

Atmos. Chem. Phys. Discuss., referee comment RC1  
<https://doi.org/10.5194/acp-2021-617-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on acp-2021-617

Anonymous Referee #1

---

Referee comment on "Novel assessment of numerical forecasting model relative humidity with satellite probabilistic estimates" by Chloé Radice et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2021-617-RC1>, 2021

---

### GENERAL COMMENTS

=====

The paper describes a new method to compare satellite measurements against model data leveraging inter quartile ranges derived from probability density functions. The method is well introduced and explained and employed to compare SAPHIR measurements against ARPEGE model data. It reveals significant and insightful differences and discusses these in detail.

The topic fits the journal.

The major short coming of this paper is the reference method chosen to be compared against the newly developed method. As demonstrated by the paper itself, it is not useful at all. The paper itself mentions the common use of second moments for such comparisons but uses a blanket, unmotivated 15\% fixed error range itself.

I recommend a major revision of the paper with a more commonly used reference method e.g., one employing second moments (standard deviations).

### MAJOR COMMENTS

=====

This paper uses a very simple, so called "deterministic" method as reference. The method assumes a blanket +-15\% error range as acceptable, independent of the

actual level 2 data quality. This reference method is not properly motivated by the paper. Also, often dominant errors in radiative transfer inverse problems are of a multiplicative nature, which would affect high and low RH values very differently. This fact alone makes a constant error range an unrealistic assumption.

The analysis of the paper itself suggests that a smaller assumed range might be more suitable.

A very common method would be to use the standard deviation supplied by the data set (or at least compute it from the available distributions, if not given directly); such a method has obvious shortcomings, particularly for non-negative quantities, but is an "industry standard".

The paper must use a more reasonable reference method to compare against or demonstrate that a blanket assumption of  $\pm 15\%$  fixed offset error is a widely used method.

#### SPECIFIC COMMENTS =====

line 43  
-----

While it is true that the forward model introduces uncertainty into the comparison in measurement space, almost all inversion schemes make use of a forward model (at least for training a statistical model with obvious implications).

Due to the ill-posedness of the inversion, this implies that the uncertainty in geophysical space is almost always larger than the uncertainty in measurement space, particularly as the representation in geophysical space might contain a large "nullspace" inaccessible to the inversion (e.g. high frequency vertical oscillations in temperature to nadir sounders).

Thus large discrepancies in geophysical space might be very small in measurement space. This is one of the reasons, why assimilation prefers assimilating radiances in contrast to geophysical quantities (which are much easier to assimilate).

The current text reads as if comparing in measurement space would be disadvantageous, while a very strong case can be made for the opposite.

A large disadvantage in comparing in measurement space is that it is much more difficult to identify the reason for a disagreement in geophysical space

and thus the "faulty" model quantity.

line 103

-----

Please provide an introduction to "beta probability density functions". The references in the vicinity do not explain the term. A mathematical beta distribution has two free parameters, which seems in principle feasible to derive for six layers from six BT measurements including error estimates.

I do not believe that most readers are familiar with the term such that it deserves a better introduction, especially as it seem to lay the foundation for the latter IQR method.

Also, one would derive by multivariate regression, under Gaussian assumptions, a maximum likelihood vector and a covariance matrix detailing correlations in the data (optimal estimation). Typically the weighting functions of the sounder are not sharp enough to neglect correlations...?

Either way, please introduce the satellite level 2 product and its supplied diagnostics/error terms in sufficient detail.

line 131

-----

Is the averaged PDF retained, which can be a rather arbitrary function (discretized in some fashion, I assume), or is effectively only mean and sigma or the IQR computed? The example PDF look very Gaussian-like in all cases and suggest such an interpretation. If the actual shapes are different, maybe some PDFs in the visualisation should look more "wild".

line 163

-----

The PDF suggests that a value of +-15\% is too generous. Staying in a Gaussian framework this

looks like a 2-sigma value, whereas the CDF based method with the CDF of 0.5 would correspond to being in an interval of even less than  $\pm 1$ -sigma (being within one sigma has a probability of 68%).

The proposed method is sound, but the chosen example seems very biased. Even without using arbitrary PDF functions, a Gaussian approximation and error analysis should be able to provide better results than shown. Only if the PDF/CDF are non-Gaussian, an improvement will be achieved.

To that end, the authors should demonstrate the the difference to the (too) common Gaussian distribution assumption is significant.

line 215

-----

The interquartile range as central concept deserves at a one-sentence explanation in addition to the back-reference.

Using an IQR instead of the PDF loses a lot of information as it boils the arbitrary shape down to two simple numbers, comparable to the Gaussian approach with mean/sigma. I do not expect large differences unless strange, e.g., bimodal distributions, distribution appear.

What is here the experience of the authors?

line 219

-----

Again the 15% uncertainty come up. The authors make a compelling argument against Gaussian models, but picking a fixed 15% uncertainty is much worse than a simple Gaussian model-based uncertainty estimation would be. With a deterministic uncertainty of, say, 30%, the proposed method would compare even more favourably.

Please provide a reference to the chosen value of 15\% being a reasonable error estimate for the level 2 product. Looking into some of the given references, I couldn't find it.

Much better would be a comparison against a traditional Gaussian error analysis. The chosen confidence interval can be compared against being within some factor times sigma of the derived value.

line 303  
-----

Showing the individual distributions of RH\_mod and RH\_obs would be interesting as well. The Talagrand diagram suggests, as the authors note, that the model assume extreme values more often than the observation. The individual distributions should show the same in a maybe more accessible/familiar manner.

Figure 8a  
-----

The colour scale of this figure hides a lot of detail as can be seen by the fact that nearly everything is gray. Another indication that the +-15\% assumption is not good. I bet a non-linear colourscale blowing up the currently gray part would reveal a lot of interesting details.

line 426  
-----

Almost all level 2 satellite products from nadir or limb sounders offer a second moment (standard deviation) as diagnostics, many go beyond that (covariance matrices, error terms from different sources). The analysis suggests that a 15\% error assumption is not reasonable for the current data set. Using a proper second moment instead would certainly deliver more useful results.

The employed method uses the more useful IQR, which is likely superior to a more simple first/second moment consideration. This is, quite sadly, not demonstrated by the paper.

MINOR REMARKS  
=====

line 36

-----

two way -> two ways

line 92

-----

km2 -> km<sup>2</sup>

line 211

-----

gaussian -> Gaussian

line 394f

-----

$P > 0.25$  ->  $P < 0.25$

Figure 9a

-----

The colour scale should reflect the three regimes that have so far been used, ie.  $P < 0.25$ ,  $0.25 < P < 0.75$  and  $0.75 < P$ .