Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC3
https://doi.org/10.5194/amt-2022-121-RC3, 2022

Comment on amt-2022-121

Paul Poli (Referee)

Referee comment on "Estimation of refractivity uncertainties and vertical error correlations in collocated radio occultations, radiosondes, and model forecasts" by Johannes K. Nielsen et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-121-RC3, 2022

This article applies a three-way error analysis or three-cornered hat method to investigate vertical correlations of ‘errors’ in refractivity profiles, using 3 sources of data that are as different as possible, namely from satellite, in situ, and model-based gridded dataset (respectively: radio occultation (RO), radiosondes (RS), and the ERA5 reanalysis).

I believe the readers will enjoy this article which raises a number of interesting questions - and systematically attempts to bring answers to them. I did not find any flaw in the work or approach, and I only have some questions to help improve the paper (and feed the discussion).

Detailed comments:

- Given that the authors make an explicit attempt to tie the terminology to established documents like the GUM, other prior relevant publications may deserve to be cited, namely those that already considered the GUM and its applicability to Earth Observation data, e.g., from the FIDUCEO project: Merchant, C. , G. Holl, J. P. D. Mittaz, and E. R. Woolliams. 2019: Radiance Uncertainty Characterisation to Facilitate Climate Data Record Creation. Remote Sensing 11, no. 5: 474. doi:10.3390/rs11050474
- Section 1.2 mentions “RO reference coordinates” and “RO reference time”: how are they defined in the present work?
- Is it possible to indicate (or cite the appropriate reference) for the step where radiosondes and ERA5 data are projected into refractivity space?
- The approach to calculate epsilon C and epsilon X needs to be detailed, and preferably with dedicated equations for clarity. This would also remove the need for forward references to sections 4.2 and 4.3 in section 1.2.
- A map showing the locations (or a density map) of the 15,997 selected collocations may be a useful information for the readers.
- Is it possible to comment on the possibility that the two steps of cubic spline

interpolation, and of refractivity computations, both applied to ERA5 and radiosonde data, may (each) introduce correlations of uncertainties between these two datasets? Similarly, given that the assimilation of RO data improves the quality of a reanalysis, what are the prospects for some structural correlation between ERA5 forecast (even if, not analyses) and RO data?

- The numbers of vertical levels in each dataset may need to be introduced in the data section.
- As the work is using model forecast, whose quality decays as the integration time increases, may one expect a sensitivity of the results to the forecast integration time?
- Relying on correlations to pick-up a signal exposes one to be sensitive to any transient or structural correlation that may exist in the input data that are correlated, whatever the reason (true signal, artefact of pre-processing, matching bias, …). In the present case, the step of bias removal seems to be limited to a mean profile subtraction, carried out at the scale of each entire dataset (RO, RS, ERA5), is this correct? If so, this would leave, present in the data, all the bias(es) that may exist within each subset of analysis. Would it be possible to consider applying the bias removal in each subset (i.e. at the step when expectation values are computed, modifying slightly equation (7) to introduce the removal of the means), and then display (or report on) how much this changes the results (or not).
- Do the brackets in the right-hand side of in Equation (7) reflect the actual implementation? (i.e. averages are computed after adding all cross-products?)
- Can you clarify how the data subsets (the sub-spaces in which expectation values and hence correlations are computed) are defined? This seems to be, at least initially, based solely by considering the vertical dimension, but then later in the paper other dimensions (for computing the correlations and presenting the results) are introduced. This may be done with various sets of subscripts (for the various dimensions: vertical, latitude band, …). In the ideal case where one would have many events, one could consider to compute these error estimates with subsets defined spatially (e.g. 5 deg x 5 deg). The resulting geographic patterns that may be obtained could be of interest.
- Would it be possible to define early on, i.e., in the methodology section, the 'raw uncertainty estimates' mentioned in the results section? Also, the sigma symbol may deserve to be introduced numerically with an equation.
- In figure 7, does the "0 m" line refer to no filter? If so, such a filter has an infinitely small width (Dirac), but is probably non-zero.
- The results in figure 7 indicate that as one filters out small-scale variability in both other datasets, the dataset that appears to be most affected in its 'error' estimate is ERA5 (this one presents the largest spread, nearly 1%, between no-filter and 1800 m filter, in the lower troposphere). This would be consistent with that dataset containing the least small-scale vertical information, given that equation (7) suggests that for an error estimate to increase, there are two pathways: the cross-products of differences with respect to the two other datasets increase (first two terms in the sum), and/or the cross-product of the differences between the two other datasets decreases (third term, negative sign). The latter may be the mechanism by which removing small-scale information in RS and RO data (thus reducing the differences RO-RS) leads to ERA5 to appear of worse quality (when in fact its quality should be independent of that, but here the method uses the other data as references). Such hand-waving comments (for lack of a better expression) may be tried with a simple toy model. Similarly, the dataset whose 'error' estimate is the least affected by filtering the small-scale variability in the two other datasets seems to be the radiosondes, which is also consistent with that data source possibly containing the most small-scale vertical information in the lower troposphere (or is the figure 7(c), truncated at a maximum of 2.0%, showing something else?)
- Figure 9 shows seemingly slightly different results because in this case one considers smoothing on only one (other) dataset a time. However, one finds consistency. When (only) RS or (only) RO data are filtered (in (a) and (c), respectively), then either one of the two may start to resemble more to ERA5 (but less to the other, i.e. RO or RS,

respectively), so, in equation (7), the three terms that make up the total error estimate sees changes of different signs in its components (respectively, for the 3 terms in the right-hand-side of equation (7): decrease of differences ERA5-RS, no change ERA5-RO, and minus an increase of differences RS-RO – the net result is then a decrease of ERA5 estimated 'errors' when only RS is filtered). Similarly, this would explain that the 'error' estimate of RO increases in (b) when (only) RS data are filtered, making them resemble more ERA5 (respectively: increase of differences RO-RS, unchanged differences RO-ERA5, and minus a decrease of differences RS-ERA5). I note in passing that one missing piece of this puzzle would be to show what is happening to the error estimates of each dataset, when one filters that dataset only, and none of the other two datasets, as the results are not entirely predictable because they involve the sum of two terms moving in opposite directions, e.g., for RO error estimates, if filtering RO data: the differences RO-RS may increase, the differences RO-ERA5 may decrease, the differences ERA5-RS would be unchanged (so the net result is hard to predict -- but such a thought experiment may help shed light on the optimal 'footprint' to characterize each dataset).

- With respect to the expected long-range correlations expected in RS data (but not picked-up as well as expected by the method), this may also be due to the choice of sub-setting considered here, analyzing together day-time and night-time data. The effects of the radiative corrections (leading to consistently positive or consistently negative differences in each profile) may, if not cancel out, possibly be reduced, when considered together. However, redoing the exercise by separating clearly night- and day- ascents (and possibly leaving aside those profiles 'in between'), may show slightly different results. Such a separation for the RS data would somehow echo the efforts made to separate between rising and setting events for the RO part.
- Figures 11-12, I fail to see the labels (a) to (f) (either add these or amend the figure caption?).