The Three-Cornered Hat Method for Estimating Random Error Variances in Multiple Data Sets

Richard A. Anthes, Therese Rieckh and Jeremiah Sjoberg
COSMIC Program, University Corporation for Atmospheric Research, Boulder, CO

EUMETSAT ROM SAF-IROWG 2019; Helsingør (Elismore), Denmark; 19-25 September, 2019

Abstract

The Three-Cornered Hat (3CH) method, originally developed by physicists to estimate the errors of atomic clocks, has been shown by Anthes and Rieckh (2018) and Rieckh and Anthes (2018) to be a powerful tool for estimating vertical profiles of random error variances from multiple atmospheric data sets that are co-located in space and time. Unlike other methods of estimating errors that compare one data set such as radio occultation to other data sets (such as radiosondes or models), which also have errors, the 3CH method uses three or more data sets to estimate the actual random errors of all the data sets, not just the differences between data sets. We use the 3CH method to compute vertical profiles of estimated errors of COSMIC and COSMIC-2 radio occultation (RO) retrievals and other observational and model data sets.

Three-Cornered Hat (3CH) Method

- Suppose we have 3 data sets, X, Y and Z of a variable (e.g. T, q or N) all co-located to the same point in time and space.
- The error variance VAR_{err} of each data set can be estimated from the equations below, where MS is the mean square difference of the data sets, b_{ij} is the bias and $COV_{err}(i,j)$ is the covariance of the errors between data sets i and j.

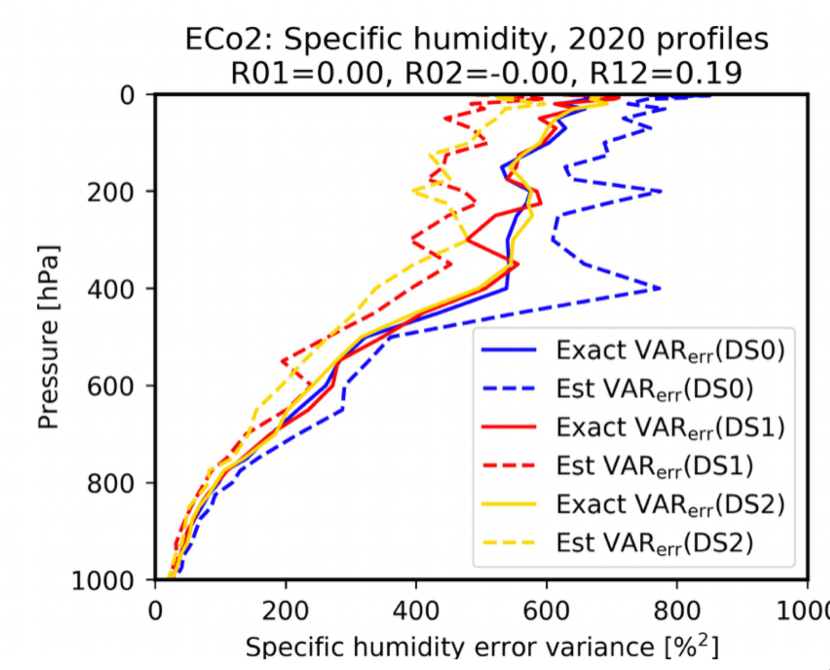
$$VAR_{err}(X) = ½\,[MS(X-Y) + MS(X-Z) - MS(Y-Z)] - ½\,[b_{xy}^2 + b_{xz}^2 - b_{yz}^2] + COV_{err}(X,Y) + COV_{err}(X,Z) - COV_{err}(Y,Z)$$

$$VAR_{err}(Y) = ½\,[MS(Y-Z) + MS(X-Y) - MS(X-Z)] - ½\,[b_{yz}^2 + b_{xy}^2 - b_{xz}^2] + COV_{err}(Y,Z) + COV_{err}(X,Y) - COV_{err}(X,Z)$$

$$VAR_{err}(Z) = ½\,[MS(X-Z) + MS(Y-Z) - MS(X-Y)] - ½\,[b_{xz}^2 + b_{yz}^2 - b_{xy}^2] + COV_{err}(X,Z) + COV_{err}(Y,Z) - COV_{err}(X,Y)$$

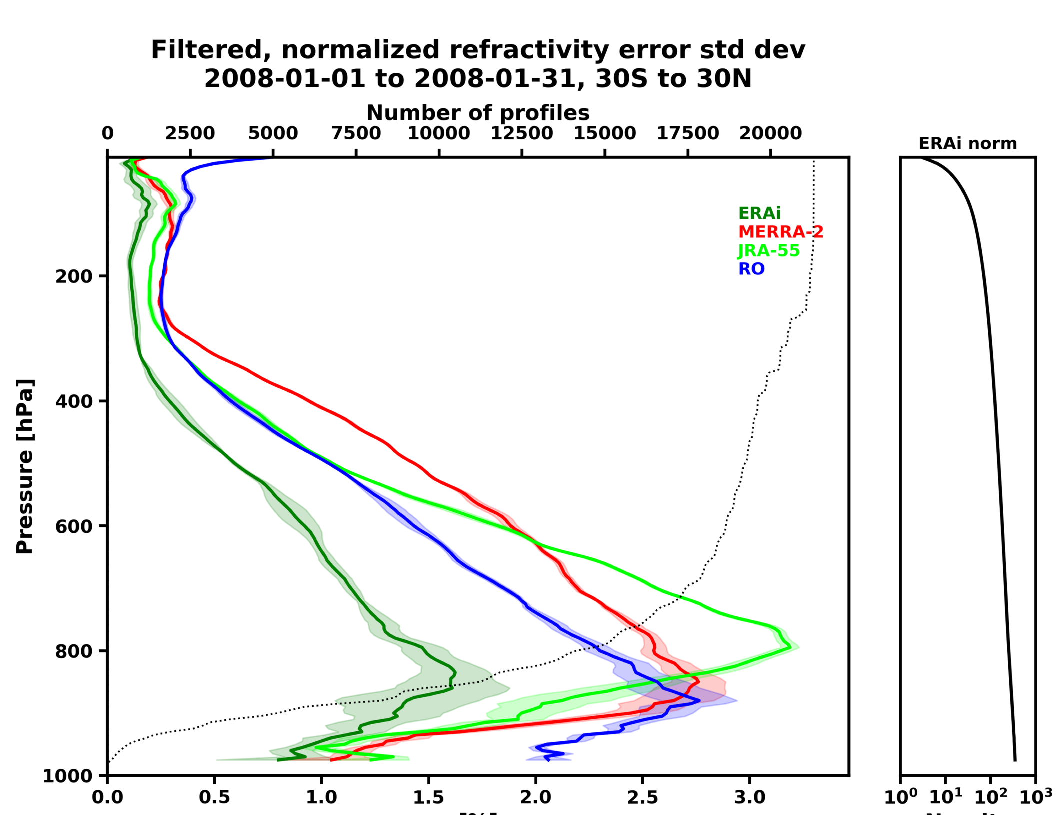
- These equations are exact. Neglect of the unknown COV_{err} terms allows VAR_{err} estimates to be computed from the data sets.
- Main sources of error are correlations of errors among two or more of the data sets and errors associated with co-location process.
- Effect of positive correlation of errors of X and Y result in underestimate of VAR_{err} in X and Y and overestimate in Z (Fig. 1)

Fig. 1: True (solid lines) and estimated VAR_{err} for three data sets. Errors for X and Y (orange and red lines) are correlated (r=0.19), which results in underestimates (dashed red and orange lines) for X and Y VAR_{err} and overestimates of Z VAR_{err} (blue dashed line). Results computed for simulated data sets with specified errors for which Truth is known.

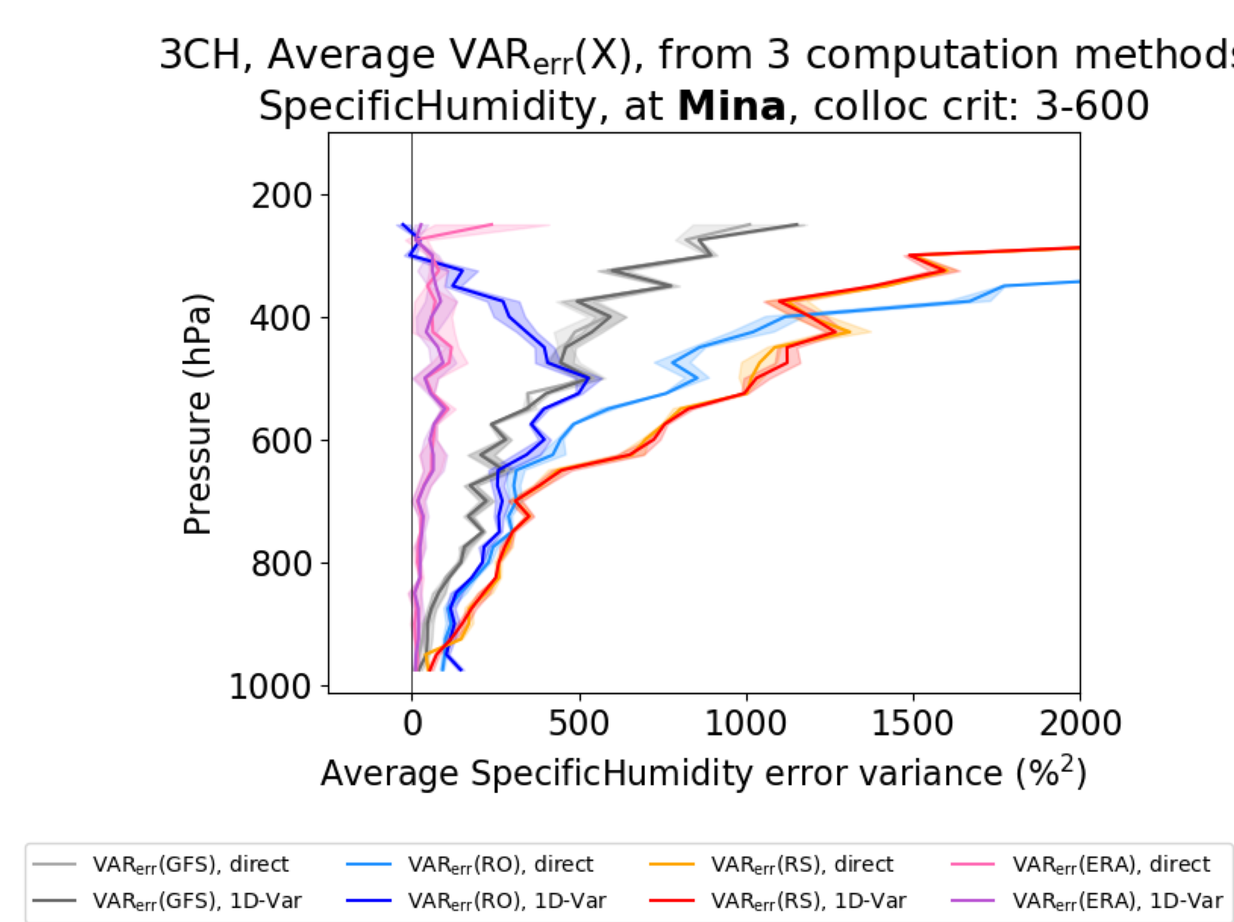


3CH error estimates COSMIC and Reanalysis Data Sets

- N data sets give (N-1)(N-2)/2 estimates of VAR_{err}
- Fig. 2 (below) shows vertical profiles of 3CH estimates of normalized refractivity error standard deviations of four data sets: (ERA-Interim, MERRA-2, JRA-55, and COSMIC RO) for Jan 2008 between 30°S and 30°N. Mean of the 3 estimates solid line, standard deviation about the mean (a measure of uncertainty) given by shading.



- Fig. 3: Estimated specific humidity (q) VAR_{err} for 5 data sets at Minamidaitojima, Japan for 2007: radiosondes, ERA-I, GFS, and two versions of RO q (direct, using GFS T, and 1D-Var (using ERA-I as background). Specific humidity normalized by 2007 mean ERA-I value. (Fig. 10a of Anthes and Rieckh, 2018, with biases removed)



3CH error estimates COSMIC-2, ECMWF and GFS forecast data sets

Fig. 4: COSMIC-2, ECMWF and GFS short-term forecast error estimates 8/8-8/14 2019. Refractivity normalized by ECMWF sample mean.

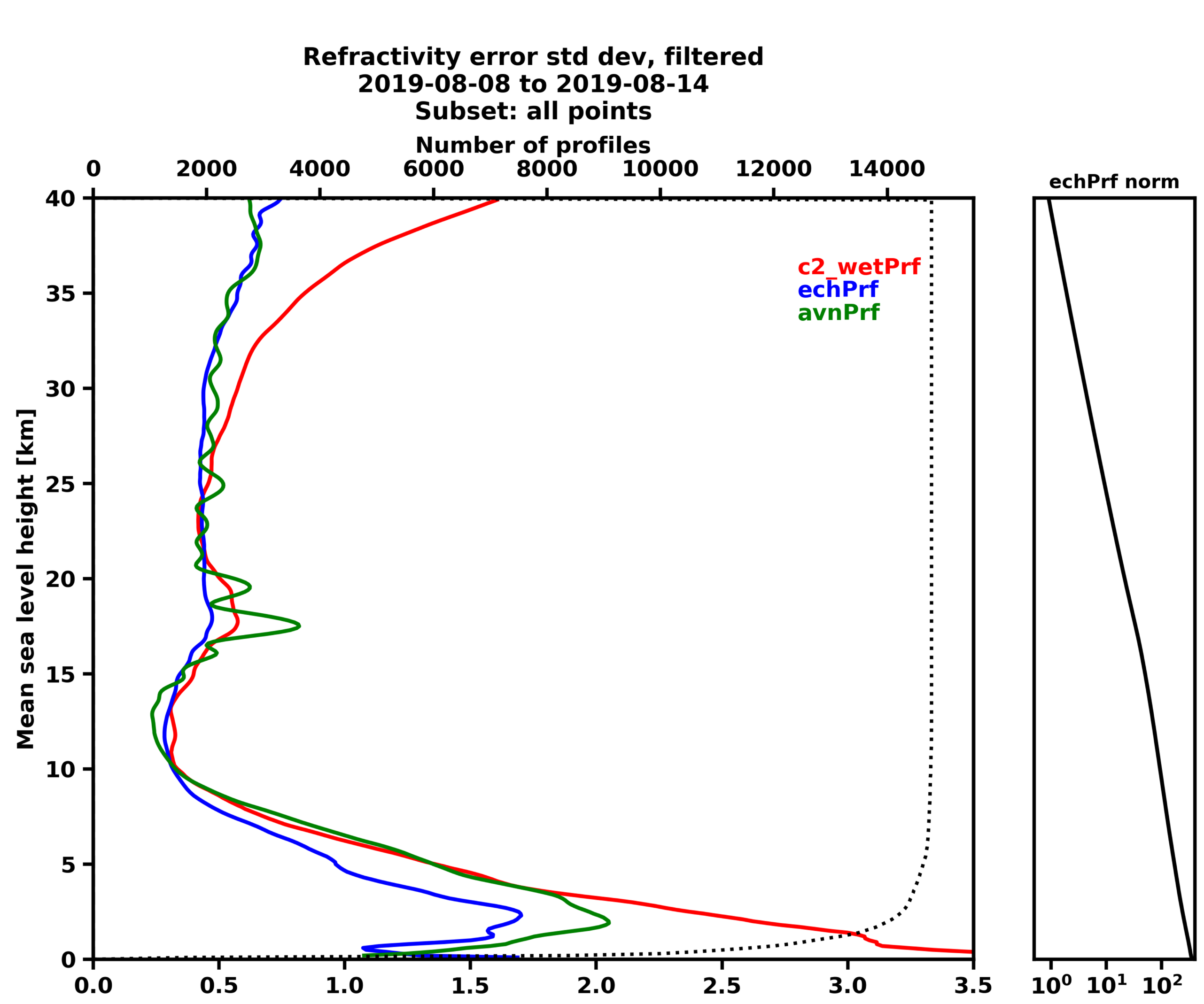


Fig. 5: 3CH and STD of differences COSMIC-1, FM1-FM6

3CH Relation to error variance estimates with two datasets

- Standard deviation (STD) of differences

$$\mathrm{Var}\,[X_n - Y_n] = \mathrm{Var}\,[\varepsilon_{X,n}] + \mathrm{Var}\,[\varepsilon_{Y,n}] - 2\mathrm{Cov}\,[\varepsilon_{X,n}, \varepsilon_{Y,n}]$$

- The STD of differences – assuming no error covariance – will always be greater than 3CH error standard deviation estimates
- With positive error covariances (most likely) the STD of differences will appear smaller than it really is.

References

Anthes, R.A. and T. Rieckh, 2018: Estimating observation and model error variances using multiple data sets. *Atmos. Meas. Tech.,* 11, 4239–4260.
Rieckh, T. and R.A. Anthes, 2018: Evaluating two methods of estimating error variances using simulated data sets with known errors. *Atmos. Meas. Tech.*, 11, 4309-4325