# Estimation of COSMIC-2 error variance using the three-cornered hat method with only radio occultation data

IROWG-8 Workshop, April 2021
Jeremiah Sjoberg (sjoberg@ucar.edu), Therese Rieckh, and Richard Anthes
COSMIC Program Office, University Corporation for Atmospheric Research, Boulder, CO

## I. Introduction

Precise estimation of error statistics is imperative in the evaluation of radio occultation (RO) data sets. The three-cornered hat (3CH) method provides a straightforward way to produce such estimates, specifically of error variance. A primary requirement for using this method is the existence of at least three data sets that contain co-located data. Gridded model data can easily satisfy this requirement, though the resultant estimates will contain the error variance contribution from representativeness differences between the data sets.

In this work, we remove the influence of representativeness by using the 3CH method with data sets consisting only of RO data (RO-RO-RO). Doing so may provide us with receiver-to-receiver and intermission estimates of the error variance. Given the small likelihood of triplets of co-located profiles in RO data sets, relatively dense observations are necessary. COSMIC-2 (C2) provides such density, and thus the primary focus here.

## II. Data

C2 atmPrf refractivity profiles
ERA5 and MERRA-2 refractivity profiles co-located to C2 following tangent point drift

Time span is 2019.274-2020.108 (2019/10/01-2020/04/17)
Only consider flight modules (FMs) 1, 2, and 3 here

## III. 3CH Method

We assume the following expansion for data set X (and similar for Y and Z):

$$X = T + b_X + \varepsilon_X$$

where $T$ is Truth, $b$ is mean bias, and $\varepsilon$ is random "error."

We can expand system of variance of differences and solve for each error variance, here for data set X:

$$\mathrm{Var}[\varepsilon_X] = \frac{1}{2}\left(\mathrm{Var}[X-Y] + \mathrm{Var}[X-Z] - \mathrm{Var}[Y-Z]\right) + \mathrm{Cov}[\varepsilon_X, \varepsilon_Y] + \mathrm{Cov}[\varepsilon_X, \varepsilon_Z] - \mathrm{Cov}[\varepsilon_Y, \varepsilon_Z]$$

In practice, we must assume the sum of covariance terms is zero. For more on the implications of this assumption, please see talk by Therese Rieckh: "COSMIC-2 random error variances using the three-cornered hat method," Thurs. at 10:30EDT.

## IV. Analysis

We use the 3CH method to estimate error standard deviations for a number of different setups. **1)** All data for FMs 1, 2, 3, or the union of all three are analyzed with co-located ERA5 and MERRA-2. **2)** The same analysis, but only for the subset of profiles for which there are FMs 1, 2, and 3 profiles within 6 hours, 600 km of each other. **3)** The same 6 hour, 600 km subset but with profiles from FMs 1, 2, and 3 as the three data sets. I.e., only RO data in the 3CH method. **4)** The same RO-RO-RO analysis but with ERA5 double differencing to remove temporal and spatial differences.

### Analysis 1

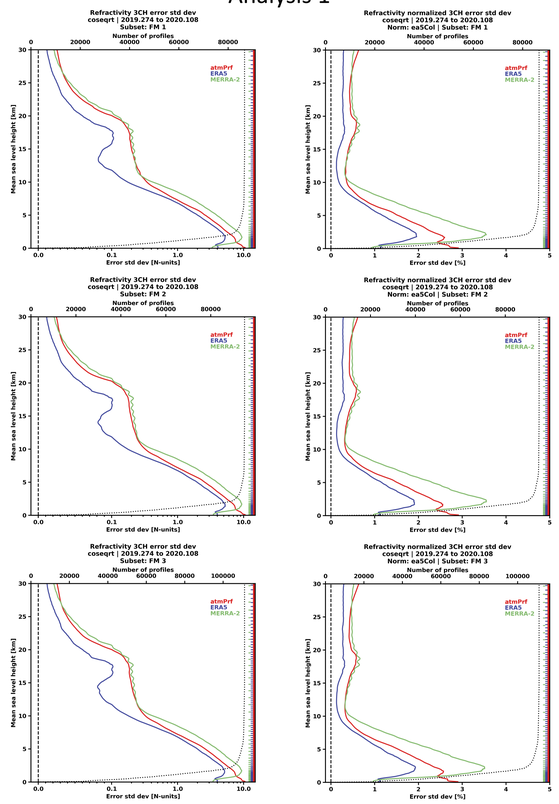


Fig. 1 – 3CH error standard deviation estimates for all profiles from FM 1 (top), FM 2 (middle), and FM 3 (bottom). Estimates are shown for C2 (red), ERA5 (blue), and MERRA-2 (green), and profile counts are shown by the dotted curve. Results are given in physical units (left) and normalized units (right). The estimates between the data sets agree with the analysis of Rieckh et al. (2021), and the estimates between FMs are very similar, as we would expect. I.e., there is very little difference in error statistics between different C2 FMs.

### Analysis 2

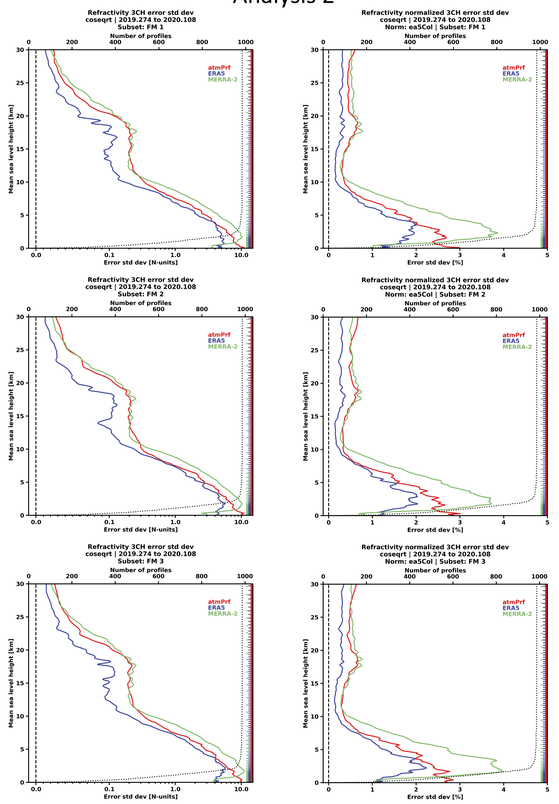


Fig. 2 – As in Fig. 1, but only for profiles from the subset where there are profiles from FMs 1, 2, and 3 within 6 hours, 600 km of each other. The sample size is two orders of magnitude smaller, but these estimates are consistent with those from Fig. 1. I.e., under very similar atmospheric sampling, the method still shows very little difference in error statistics between different C2 FMs.

### Analysis 3

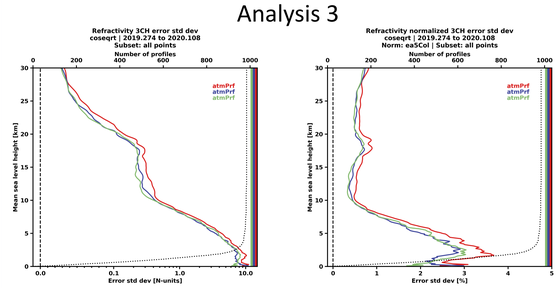


Fig. 3 – 3CH error standard deviation estimates from the subset where there are profiles from FMs 1, 2, and 3 within 6 hours, 600 km of each other. Estimates are shown for FM 1 (red), FM 2 (blue), and FM 3 (green). Results are given in physical units (left) and normalized units (right). Despite the spatial and temporal differences, the estimates for each FM are consistent with but larger than those using reanalyses co-located to the C2 data (cf. Figs. 1 and 2).

### Analysis 4

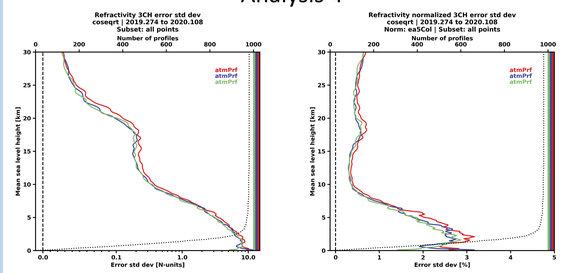


Fig. 4 – As in Fig. 3, but with double differencing by ERA5. The double differencing reduces the spatio-temporal differences between samples, but adds in representativeness differences. Despite this, the estimates in this figure are smaller than those where there is no double differencing applied (cf. Fig. 3). The estimates are larger than those from the estimates produced using RO with co-located reanalyses (cf. Figs. 1 and 2).

### Synthesis

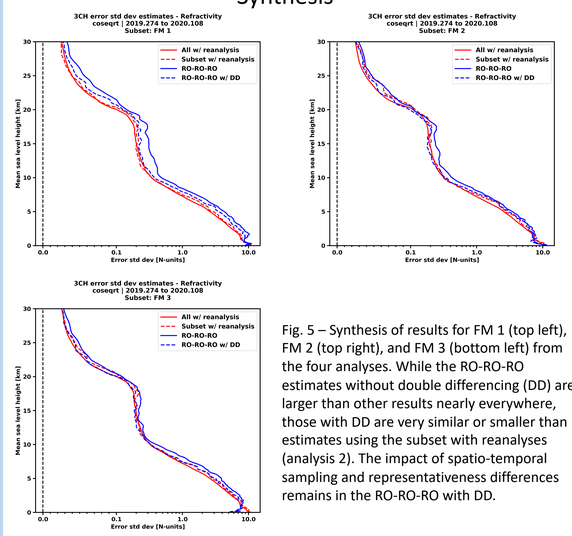


Fig. 5 – Synthesis of results for FM 1 (top left), FM 2 (top right), and FM 3 (bottom left) from the four analyses. While the RO-RO-RO estimates without double differencing (DD) are larger than other results nearly everywhere, those with DD are very similar or smaller than estimates using the subset with reanalyses (analysis 2). The impact of spatio-temporal sampling and representativeness differences remains in the RO-RO-RO with DD.

## V. Summary

All of our analyses show considerable consistency between the results, supporting using RO-RO-RO with the 3CH. Double differencing reduces the impact of spatio-temporal differences between roughly co-located RO profiles, but builds in representativeness differences that we seek to remove by moving to the RO-RO-RO framework. Additional studies will look at how to further, simultaneously minimize these two sources of differences so as to get better estimates of the intrinsic error statistics of C2 and other RO missions.