 km. However, the BIAS shows a non-zero behavior as a function of height with opposite sign (Fig. 11). This behavior varies between -0.16 and +0.3 K at the surface and shows minima around +0 K at 250, respectively 1300 m. This effect is due to the fact that the derived retrieval coefficients have been derived without discriminating between day and night. Systematic temperature offsets very similar in size and vertical structure are obtained when applying the retrieval exactly to the simulated TBs of the 1992-2009 data set from which the regression coefficients are derived. However, when differentiating again by day and night, the magnitude and height of the minimum/maximum bias depends strongly on the random TB uncertainty chosen when deriving the linear regression coefficients. As only radiosonde ascents at 0 and 12 UTC are available, BIAS errors will always be an issue, even if one would derive separate retrieval coefficients for day and night. A way to partially reduce these effects is to include 2m-temperature observations (Fig. 11). Instead of using only calculated TBs as predictors for the temperature profile in the multi-linear regression, the 2m-temperature is additionally included and the BIAS values are much lower at the surface. The positive amplitude of the
BIAS values with height is also slightly reduced to 0.2 K throughout the lowest 4 km.

5) On the significant weather event, the error is lower than for the general inter comparison. The authors mention that this might be due to the lack of elevated inversion in the profile for which radiometers have indeed low skill. Could the authors show some examples of profile (retrieved profile and associated radiosonde) and how do they compare with the average climatology shifted using the surface temperature. Are the conditions really extreme?

R3x: We have extended the extreme case analysis to include inversion cases, cold and warm extremes (see also R1g and R2x). Instead of showing one typical example as proposed by the reviewer, Fig. 12 now shows i.e. the frontal cases in comparison to the other extremes in a statistical sense. Here it becomes directly clear, that the inversions are a major limiting factor to the MWR retrieval performance and that the frontal passage only contain a very small number (< 5 %) of these cases.

6) The data availability is quite high, the cross check by eye is not an option for operational radiometer as mentioned by the authors. The authors quote a 12 % rejection of data, it would be nice to know how much of this % come from the “by eye” check. The authors mention that the new generation of HATPRO will reject automatically this data, but this would have to be evaluated.

R3x: Our revised statistics give a total data availability of 2107 simultaneous radiosondes and MWR measurements of which 291 cases were excluded due to quality control. This corresponds to a 14 % rejection data rejection rate. 85 of the 291 rejected case were due to filtering “by eye”, whereas the other 206 cases very automatically rejected by the internal precipitation sensor detector. At the end of Section 4 we now note:

These quality controls lead to a reduction of the total available MWR data for the radiosonde comparison of about 14 %, whereby 4 % were rejected due to cross-checking “by eye”. The latter point still remains an open issue for all operational MWR applications: cross-checking “by eye” is certainly not an option for an operational MWR application, although necessary for the analysis of the current data set. This clearly demands for sophisticated automatic RFI filters, as well controls concerning the sanity of the receiver system.

Type
In the text it is mentioned STDEV, but in all the inter comparison plots we can see RMS, could you precise what are the number.

R3x: In all the figures, RMS has been changed to STDEV (see R1g, R1q, R2m)

Line 85, one and, should be removed
R3x: Sorry, we cannot identify the passage the reviewer is referring to.

Line 584, 592 TB org, should be TB orig
R3x: The terminologies TB_orig and TB_org are no longer used in the paper.