

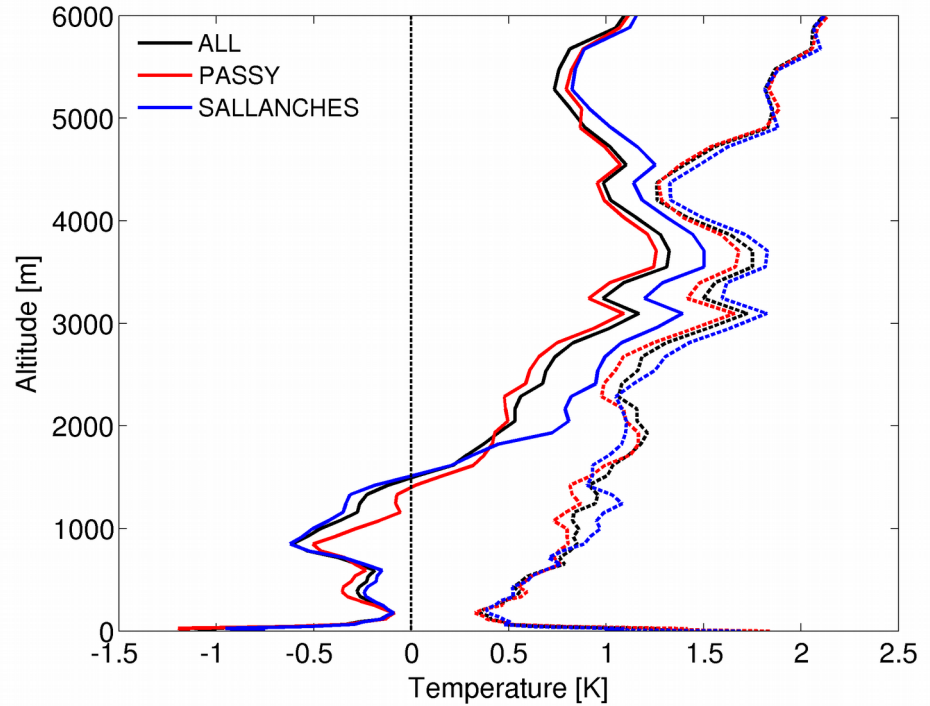
We would like to thank the reviewers for their time to review this manuscript and helpful suggestions to improve the manuscript. The reviewers questions are highlighted in bold, and the modifications to the manuscript in red. Figures 6 and 7 have been recomputed with a smaller number of elevation angles, and new colors have been chosen for figures 6, 7 and 9 to 13. To answer to reviewer 2 comments, original figures 9 to 15 have been modified. We hope that the improvement we brought to the figures will now fit the reviewer comments to make the manuscript suitable for publication.

Reviewer 2 comments :

1) In the introduction, the authors pose the question whether the surrounding mountains in the narrow valley affect the microwave observations. However, this is not really investigated in this study as only a comparison with radiosondes is made. Neither the atmospheric volume observed by the radiometer nor the flight track of the radiosonde is considered. One would need to calculate the effect of the antenna pattern or perform azimuth scans to see at which point the mountain slopes are in the field of view. In fact a simple calculation shows that even for transparent channels which receive radiation over the full extent of the atmosphere, mountains in 2 km wide valley should not provide a contribution to the main beam (2.5 deg FWHM). As side lobe suppression is -30 dB this is unlikely to contribute. More interesting is the question how strongly the true temperature field varies across the valley, e.g. is there any influence of mesoscale circulations or solar insolation? I do not expect the authors to perform an elaborated analysis in this respect but a more careful wording is necessary, e.g. "...thus can be safely deployed in complex terrain.."(p8, l5) or at p13,l1. It would be very interesting to know if the boundary layer scans performed in the two different directions as indicated in Figure 1 differ from each other?

We agree with the reviewer that the paper does not clearly investigate the effect of the mountain emission as azimuthal scans should have been performed. It is also true that mountain slopes should not contribute to the main beam at zenith. However, the mountain slopes were quite close in the Passy direction and it was interesting to investigate to what extent it impacts the MWR measurements. In fact, measurements at 54.94 GHz receives a small contribution of emission further than 6000 m and the first mountain slopes in the Passy direction are found at a distance of 4.5 km from the radiometer. As the sensitivity to the mountain slope is not detailed in the paper, we propose to remove « without being affected by surrounding mountain » in the introduction. The sentence « MWR can thus be safely deployed in complex terrain and similar temperature accuracy to that of flat and less complex terrain can be expected. » has been changed into : **MWR can thus be safely deployed in complex terrain and then similar temperature accuracy to that of flat and less complex terrain can be expected, at least if the line of sight of the MWR is free of obstacles over distances larger than about 5 km.**

Figure 1 : Vertical profiles of bias (solid line) and root-mean-square-error (dashed line) of temperature profiles retrieved by regression against radiosondes. Profiles retrieved using all measurements (black) or only measurements made in the Passy direction or only measurements made in the Sallanches direction



The reviewer is right that it is interesting to investigate the differences in boundary layer scans. It was the reason for this alternance of observations in the valley.

Figure 1 shows that most differences between temperature profiles retrieved in each direction are located above the boundary layer. The agreement with radiosonde is degraded between 2 and 5 km with measurements in the Sallanches direction. Figure 2 also investigates the differences in brightness temperatures between Passy and Sallanches. We can see that very few differences are found for opaque channels. On the contrary, the lower the elevation angle is, the larger the BT differences are for transparent channels. The maximum of differences is found at 5.4° with BT measurements colder in the Sallanches direction than Passy. Figure 3 shows a time serie of BT measurements at 51.26 GHz and 5.4° of elevation and we can observe a time delay in the diurnal cycle between Sallanches and Passy, Sallanches getting warmer before Passy.

The brightness temperatures are also warmer in the direction of Passy all day long that can be explained by the local orography differences between Passy and Sallanches (for example the altitude of the valley bottom is lower at Sallanches than Passy and the valley is narrower at Sallanches) and maybe by the impact of the urbanization in the direction of Passy. Another source of explanation could be the formation of cold pool in Sallanches.

If we look at the time series of BT measurements at 58 GHz (figure 3, bottom panel), we observe the same diurnal cycle and measurements in both directions. The city of Passy at 2.5 km is too far to affect the measurement of opaque channels and the mountain emission is totally absorbed by previous atmospheric layers. We observe thus the diurnal cycle of the atmospheric temperature in the valley at less than 2 km of the radiometer. Even though temperature heterogeneities probably exist in the valley, it is likely that they are not strong enough over such a short distance to be captured by the MWR. This will be investigated in future work.

The differences observed at 5.4° should be investigated more in details in a future work to understand and explain how temperature heterogeneity in the valley can be linked to the atmospheric circulation. This perspective has been described in the discussion:

Scanning in two different directions of the valley, MWR observations also offer the possibility of investigating temperature heterogeneity in the valley and how these differences are linked to the mesoscale circulation. This will be further investigated in a future study.

Figure 2 : Bias (solid lines) and standard deviations (dashed lines) of BT observations between Sallanches and Passy (BT Sallanches – BT Passy)

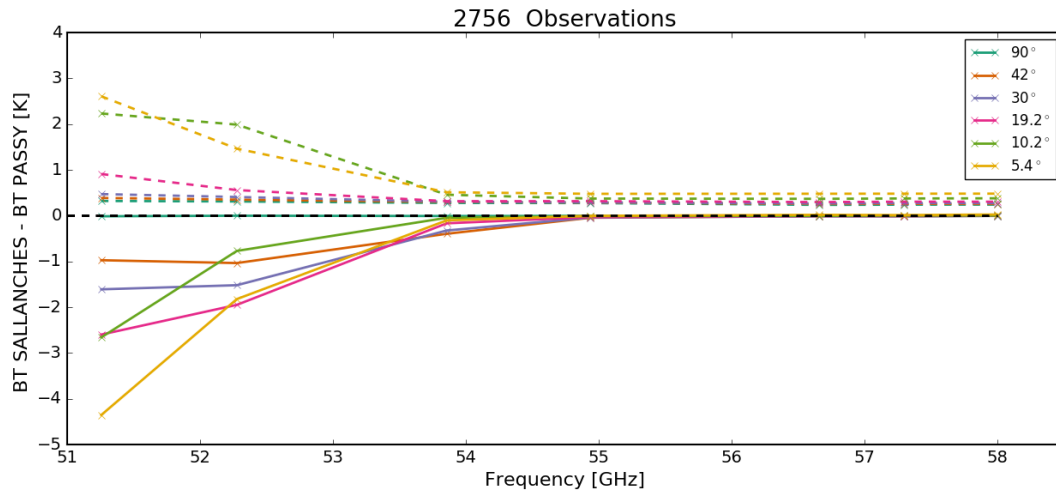
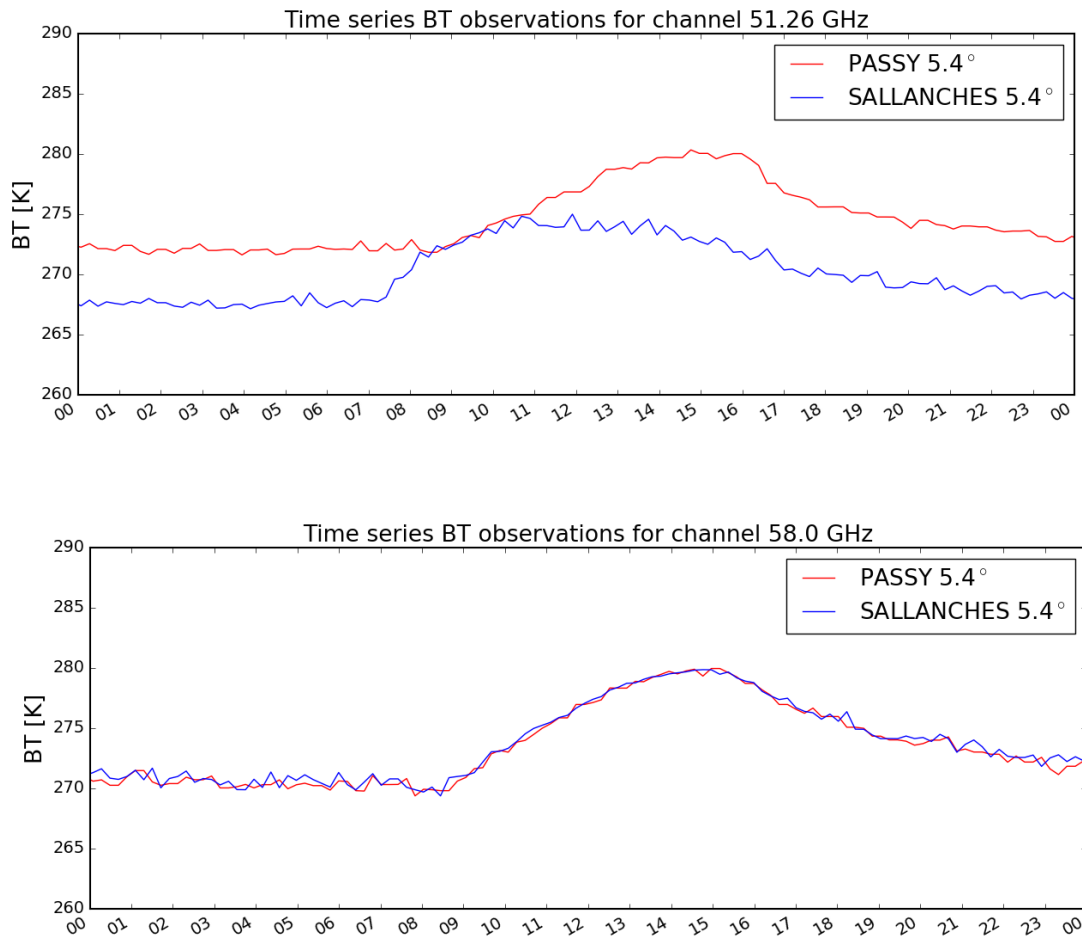


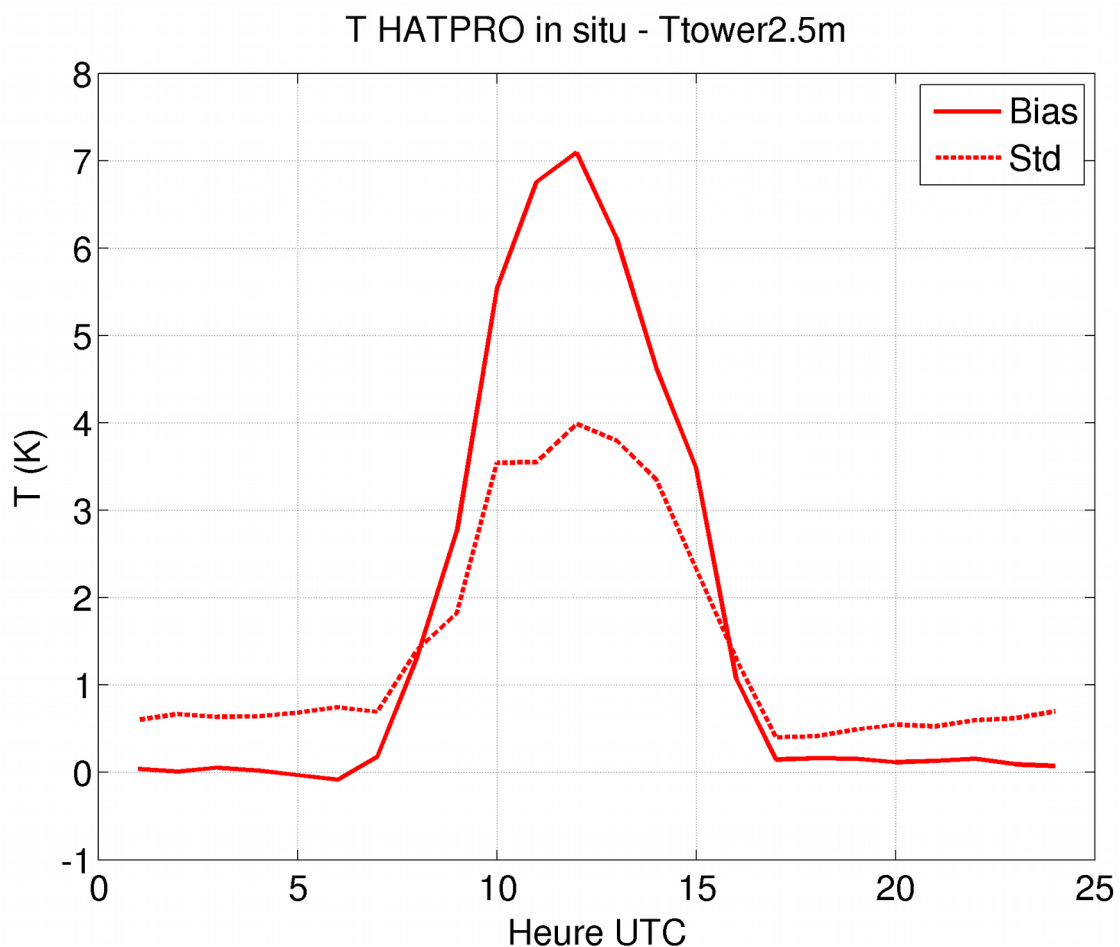
Figure 3 : Time series of BT observations at 51.26 GHz (top panel) and 58 GHz (bottom panel) in the Sallanches direction (blue) or Passy direction (red) at angle 5.4°



As far as I know the HATPRO standard regression boundary layer temperature retrieval makes use of the inbuilt in-situ temperature measurements. This might explain the good performance despite the lack of bias correction. However, the HATPRO sensor should not be as accurate (representative) as the weather station - did you intercompare them?

This is right that the inbuilt in-situ temperature measurements can normally be included in the regressions. However, the HATPRO sensor is known to suffer from large errors due to an inappropriate ventilation of the sensor. Figure 4 shows the bias and standard deviation of the HATPRO in-situ sensor minus the 2.5 m tower measurements. We observe a large bias during the day with a maximum of 7 K around 12 UTC. The use of this in-built sensor would thus degrade the retrievals. For that reason, the manufacturer regressions are generally provided without the inclusion of this in-situ sensor and it would need a lot of extra work to properly recompute the regression coefficients including the surface temperature which could be applied by replacing the HATPRO in-situ station by an accurate weather station. However, we made some testing during our previous campaign and the inclusion of the temperature sensor only improves the result in the first 200m where regressions are already very accurate in our experiment.

Figure 4 : Bias and standard deviation of HATPRO in-built station minus tower measurements at 2.5 m.



The authors use radiosonde measurements down to 10 m, however, the first roughly 100 m of the radiosonde ascent suffer from the fact that the sensors are not fully vented, Do the authors have some information on this from the tethered balloon measurements ?

The Vaisala probe used in the radiosonde has a very short time response of less than 0.5 s and is naturally ventilated by the balloon ascent at a 5m/s speed. We are thus quite confident in the accuracy of the radiosonde measurements. We do not expect the measurements to be degraded in the first 100 m but only in the first 10 to 20 m.

The Vaisala probe used under the tethered balloon has a similar time response but the tethered balloon at about 1 m/s so it is even less ventilated than the radiosonde probe.

However the temperature profiles from RS and tethered balloon (both from Vaisala probes) have been compared from the ground to 45 m (maximum height of the tethered balloon when it was on the same site than the RS from 6/02 3pm to 10/02 noon). Maximum differences observed are lower than the sensors accuracy (0.5 deg C).

A short discussion on the vertical resolution of the MWR should be included. Löhnert and Maier (2012) smooth the radiosonde profile with the averaging kernels for comparison to take these effects into account but here you are interested in the optimal retrieval. In fact, this discussion would support your outlook that the inclusion of infrared radiometers and lidar could improve the vertical resolution (p26, l27), cf Barrera et al., AMT, 2016. I do not agree that this would help only below clouds: IR and lidar give information below clouds and thus the information content from the MWR could be exploited for the higher levels.

The reviewer is right and a small discussion about the limited vertical resolution of the MWR has been included. Firstly in section 3.3 :**It is important to note here that the retrieval grid is finer than the true instrumental resolution but matches the AROME model vertical resolution.**

and section 6.2 :

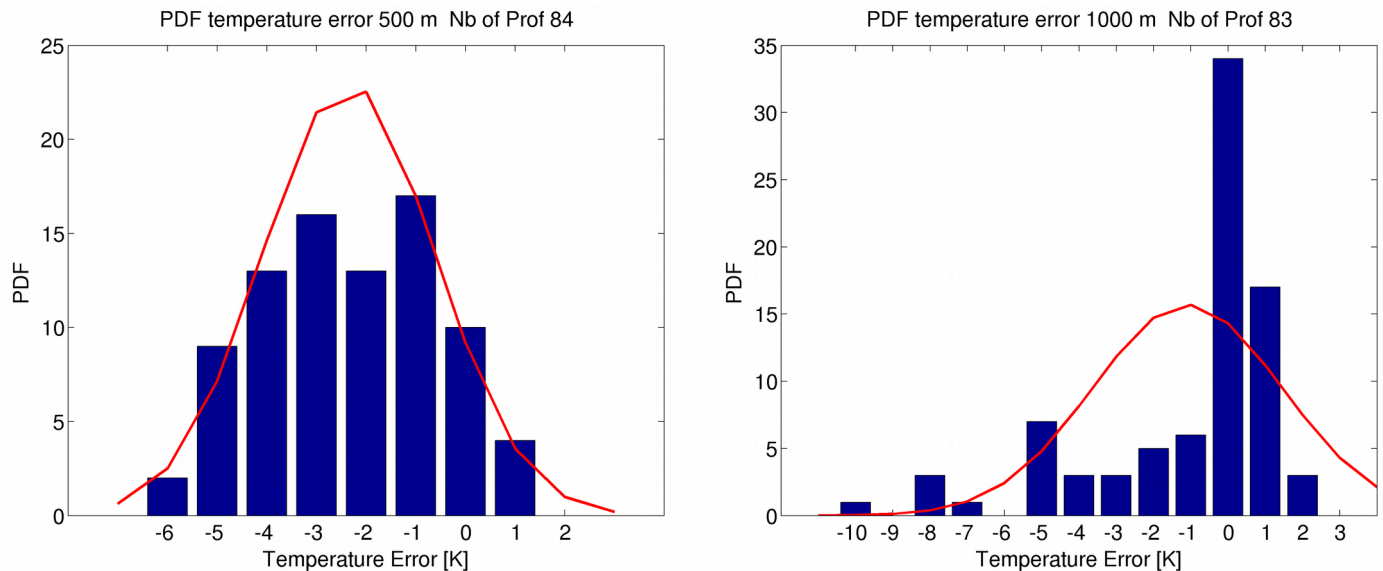
Here the radiosonde profiles are interpolated to the retrieval grid without taking into account the smoothing errors due to the limited vertical resolution of the MWR. In fact, this resolution is approximately between 50 m and 500 m and only 4 independent pieces of information can be extracted from the signal {Lohnert et al 2012}. On the contrary, the temperature profile is sampled approximately every 10 m by the radiosonde. In the future, the averaging kernel matrix could be used to bring the radiosonde profile onto the MWR vertical resolution.

Fig. 8 shows a bias of about -5 K with STD of about 3 K. Looking at Fig.3 the difference between Arome and RS is certainly not Gaussian distributed - what is the impact on the B matrix? This might be discussed in respect to the question what the optimal way to build the B-matrix is, e.g. dependence on flow and diurnal cycle?

Even though a bias is observed in the AROME forecast errors it does not necessarily imply that the forecast errors are not Gaussian. The only accurate way that is known nowadays to infer the forecast error distribution is to use an ensemble assimilation with an adequate number of members. It allows the evaluation of the Gaussianity by computing the distribution of the member differences with respect to the mean ensemble, all members being valid for the same forecast time. In our case, this ensemble assimilation is not available and we can only plot the frequency distribution of AROME minus RS differences mixing different days and forecast hours. This comparison is thus not-optimal for this estimation. However, we computed these differences for levels below 1000 m because the sample is too small above due to fewer radiosondes reaching high altitude (figure 6). We can see that the approximation of Gaussian distribution is respected for the lowest levels but a

negative skewness is observed at 1000 m. This is due to the small number of large errors at 1000 m at the beginning of the period corresponding to the cloud-based temperature inversions.

Figure 6 : Frequency distribution of AROME minus radiosonde differences at 500 m (left panel) and 1000 m (right panel) with corresponding Gaussian Distributions (red line).



Legrand et al 2016 recently evaluated this non-Gaussianity for the AROME model with a 90 member assimilation system. It was found that all control variables present some non-gaussianity but vorticity and divergence are more affected than temperature and humidity. This non-Gaussianity is reduced by the analysis process in areas constrained the most by observations. Even though non-gaussianity exists, current 3D-Var and 4D-Var system does not take into account this error. In addition, the B matrix should not be affected by the non-Gaussianity as it only affects the higher moments of the distribution (skewness and kurtosis) and not the ones used to compute the B matrix (mean and standard deviation).

The optimal way to build the B-matrix and to make it flow dependent and evolve with the diurnal cycle is again to use an ensemble assimilation to compute a new B matrix at each assimilation cycle (see Ménétrier et al 2014). This system has been extensively developed at Météo France in the last 5 years and should be operational by 2018. A B matrix flow-dependend and varying in time will thus be available in the future and the impact on the 1D-Var will also be possible.

This problem has been discussed in the manuscript :

Section 6.1 :

Non-Gaussianity can also affect forecast errors. Recently, Legrand et al 2016 evaluated the non-Gaussianity of analysis and forecast errors using a 90 member AROME ensemble assimilation. It was found that for all variables, non-Gaussianity exists but dynamical variables (vorticity and divergence) are more affected than temperature and humidity. The data assimilation reduces this non-Gaussianity at each cycle in regions well covered by observations. This ensemble assimilation does not exist for our period making complicated the evaluation of this Gaussianity in our context. However, it should affect higher moments of the error distribution than those used in the B matrix.

Section 6.2 :

The flow-dependency and diurnal cycle of forecast errors can be determined by implementing a real-time AROME ensemble assimilation system (Menetrier et al 2014). This is under-development and should be available next year.

Minor Comments

All comments have been taken into account. The changes are highlighted in red in the manuscript.

P10, I3 : A more

Not clear to us what the referee refers to here; thus, we take no action

p7, I19 : One sentence explaining O-B would be helpful.

O-B monitoring has been introduced at the end of section 3.3 :

Information about instrumental errors can be obtained by investigating differences between observations and simulations from background profiles (short-term forecasts or radiosondes). The monitoring of these differences called O-B (observations minus background) departures is essential to remove any systematic errors in the measurements, the forward operator or the background profiles (De Angelis et al 2017). They are investigated in section 5.

p12,I4: Did you look at the variability of Arome within the valley?

The variability of the AROME profiles has been studied and was found very homogeneous justifying the use of only the closest AROME grid point in the valley. Figure 7 shows a time serie of the temperature values extracted at different levels for all AROME grid points within the Valley. Figure 8 shows a comparison of the different temperature profiles within the valley for two different days.

Figure 7 : Time series of AROME temperature values at three different levels (300 m top panel, 800 m middle panel, 1500 m bottom panel) for all AROME grid points within the valley (grey dots) and the AROME grid point closest to the microwave radiometer (black dots). This is compared to the radiosonde (red) and the 2.5 m tower (orange) measurements.

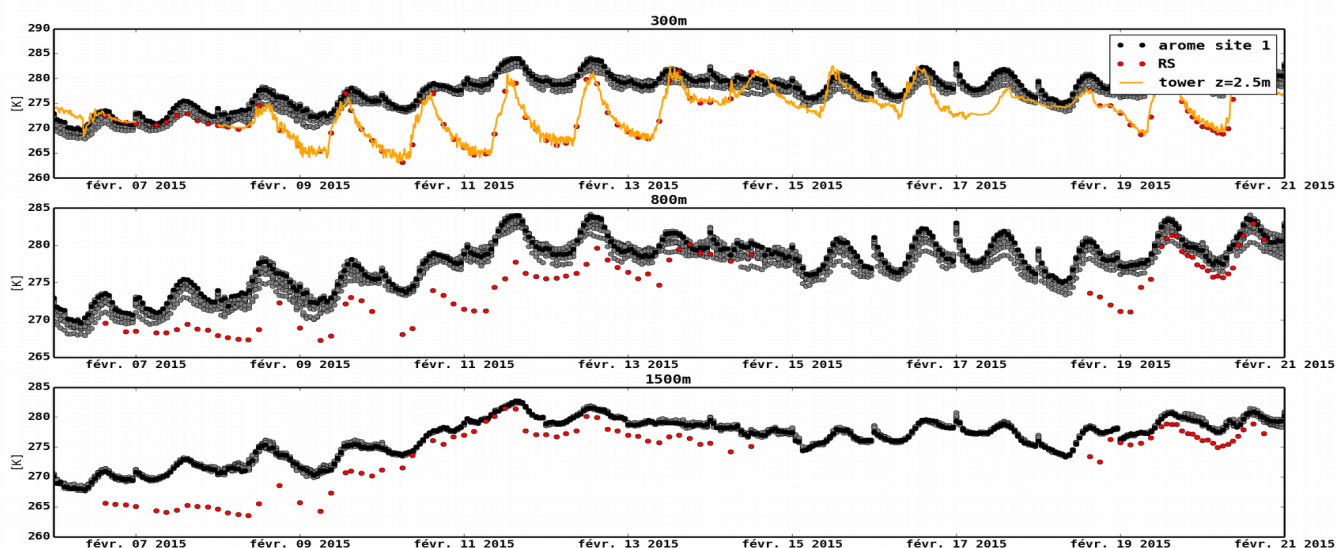
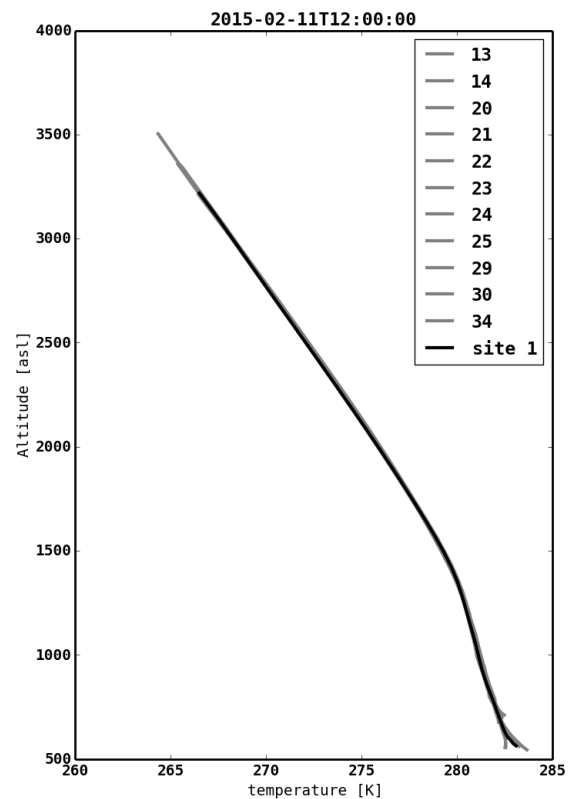
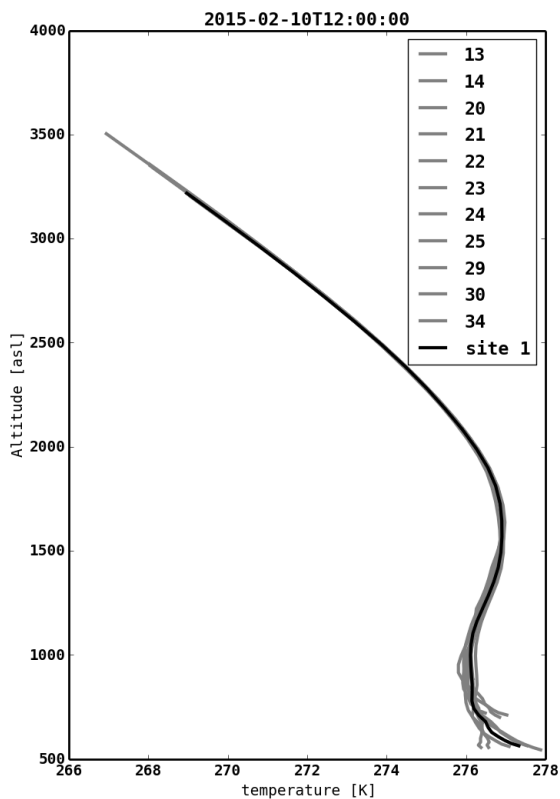


Figure 8 : Temperature profiles for all AROME grid points within the valley (grey lines) compared to the closest grid point to the microwave radiometer (black line) during two different days.



P15, 119 : How is this done in detail, should be reproducible

As the standard deviation of RS minus AROME differences are on top of the operational standard deviation above 2 km altitude, a new standard deviation vector is obtained by combining the operational standard deviation above 2 km and the one computed from AROME minus RS differences below 2 km altitude :

$\text{sigma_new} = [\text{sigma_oper}(1:61), \text{sigma_Passy}(62:90)]$

Covariances are then computed with the usual formulation :

$\text{cov}(i,j) = \text{cor}(i,j) * \text{sigma_new}(i) * \text{sigma_new}(j)$

with the operational correlations.

In order to make this more clear, we modified the sentence :

As the 1DVAR retrieval accuracy depends on how well the B matrix is defined, the diagonal terms of the B matrix (auto-covariance of the temperature errors) were simply replaced by the variance of the radiosonde minus AROME differences (i.e. the square of standard deviation values in Figure 8) below 2 km.

Tower measurement differences between 2.5 and 5 m :

The values were not mixed up but we are more confident with the accuracy of the 5 m measurements. They are based on a Socrima shield which is naturally ventilated whereas the 2.5 m sensor is a new development undertaken in our laboratory. It is composed of a PT100 probe and a thin wire in a prototype shield with a forced ventilation to retrieve temperature at a high temporal frequency. The experiment highlighted some problems that would request an improved shield

design. We suggest to modify the figure to remove the comparison with the measurements at 2.5 m and only compare with the measurements at 5 m. Figure 11 has been changed accordingly.

p18, l10 : **A bit provocative: When you use the RS as a priori and also evaluate with an RS one could argue that systematic RS errors (time lag, calibration..) might be similar and therefore Arome has no chance**

Of course using radiosonde measurements which show less background errors than AROME was expected to improve the retrievals especially at the cloud-based temperature inversions. However, the systematic RS errors should be negligible in this comparison. Here, we just want to illustrate how the 1DVAR can be applied in an experimental campaign adapting the background profile to obtain the best estimation of the true atmosphere and how to deal with elevated cloud-based inversion by using a more appropriate background. This has been highlighted in section 6.4 :

Our study shows that another way of improvement is to use an external information to infer the presence of an elevated temperature inversion that will be incorporated in the background of the 1DVAR algorithm.

P24, l7 : **I do not understand how the valley constrained the measurement configuration - do you mean the difficulty to find a site with a free view?**

By « constrained the measurement configuration » we mean a direction free of obstacles. We changed the sentence into :

Within the Passy-2015 field campaign, a HATPRO ground-based microwave radiometer was operated in a deep Alpine valley making complex the instrumental deployment due to surrounding mountains.