20 May 2022

Review of:
"Estimation of refractivity uncertainties and vertical error correlations in collocated radio occultations, radiosondes and model forecasts**",** by Nielsen et al., submitted to AMT. AMT-2022-121. Version dated 9 May 2022.

General comments

1. This paper presents a generalisation of the three-cornered hat (3CH) method for assessing the error statistics of 3 independent collocated data sets. The generalisation (which is new) extends the method to error covariances of vector quantities (rather than error variances of scalar quantities). The method is applied to the following data: refractivity profiles retrieved from radio occultation (RO) data, refractivity profiles calculated from radiosonde profiles (GRUAN RS92 data), and refractivity profiles calculated from ERA5 (ECMWF reanalysis) short-range forecasts. The error covariance of each data set is derived using the 3CH method. The sensitivities of the method to collocation criteria and to vertical smoothing are assessed. The implications of the results for providing new and useful estimates of RO observation uncertainties are discussed.

2. This is a very interesting paper, and the results should be useful to the RO and NWP data assimilation communities. The results appear to represent a sound application of the method, and they are discussed thoughtfully.

3. I have some concerns over aspects of the discussion, particularly with respect to the concept of the "true" profile and the discussion of error correlations between data sets. However, addressing the issues in the text should not have any effect on the results.

4. I therefore recommend that this paper could become suitable for publication subject to minor revision to address the detailed comments and editorial points below.

Detailed comments

5. p.1-2, l.23-25: "This is likely because of the requirements — that the errors of the three data sets must be uncorrelated, and that the data sets must truly represent the same property with the same footprint in time and space — that are seldom met." This is not the main reason; it is that NWP DA theory requires that the errors of representation (all of them) are considered as part of the observation error, not part of the NWP forecast error, even though they arise because of the NWP system's limited ability to represent the real world. With the 3CH method it is not clear how the errors of representation will be portioned between the 3 data sets. (The Desroziers method does not have this problem.)

6. p.2, l.56-56: "The term vertical footprint of a data set is used here in the meaning width of an ideal physical refractivity feature, shaped as a delta function, mapped to the resolved representation of refractivity, for the given data set." This is not very clear, partly because a (Dirac) delta function has zero width. Please define more clearly what is meant to "vertical footprint", how it differs from "vertical resolution", and how it is related to the vertical smoothing discussed later.

7. p.3, l.62-63: "t is the actual refractivity at a vertical line above the RO reference coordinates at the RO reference time". 3CH method, i.e. the solution of simultaneous equations described later in the paper, makes no assumption about exactly what the

reference profile is. In fact, this is the paradox of the 3CH method, as discussed by O'Carroll et al. 2007 (appendix to their paper). The solution to the paradox is through the appreciation that non-zero correlations of error (assumed to be zero in the 3CH method) arise because each of the data sets represents different spatial scales. Alternatively, these can be considered as correlated errors of representation when the data are assumed to measure the same scales. There are also the error correlations caused by space/time collocation differences. In general, the paper discusses very well the issues of scales and error correlations, but it would be helpful to point out that they also are also related to the problem in defining the "true" profile.

8. p.3, l.66: "we assume that". This is a little confusing, because you say later that you actually processed the data to ensure that this was the case. (You did not just assume that it was true.)

9. p.3, l.68-69: "The ε^R component represents the distortion of the underlying truth in a data set, as it is being mapped to the observation grid." This is not clear. To which observations does it apply?

10. p.3, l.71-72: "representing the departure of the RO and RS92 trajectories in time and space from the vertical profile at the RO reference time." Again, this is identifying the reference profile with the "true" profile, and the problems discussed in point 7 above apply.

11. p.3, l.83: "intrinsic error". Is this the same as epsilon, as introduced on line 61?

12. p.3, l.88: "ε^R". This is the representation error for this definition of the true profile, but it would not be the appropriate representation error for use in NWP DA. See point 5 above.

13. p.4, l.105-107: "model forecast …". What is the range of this forecast, e.g. 3-, 6- or 9-hour forecasts interpolated to the RO observation time?
Also, "is model forecast" → "contains forecasts"?

14. p.4, l.125: "we assume no cross correlation components". Yes, this is what the algebra of the 3CH method assumes, but it is the weaknesses in this assumption that represent the problems with the method – see point 7 above.

15. p.4, l.127: "it can without loss of generality be assumed that all three data sets are bias free". Again, this is ambiguous – you did not just assume it – you actually ensured that it was true through the data processing (lines 151-152).

16. p.6, l.167-168: "However, if two data sets have similar vertical footprints, differing from the vertical footprint of t, these two data sets will have cross-correlated errors". Yes, they will have correlated errors, almost independently of how t is defined – see point 7 above.

17. p.7, l.180-191: "Uncertainty estimates … . In the three cornered hat analysis, the data set with the largest footprint determines the common footprint to be used for all three data sets.". I don't think this is true – see point 7 above. It is certainly not consistent with the assumption that the "true" profile (relatively to which all errors are assessed) is a vertical profile at the nominal RO location.

Also, the use of the ERA5 scale as a common scale to which all data are smoothed is certainly a good strategy for the reasons stated but, again, it is not consistent with the definition of the true profile.

18. p.8, l.205: "between 50 km and 300 km". This must reduce the sample size by a factor of 36. It raises the question of question the sample size is still big enough. From the results it appears to be so, but this may be worth a comment.
Also, please could you comment on the related problem of temporal collocation window.

19. p.8, l.213-214: "For each data set filtering has been performed, not on the data set itself, but on the two other complementing data sets (see figure legends)." This is not very clear, and the figure caption is not any clearer. I think I understand what has been done, but a few more words of explanation would be helpful, e.g. for the calculation for the curve shown for RO, only the RS92 and ERA5 data have been smoothed?

20. p.9, l.253-254: "It is worth noticing that the estimated vertical correlations of RS92 are larger for setting than for rising RO at high latitudes, especially between 6 and 22 km. So the G3CH fails to give an independent estimate of the RS92 correlations." This is a helpful warning; it is another illustration of a weakness in the 3CH and G3CH methods, through their implicit assumption that correlated errors (between data sets) are zero.

21. p.10, l.269-273: "In the derivation of G3CH representativeness is defined with reference to given scales in space and time of the truth. The truth is assumed to have smaller footprint 270 than any of the involved data sets. We choose for all data sets to report the estimated uncertainty boundaries with reference to the estimated footprint of the ERA5 data set. … This operation is equivalent to define the truth t with reference to the ERA5 footprint if one will." This is not consistent with 1.2 para 1. It again illustrates that the choice of t is somewhat arbitrary with this method, and that the error correlations will change according to the scale of the "truth".

22. p.10, l.287: "The increase of uncertainty in the troposphere is smaller at high latitude". Why is this? It may seem an obvious point, but you would expect larger % errors in the tropical lower troposphere because the absolute values of humidity are highest there and hence the collocation errors in refractivity will also be highest.

23. p.11, l.331: "model" → "NWP"?

Editorial comments

24. Throughout. "data sets", "datasets" or "data-sets". Consistency.

25. Throughout. "data" is usually used as plural, but in a few places as singular. Consistency.

26. p.4, l.98: "has" → "have". Also, l.103 and l.104.

27. p.4, l.109: "spans" → "span".

28. p.6, l.155: "is" → "are".

29. p.8, l.224: "happens" → "happen".

30. p.10, l.274: “includes” → “, include”?

31. p.10, l.289: “does” → “do”.

32. p.11, l.321: “promises” → “promise”.

33. p.11, l.325: “forecast” → “forecasts”.

34. p.22, Figure 7 caption: “easiest” → “most easily”.