Integrated water vapor and liquid water path retrieval using a single-channel radiometer

amt-2020-211

Responses to Reviewer 3

Anne-Claire Billault-Roux and Alexis Berne

January 2021

We would like to thank Reviewer 3 for their additional comments, which were constructive and helpful for our analysis. In the present document, the comments of the referee are reported in italic, our responses in normal font and the corresponding modifications in the manuscript in blue.

General

I like how the authors have improved the manuscript – specifically concerning a more detailed discussion of the results. I do have some remaining further general and specific points, which I would like the authors to bring forward more clearly.

For my point of view, the authors don't make it clear enough, why 89 GHz TB impacts the LWP retrieval more than IWV compared to the non-radiometric observations. To that extent it is important to note that LWP and atmospheric water vapor (mainly IWV and to lesser extent the profile shape) are basically the only contributors to the TB signal at 89 GHz. Since you can't retrieve two largely independent parameters (IWV and LWP) from only one measurement (89 GHz TB), you need to consider additional information that constrains the retrieval. The additional information the authors have (correctly) chosen (e.g. near-surface temperature and humidity, geographical location and altitude) is mostly correlated to IWV and hardly to LWP. This is the actual reason why LWP is more impacted by the 89 GHz TB: the additional information constrains the IWV and so the 89 GHz information can be used to retrieve the LWP. This needs to be made more clear. The authors, in contrast, also argument (see abstract, conclusions and within the main text) with a lower sensitivity of 89 GHz TB to IWV compared to LWP. This is a bit like comparing apples with oranges. If the authors would like to use this argument, they need to explicitly state the sensitivities (in TB/kgm-2) but these numbers then need to be related to the absolute variability of IWV, respectively LWP which again are not directly comparable.

We thank the reviewer for this useful comment which brings more subtle insight into the problem. We reformulated some of our analysis to take this comment into account.

Abstract While 89-GHz brightness temperature is crucial to LWP retrieval, only moderately does it contribute to IWV estimation, which is more constrained by the additional input features.

Section 4.1 Adding further information allows to disentangle IWV and LWP, which could not be achieved from the sole measurement of 89-GHz T_B. In this study, several categories of variables were included in the input features. The first category consists of T_B and higher order polynomials (up to fourth degree) and is expected to have the greatest importance in the retrieval of LWP, while the other categories would likely be more correlated to IWV.

Section 5.1 This highlights that while environmental features are well correlated to IWV, they are not sufficient to provide a reasonable estimate of LWP, for which microwave radiometer measurements are critical.

Section 6.2 The algorithm largely relies on non-radiometric features, and this is even more the case in cold and dry environments like that of ICE-POP, where IWV is low.

Error analysis: could the authors explain why they use R2 and not just R? R2 is often termed "explained variance". Also, I assume the RMSE is calculated in a way that includes the bias? Including information how RMSE is calculated would be great. I'm a bit puzzled about the "relative error" that is given for LWP and IWV. How was this calculated? E.g. the authors claim that the "relative error" is 7.2% for LWP values larger than 30 gm-2. But, if I look at Fig. 5, the RMSE is about 50 gm-2 at a target value of 100 gm-2 and about 100 gm-2 at a target value of 500 gm-2 which I would interpret as relative uncertainties of 50%, respectively 20%. Please clarify.

We truly thank the reviewer for this relevant comment that allowed us to identify a flaw in our calculations of the relative error, which, indeed, were not coherent with the other error metrics shown.

- R^2 was replaced with R, for the sake of clarity, as suggested by the reviewer. This variable is simply used to illustrate the quality of the correlation between predicted and target values, which is not necessarily well captured by RMSE.

  Please see changes in Fig. 5, 6, 9, 10, 11, 12, 13.

- The equations for the error metrics are stated below (same for IWV). $\mathrm{LWP_{retrieved}}$ and $\mathrm{LWP_{target}}$ are length-N vectors with the values of predicted (i.e. algorithm-retrieved) and target LWP values, respectively. With to this definition, RMSE indeed include bias. Note that for the calculation of Relative error, $\mathrm{LWP_{target,k}} = 0\ \mathrm{g/m^2}$ are excluded from the dataset, and that all LWP values are by definition positive. This information was included as a table in Appendix.

$$\mathrm{RMSE} = [\frac{1}{\mathrm{N}}\sum_{\mathrm{k=1}}^{\mathrm{N}}(\mathrm{LWP_{retrieved,k}} - \mathrm{LWP_{target,k}})^2]^{\frac{1}{2}}$$

$$\mathrm{RelErr} = \frac{1}{\mathrm{N}}\sum_{\mathrm{k=1}}^{\mathrm{N}}\frac{|\mathrm{LWP_{retrieved,k}} - \mathrm{LWP_{target,k}}|}{\mathrm{LWP_{target,k}}}$$

$$\mathrm{Bias} = \frac{1}{\mathrm{N}}\sum_{\mathrm{k=1}}^{\mathrm{N}}(\mathrm{LWP_{retrieved,k}} - \mathrm{LWP_{target,k}})$$

  For reference, the definitions of the error metrics that are used in this section and further on are recalled in Appendix A1.

- As indicated before, there was a significant mistake in the values of relative error presented in the revised version of the manuscript. The values did not match the rest of the error metrics shown. This was corrected for. The correct values of the relative error on the test dataset, as calculated with the formula indicated above, are 29% and 18% when excluding LWP=0 and LWP>30 g m^{-2}, respectively. For IWV, the relative error is of 6.5 %.

  Abstract The algorithm is shown to be quite robust although its accuracy is inevitably lower than that obtained with state-of-the-art multi-channel radiometers, with a relative error of 18 % for LWP (on cloudy cases with LWP > 30 g m^{-2}) and 6.5 % on IWV.

  Section 5.1.1 The IWV retrieval algorithm yields a RMSE of 1.6 kg m^{-2} on the testing set, which corresponds to a relative error of 6.5 %.

  Section 5.1.2 The LWP retrieval algorithm has a RMSE of 86 g m^{-2} at best on the testing set (training set: 84 g m^{-2} and validation set: 86 g m^{-2}). This corresponds to a relative error of 29 % on the testing set. Let us underline that the subsampling which is performed on the dataset for the retrieval of LWP is applied to training, validation and testing sets: the results that are presented here are therefore computed on the testing set with a truncated distribution – i.e. after subsampling. Additionally, if clear-sky cases are removed using 30 g m^{-2} as a threshold value, following Loehnert and Crewell (2003), the relative error is 18 %.

And last but not least: the presented IWV retrieval results should be briefly discussed in the context of ground-based GPS receivers, which are a world-wide standard methodology for deriving IWV.

We included one sentence on this technology in the introductory section, and another one in the conclusion.

Introduction It should be noted that IWV retrievals with similar accuracy are obtained using GPS sensors, as first proposed by Bevis et al. (1994), but this widely used technique does however not allow for joint retrieval of LWP.

Conclusion If available through a separate sensor such as a GPS receiver, independent IWV measurements could be included in the algorithm which may lead to an enhanced precision of the LWP retrieval.

Specific points

2.1 Radiosonde data set: please state how high a radiosonde ascent had to be in order to be used for NN training

The information was added to this paragraph.

The vertical extent of the atmospheric profiles ranges from 1 to 50 km, with a 0.25 quantile of 11km, meaning the profiles largely cover the lower troposphere.

3.1 Cloud liquid model: I see the ERA5 LWP comparison critical because, if I understand correctly, you assume the reanalysis LWP to be bias free. Could you comment on this in the text?

This is a true limitation of this criterion, which should indeed be highlighted.

Inevitably, when using this criterion for the choice of the cloud liquid model, it is assumed that reanalysis values of LWP are themselves bias-free, which could be questioned, especially in extreme environments (e.g. Lenaerts et al. 2017).

3.2 Radiative Transfer model: For completeness, please state which liquid water absorption model you have used. Also, I find the third paragraph difficult to follow and too lengthy. How can a LWC value correspond to a LWP value? This is dependent on cloud vertical extent. Can't you just justify a threshold in LWP above which with X% probability you'd expect significant precipitation leading to an additional drop size and DSD dependency of your TB?

The information on the liquid water absorption model was included.

Liquid water absorption is modeled according to Ellison et al. (2007).

To our knowledge, there is no clear-cut relation between high LWP values and the occurrence of precipitation; indeed, an atmospheric profile featuring a cloud with a large vertical extent might have a high LWP, even though there is no precipitation. Besides, we have no direct way to identify precipitating cases in the radiosonde dataset. The method we proposed here was to use the cloud classification proposed by Karstens et al. (1994), to identify in our dataset the atmospheric profiles which might correspond to precipitating cases, or at least to clouds which diverge from Rayleigh DSDs. We tried to reformulate the paragraph to make this more clear. An LWC value as such does not correspond to an LWP value; the proposed method flags atmospheric profiles in which the LWC exceeds a threshold, and then the LWP of those profiles is calculated.

There is no clear-cut relation between LWP values and the occurrence of precipitation, although the general trend is that higher LWP is related to more likely rain: as such, deviation from the Rayleigh regime is likely in high-LWP cases. In order to have a more rigorous grasp on when and how this drawback might affect the retrieval,criteria from Karstens et al. (1994) were used. In their study, the authors distinguished three types of liquid water clouds based on the value of LWC at a given altitude; for each category of cloud, a different characteristic radius is chosen for the DSD. **Mie effects can start to become an issue in the second category of clouds (*cumulus congestus*), identified for LWC $>$ 0.2 $\mathrm{g\,m^{-2}}$; in our dataset, the atmospheric profiles where this LWC theshold is exceeded (in at least one range gate) have, on average, a total LWP $\geq$ 830 $\mathrm{g\,m^{-2}}$, and around 2% of the entire dataset fall into this category**. Taking the third category (*cumulonimbus*) with LWC $>$ 0.4 $\mathrm{g\,m^{-2}}$, this applies to 1% of the entire dataset and **the average LWP threshold increases to 1400** $\mathrm{g\,m^{-2}}$. Those values can serve as a benchmark to identify LWP values where Mie effects can typically contaminate the retrieval. However, edge cases can also exist where the total LWP is quite low, but a small layer of nearly-precipitating or drizzling cloud still contaminates the retrieval, without featuring extremely high total LWP.

5.1.2 LWP algorithm: the authors write: "Let us highlight that in the case of a linear regression one would expect the error to diverge when high-order polynomials are included. This is not the *case here because of the non-linear behavior of the neural network." I don't understand this argumentation. Doesn't the inclusion of higher order terms in a linear regression actually lead to a retrieval improvement if the original dependencies between observation and target variable are non-linear? And shouldn't the NN actually capture this non-linearity without having to include higher-order terms?*

- Including high-order polynomial terms in a linear regression is often beneficial, however it can lead to erroneous results outside of the training domain. For instance, if a linear algorithm were implemented on a case with high TB (higher than, say, the maximum TB considered in the training dataset), this could lead to a strong divergence. This would especially be an issue in the case of a miscalibrated instrument. The neural network somehow saturates this effect and prevents this excessive divergence.
- The NN could indeed capture the non-linearity without including polynomial terms in the input features (the results of the analysis show that including them brings some improvement, but not massive improvement). However, for it to fully capture the polynomial behavior without including those terms in the input features, it would require more NN parameters (hidden layers and/or neurons), which comes at the risk of overfitting. Since general knowledge of the "physics" of the problems indicates that the dependence of LWP on TB is approximately polynomial (e.g. Loehnert and Crewell 2003), it is reasonable to include those features in the input of the neural network.

To make this less confusing, we changed the word non-linear to saturating.

5.2 Instrument calibration: I'm missing a discussion why in Fig. 7 there is practically no dependency on TB-offset (even if TB-offset = 5 K, corresponding to a delta LWP of say 250 gm-2? Please check...) for the TB-only retrieval, however very well on the TB + additional information retrievals. I'm pretty sure this has to do with the combined IWV and LWP dependency which I already commented on in the beginning.

This is an interesting aspect of this Fig. 7, on which we propose a discussion:

It is noteworthy that for T_B-only retrievals, the addition of a T_B offset does not result in a large increase of the error: for IWV, the addition of a 5K offset increases the RMSE from 5.6 to 6.2 kg m^{-2}; for LWP, the same offset leads to an increase from 139 to 142 g m^{-2}. This behavior is also observed when looking at how the bias, instead of the RMSE, increases with the addition of a T_B offset (not shown). In both cases, the error increases more drastically when multiple features are included, than when only T_B is used as input. One possible explanation for this effect is the following: when incorporating numerous input features, the algorithm is able to narrow down the range of possible IWV and LWP values in a given environmental context; in this constrained configuration, the correlation and sensitivity of the retrieval to T_B are then enhanced, leading to a stronger influence of a T_B offset.