Response to Reviewer 2

The authors would like to thank the reviewer for his/her careful review and constructive comments, which we believe will help improve the content of the manuscript. Below, under each bullet point, we provide a point by point response to each comment/question. The author responses are given in blue, and textual changes are italicised.

General Comments

- As I understand it, this study uses only simulated measurements (either with or without hydrometeors included), even for MWHS-2 for which actual measurements are available. This should be made clearer, especially given line 66, which states that no "background" data is required for the method. I'd also like to see a bit more discussion about whether the results you present should be expected to hold for real data. An assumption that you are making is that your forward model can represent real clouds with enough fidelity that a model trained on simulated data will still work on real-world data. Can you point to any evidence to back up this assumption? Have you tried comparing the histograms of model-simulated Tbs with real-world MWHS-2 Tbs?

  Reply: We agree with the reviewer, the line 66, can be confusing. We have re-written the sentence to:
  Pg 3, line 68: *"This is done for each channel separately and by only using the measurements, although the scheme is demonstrated in the study by using simulated observations."*

  Regarding the discussion about simulations holding well for the real-data, we agree with the reviewer that our assumption is a strong one and requires some evidence. For MWHS-2, we have added a new figure (Pg 6, Fig. 1, also shown below) comparing the histograms of background and bias corrected observations for one of the lower peaking channels of 183 GHz channel. The main deviations in the distributions arise from the hydrometeor scattering. With limited scope of particle size and shape variation in current NWP microphysical schemes, the true cloud variability in radiance space is likely underestimated, though this is but one factor among many when it comes to the challenge of modelling clouds. Also, ECMWF has real time monitoring of many satellite sensors including FY-3D MWHS-2, thus it is possible to assess how well the system (and RTTOV-SCATT) is performing.
  For ICI and SMS, though a qualitative assessment of the simulations is not possible, we expect underestimation of the cloud variability owing to only single PSD and habit assumption. However, it is important to consider that QRNN can easily adapt to changes in brightness temperatures introduced by PSD and habit variation.
  We have added this discussion in the revised manuscript in Sect 2.2.

- What is the computational cost of the QRNN method compared to simpler cloud- clearing filters? Could it feasibly be implemented into existing NWP models given current compu-
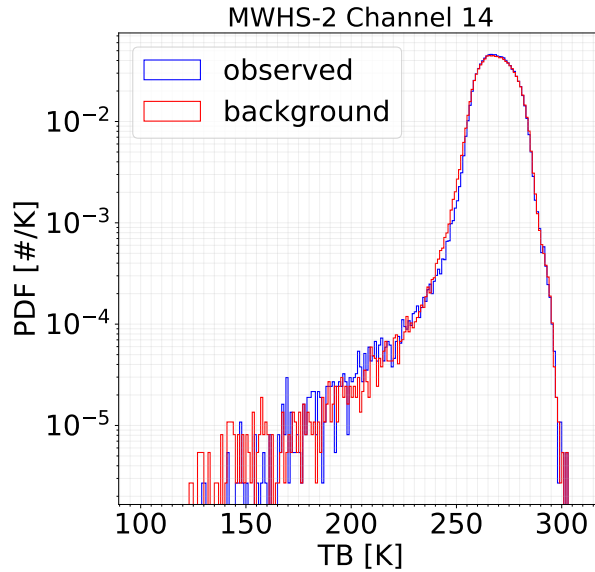
Figure 1: Probability distribution functions (PDFs) of simulated and observed brightness temperatures for MWHS-2 channel 14. The data covers latitude range $60°$ S to $60°$ N, and satellite zenith angle less than $7.5°$.

tational constraints?

Reply: Although the method is probably computationally more complex than other methods based on, for example, a scattering index, we consider it unlikely that the computational cost of the scheme would be an issue in an operational context.
The only demanding part of the scheme is the training process, but it is performed offline. Evaluating the network during operational processing would require only a forward pass through the network. In the study, we employ a simple, fully-connected network architecture so that the complexity of a forward pass is dominated by a low number of matrix-vector multiplications (one for each layer the network). The matrix-vector multiplications are typically combined into matrix-matrix multiplications to evaluate the network for multiple observations in parallel. Due to the recent popularity of neural networks, highly optimized implementations of these methods are available for all common computing architectures. We also include the following text to the Conclusions.
Pg 29, Line 591: *"Due to its low computational cost, implementation of this scheme should be feasible in NWP models given their computational constraints. Although the method is probably computationally more complex than existing cloud clearing methods, the demanding part of the scheme, the training, is performed offline. The operational processing only requires a forward pass through the neural network, for which highly-optimized implementations are readily available on all common computing platforms."*

- Lines 135-143: Why the large difference in number of cases simulated for ICI ($220\,000$) vs. SMS ($143\,000$), if they are both coming from the same population of CloudSat profiles? For the training dataset, you use about 75% of total cases for MWHS-2, about 80% of cases for ICI, and about 84% of cases for SMS. Why this difference in proportions?

Reply: Yes, both ICI and SMS simulations come from same population of Cloudsat. For

2

SMS, we had access to a smaller database, which was used for a earlier study. Additionally, SMS has fewer channels involved so it can be handled by smaller databases.
Due to difference in sizes of the available observations, maximum possible observations were incorporated in the training data to allow cloud variability. Even with differences in the database sizes, the results show that the predicted quantiles are well calibrated for all three sensors. With a larger database, one could expect similar or even better performance.

- Line 340: "With a test dataset of 70 000 samples, we cannot represent the far wings of the distribution accurately." Couldn't one apply this same reasoning to Figures 2,3, or 8, which have density values below 10^{-4} included?

  Reply: Yes one could apply the same reasoning to the Figures 2, 3, or 8, but we prefer to show the complete distribution to emphasize that the cases with large errors occur infrequently.

- Figure 13: I wonder if there might be an easier, more concise way to evaluate whether the uncertainty intervals are properly calibrated. Namely, could you simply calculate how often the true value falls within the $\pm 2\sigma$ uncertainty range, for each of the uncertainty bins $(0 - 3K, 3 - 8K, 8+ K)$ that you've included on the plot? If this percentage is significantly less than 95%, it would suggests that your uncertainties are too small, while if it were closer to 100% it would suggest your uncertainties are too large. Even if you don't get rid of Fig. 13, I think this would be useful information to include in the paper.

  Reply: As per suggestion of the reviewer, we have included the information highlighting how well calibrated the uncertainties are. We provide the occurence percentage of true value for each uncertainty bin and add this information in Fig. 13 (also shown below, as Fig.2). For each bin considered, the true value falls in the $\pm 2\sigma$ uncertainty range almost 94% of the time, indicating that the uncertainties are well calibrated. Only for channel I3V and the outermost bin, occcurence percentage is 88%.
  These results are in line with Fig.7 and Fig.11, which also indicate that the uncertainty intervals are well calibrated.

- Lines 498-499: ". . . other underlying uncertainties not considered here." Whether here or elsewhere in the paper, I think you should talk a bit more about what these other uncertainties are (radiative transfer model errors should certainly be discussed), and what effects they might have on your results.

  Reply: Among the errors which affect the radiative transfer calculations are underestimation of cloud variability due to limited particle size distribution (PSD) and habit variation. Other factors include neglected antenna pattern and the limitations associated with input data, both Cloudsat and ERAInterim. For example, the simulations could have tendency to be biased towards the Cloudsat geographical sampling. All these issues can contribute to mapping errors between 183 GHz and higher frequency channels. However, it is possible improve the cloud representation by incorporating PSD and habit variation, and QRNN based cloud correction can easily adapt to changes in brightness temperatures introduced by the local variability. This is not the case with all-sky data assimilation systems, which may assume only single PSD and habit combination.
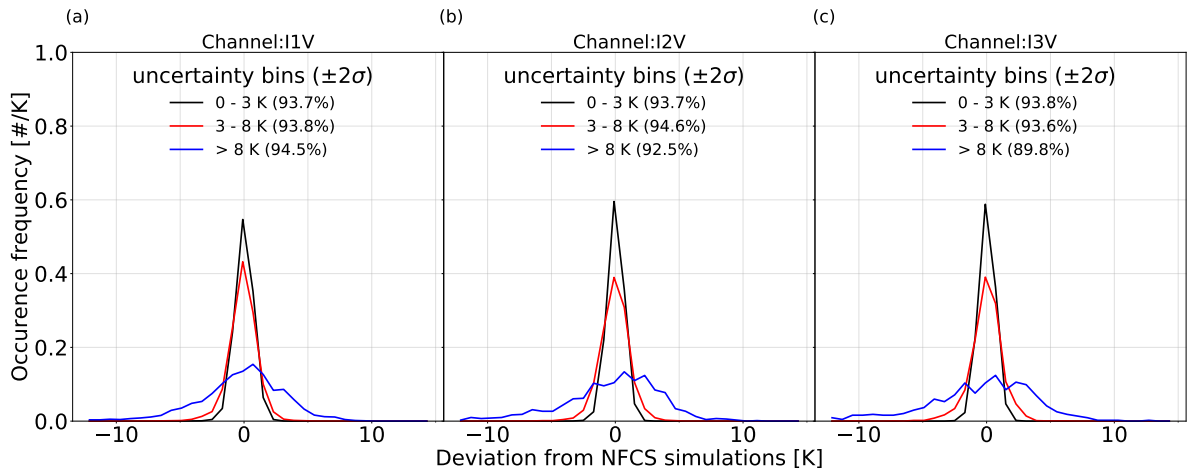
Figure 2: Distribution of errors binned according to their uncertainty. The percentage values in the parenthesis represent the occurence of true value within the $\pm 2\sigma$ uncertainty range, for each of the uncertainty bins. Results are from QRNN-single for channels I1V, I2V and I3V.

We have edited Sect 2.2 and Conclusions to describe these different errors and their implications.

Typos

- Line 26: The phrase "weather satellites are since some time equipped" is confusing to me. I suggest "weather satellites have for some time been equipped. . ."

  Reply: The sentence is re-written as per suggestion.

- Line 51: I believe you are missing the word "on" between "predicated" and "Gaussian."

  Reply: The typo is corrected.

- Line 154: Should 85% be changed to 84%? That would be +2 sigma for a normal distribution.

  Reply: Yes, the predicted percentile should be 84% instead of 85%. The typo is corrected.

- Line 455:". . . provide coverage to the humidity channels in, lower and mid troposphere" This is confusing - get rid of the comma and add the word "the" instead?

  Reply: The comma is removed and the word "the" is added.

- Line 494: "QRNN predictions are weighted mean. . ." I think this should say, "QRNN predictions are the weighted mean. . ."

  Reply: The word "the" is added.

- Line 508: Missing "the" before "same"

  Reply: The word "the" is added.