Estimates of Errors in Radio Occultation and multiple (models and) Reanalyses

Jeremiah Sjoberg, Richard Anthes, and Therese Rieckh

UCAR/COSMIC

2019/09/20

Overview

- Want to describe error statistics of 1 or more datasets
  - Actual error values would be nice, but impossible to measure
  - Error variance, alongside mean bias, is a useful statistic
  - E.g., data assimilation systems weight observations using error variance
- In this presentation, we will show
  1) A method for simultaneously estimating error variance for 3 or more datasets
  2) Application of this method to RO and models

Theory

- Assume we have three co-located datasets that can be cast as

$$X_n = T_n + b_X + \varepsilon_{X,n}$$
$$Y_n = T_n + b_Y + \varepsilon_{Y,n}$$
$$Z_n = T_n + b_Z + \varepsilon_{Z,n}$$

  - T: reference dataset with N elements
  - b: mean bias, constant for each dataset
  - ε: random variations with mean of 0
- For us
  - T will be “Truth”
  - ε will be “errors”

Theory

- The variance of their differences can be written

$$\begin{aligned}
\mathrm{Var}\left[X_n - Y_n\right] &= \mathrm{Var}\left[\varepsilon_{X,n}\right] + \mathrm{Var}\left[\varepsilon_{Y,n}\right] - 2\mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] \\
\mathrm{Var}\left[X_n - Z_n\right] &= \mathrm{Var}\left[\varepsilon_{X,n}\right] + \mathrm{Var}\left[\varepsilon_{Z,n}\right] - 2\mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] \\
\mathrm{Var}\left[Y_n - Z_n\right] &= \mathrm{Var}\left[\varepsilon_{Y,n}\right] + \mathrm{Var}\left[\varepsilon_{Z,n}\right] - 2\mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right]
\end{aligned}$$

- The linear combinations give solutions for the error variances

$$\begin{aligned}
\mathrm{Var}\left[\varepsilon_{X,n}\right] = \frac{1}{2}&\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[X_n - Z_n\right] - \mathrm{Var}\left[Y_n - Z_n\right]\right) \\
&+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] + \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] \\
\mathrm{Var}\left[\varepsilon_{Y,n}\right] = \frac{1}{2}&\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Z_n\right]\right) \\
&+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] + \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] \\
\mathrm{Var}\left[\varepsilon_{Z,n}\right] = \frac{1}{2}&\left(\mathrm{Var}\left[X_n - Z_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Y_n\right]\right) \\
&+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] + \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right]
\end{aligned}$$

- These are the three-cornered hat (3CH) error variance relations

Theory: 3CH method key points and caveats

- Established: history in atomic clock (Gray and Allan 1974), SST (O'Carroll et al. 2008) error estimations
- Exact
- Straightforward to compute
- Does not rely on knowing truth
- Removes the impact of mean biases
- Smaller estimates than variance of differences

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

... except when error covariance is large

- Includes all sources of "error" ε : instrument, co-location, representativeness, etc.

$$\begin{aligned}
\mathrm{Var}\left[\varepsilon_{X,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[X_n - Z_n\right] - \mathrm{Var}\left[Y_n - Z_n\right]\right) \\
+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] + \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] \\
\mathrm{Var}\left[\varepsilon_{Y,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Z_n\right]\right) \\
+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] + \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] \\
\mathrm{Var}\left[\varepsilon_{Z,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Z_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Y_n\right]\right) \\
+ \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Z,n}\right] + \mathrm{Cov}\left[\varepsilon_{Y,n}, \varepsilon_{Z,n}\right] - \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right]
\end{aligned}$$

Theory: 3CH method key points and caveats

- Established: history in atomic clock (Gray and Allan 1974), SST (O'Carroll et al. 2008) error estimations
- Exact…if ε are not correlated
- Straightforward to compute
- Does not rely on knowing truth
- Removes the impact of mean biases
- Smaller estimates than variance of differences

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

... except when error covariance is large

- Includes all sources of "error" ε : instrument, co-location, representativeness, etc.

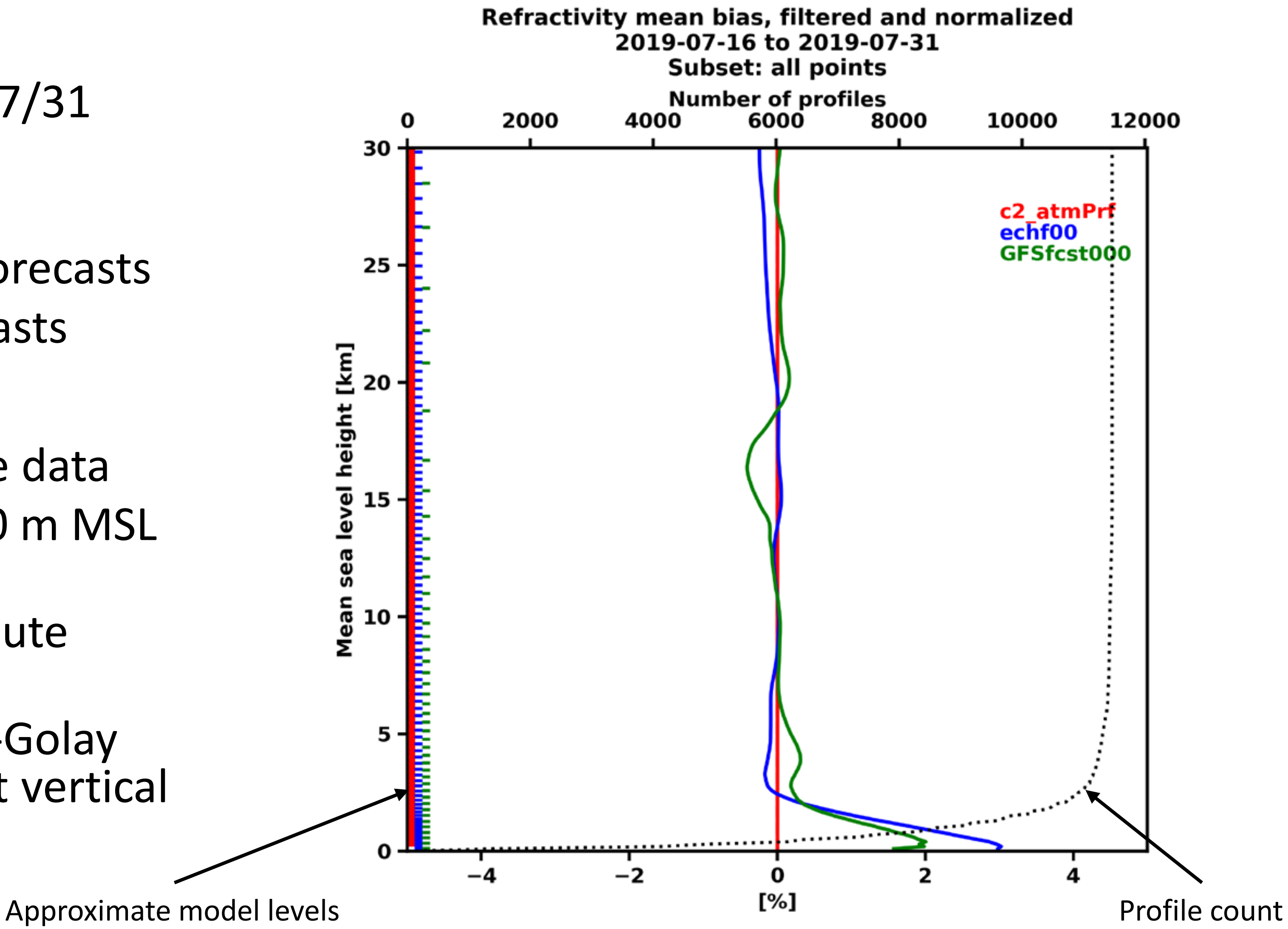
$$\mathrm{Var}\left[\varepsilon_{X,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[X_n - Z_n\right] - \mathrm{Var}\left[Y_n - Z_n\right]\right)$$

$$\mathrm{Var}\left[\varepsilon_{Y,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Z_n\right]\right)$$

$$\mathrm{Var}\left[\varepsilon_{Z,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Z_n\right] + \mathrm{Var}\left[Y_n - Z_n\right] - \mathrm{Var}\left[X_n - Y_n\right]\right)$$

Initial COSMIC-2 analysis

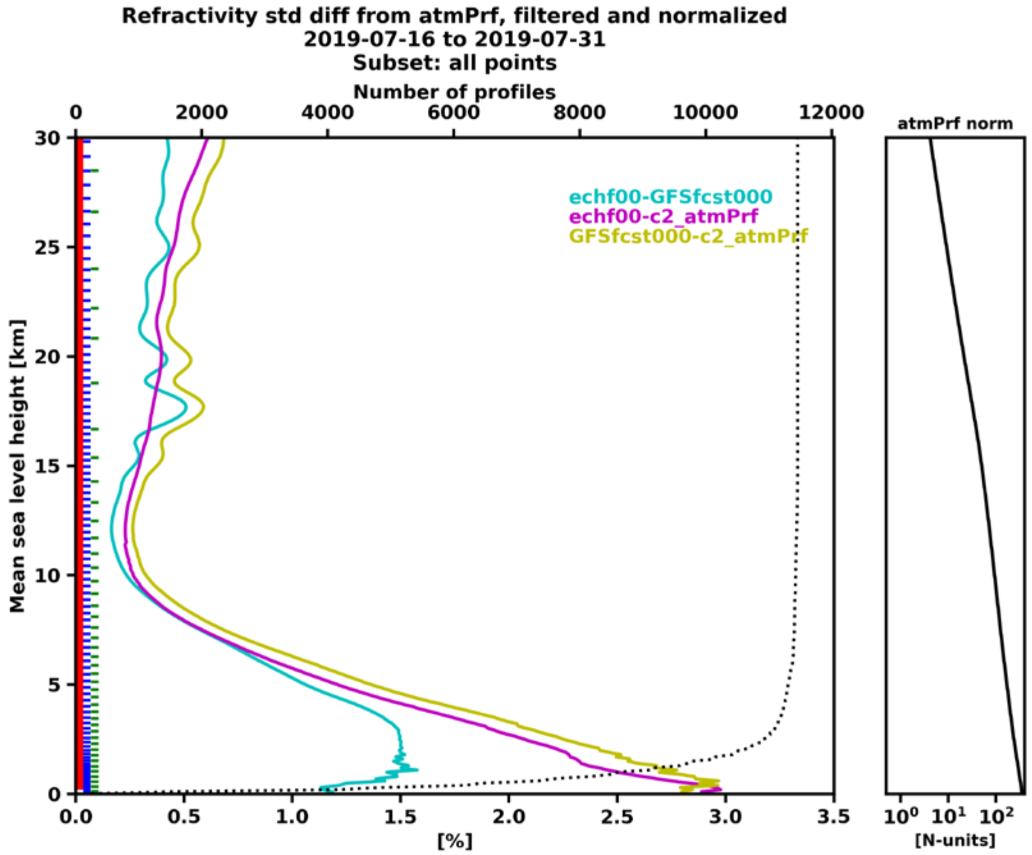
- Data: refractivity
  - 2019/07/16 (2019.197) – 2019/07/31 (2019.212)
  - All C2 atmPrf
  - Co-located ECMWF analyses or forecasts
  - Co-located GFS analyses or forecasts
- Method
  - Horizontally, temporally co-locate data
  - Interpolate fields to common 100 m MSL height grid
  - Detect and remove median absolute deviation outliers
  - Apply 4500 m, 3rd-order Savitzky-Golay filter to help account for different vertical resolution
  - Calculate 3CH relations



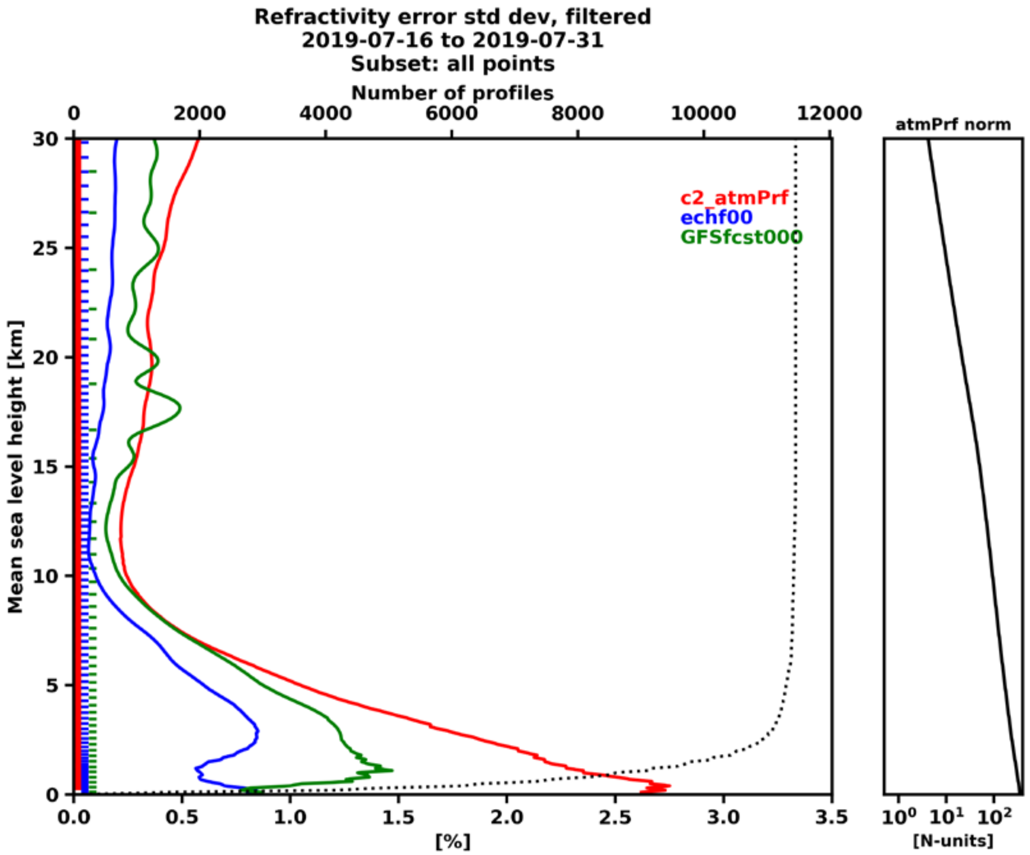
Initial COSMIC-2 analysis: refractivity

C2 atmPrf + EC analysis + GFS analysis

Std dev of differences



3CH estimates

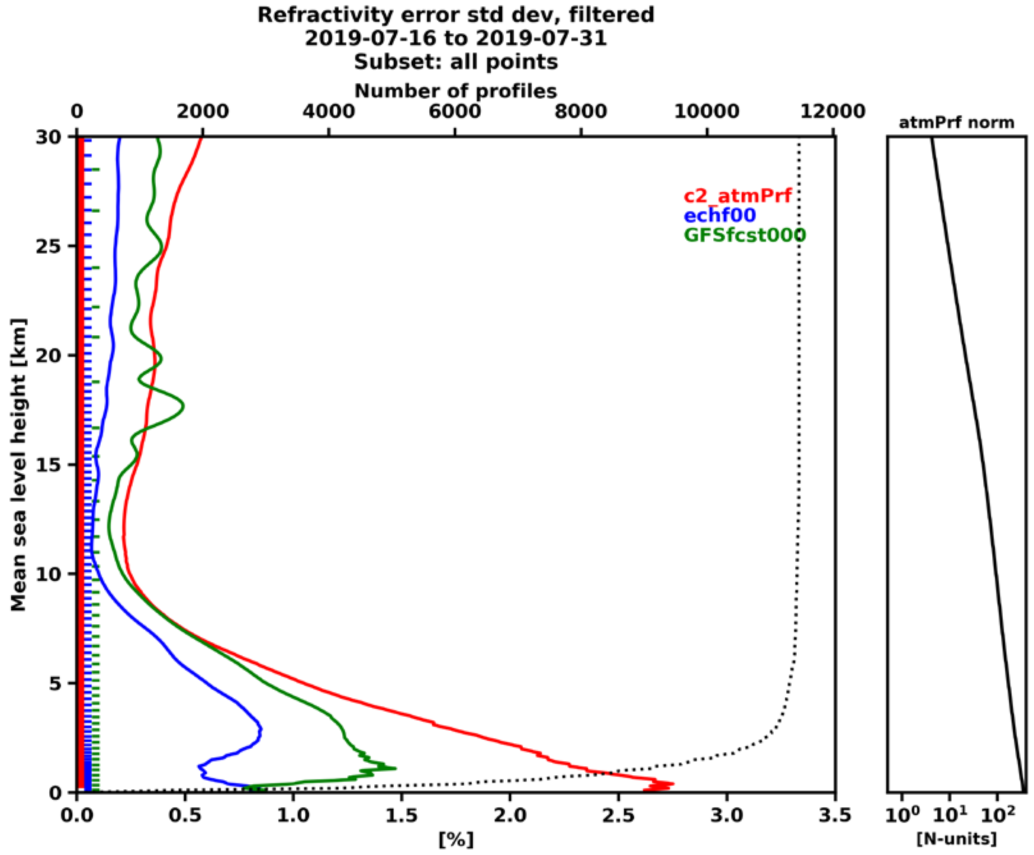


Normalized error standard deviations comparable to (O-B)/B relationship

Initial COSMIC-2 sensitivity analysis: refractivity

- We can test sensitivity of our results by analyzing other triplets of data sets
    - In particular, how good is our assumption of zero (minimal) error covariance?
- Here: substitute forecast data for model analyses
    - Error covariance should be reduced

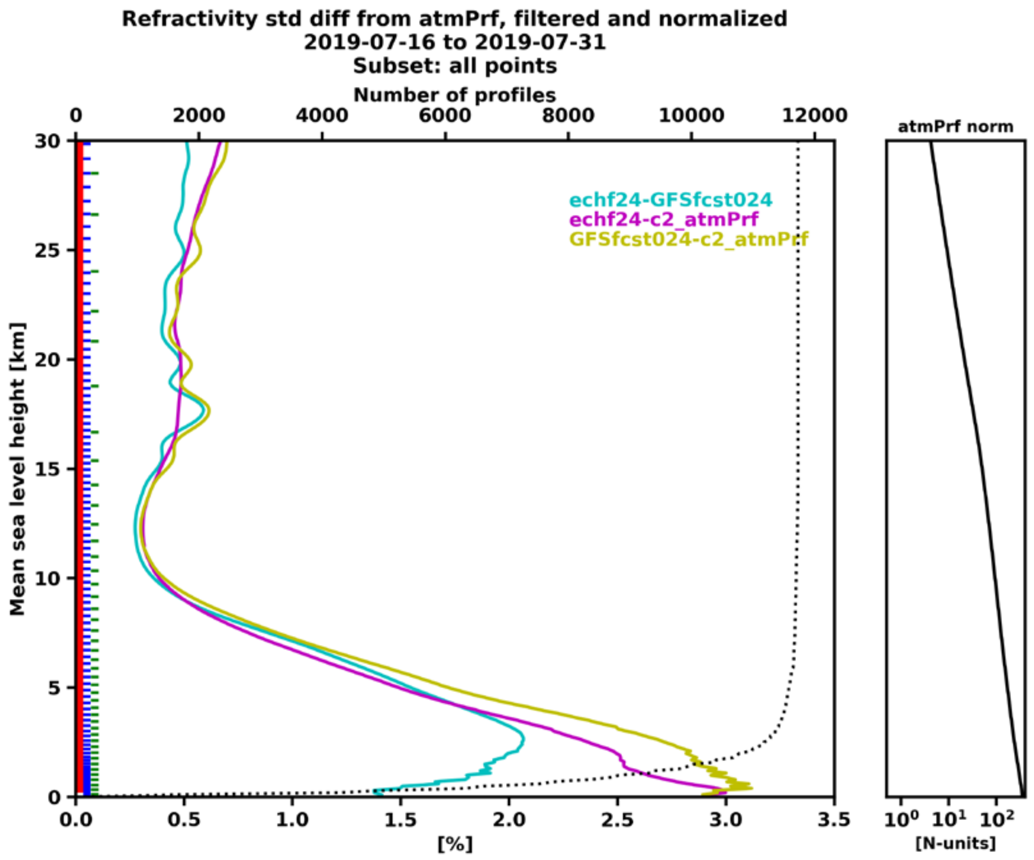
3CH estimates



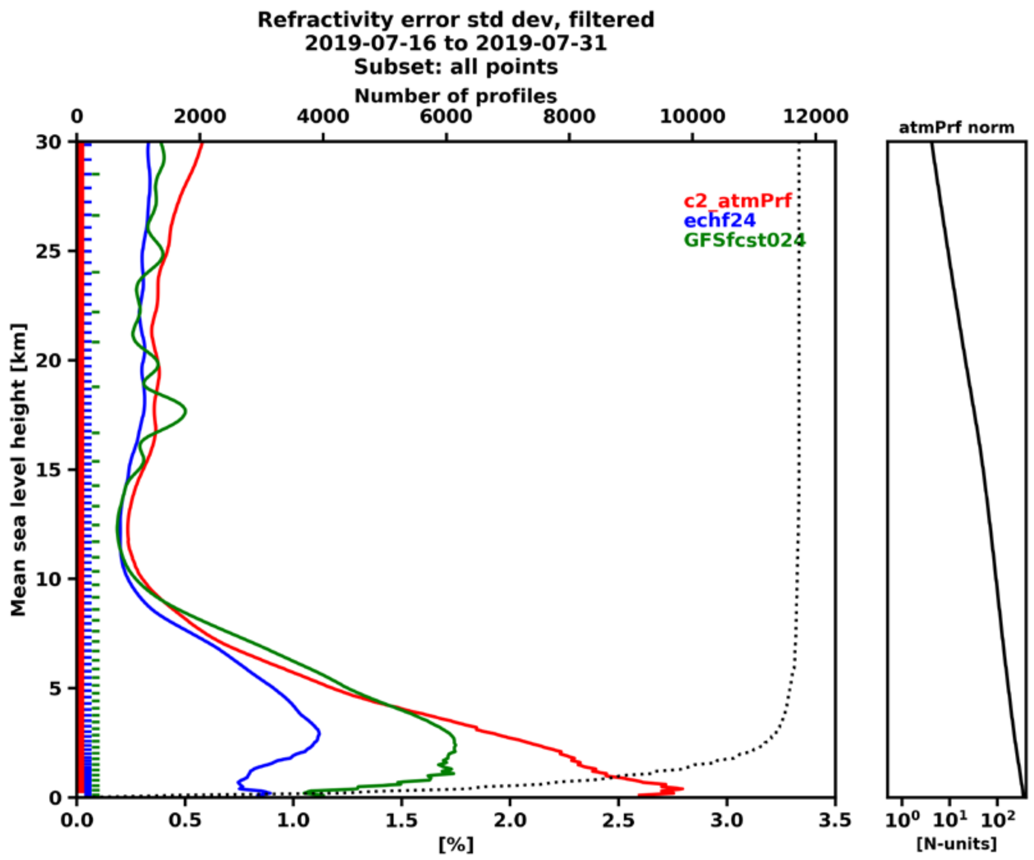
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 24 hr fcst

Std dev of differences



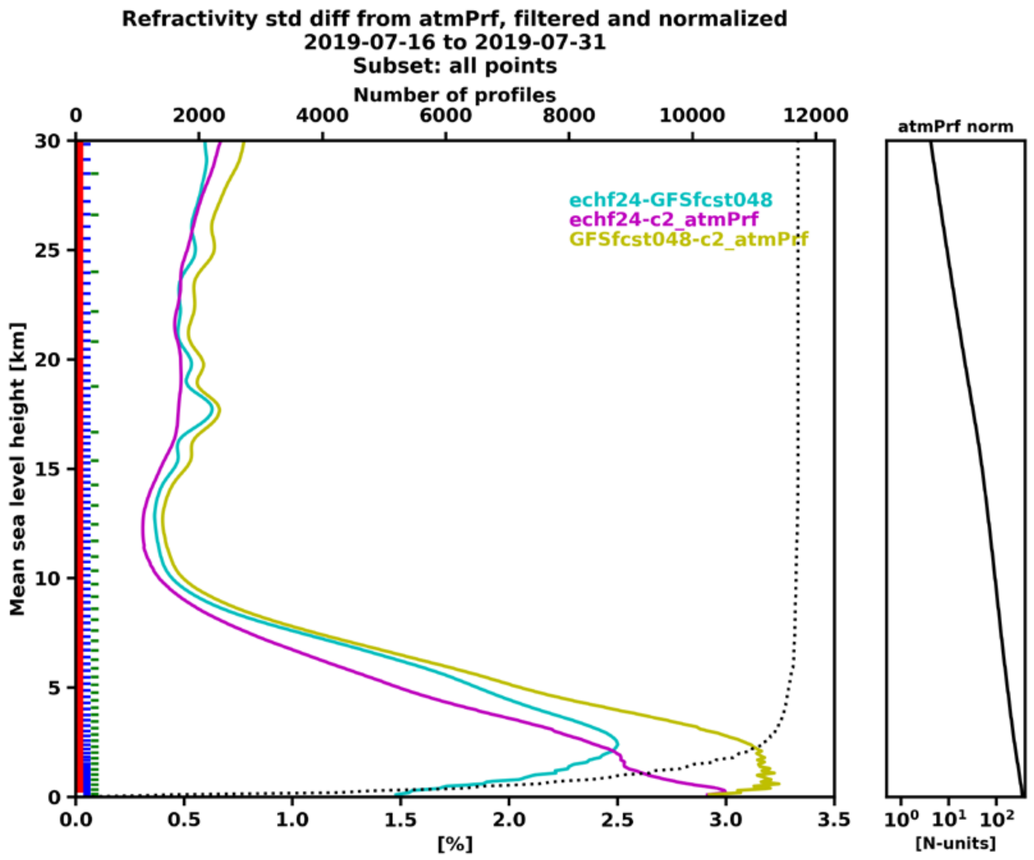
3CH estimates



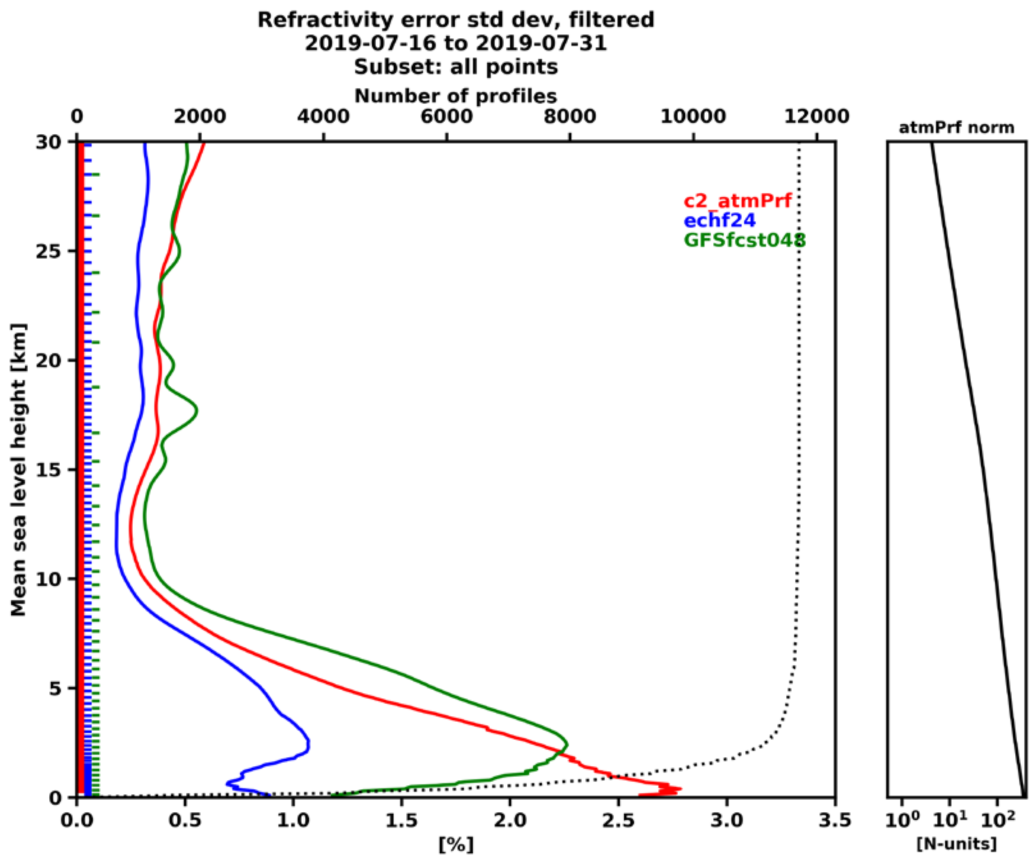
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 48 hr fcst

Std dev of differences



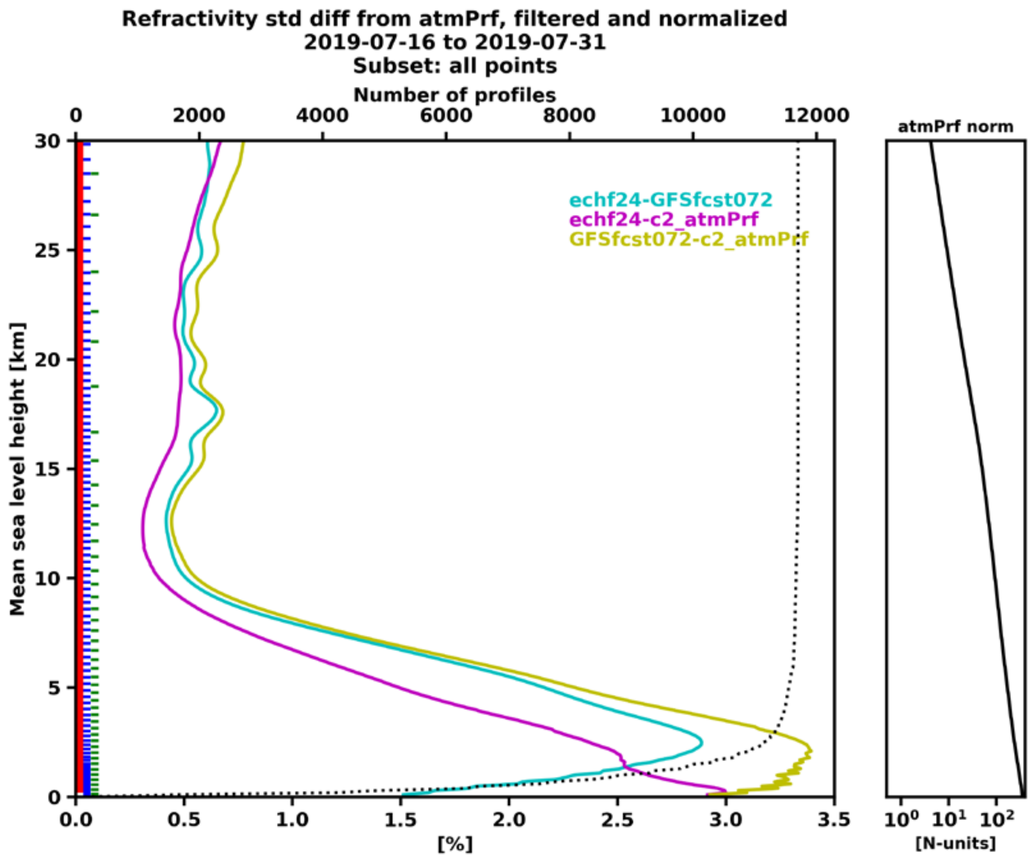
3CH estimates



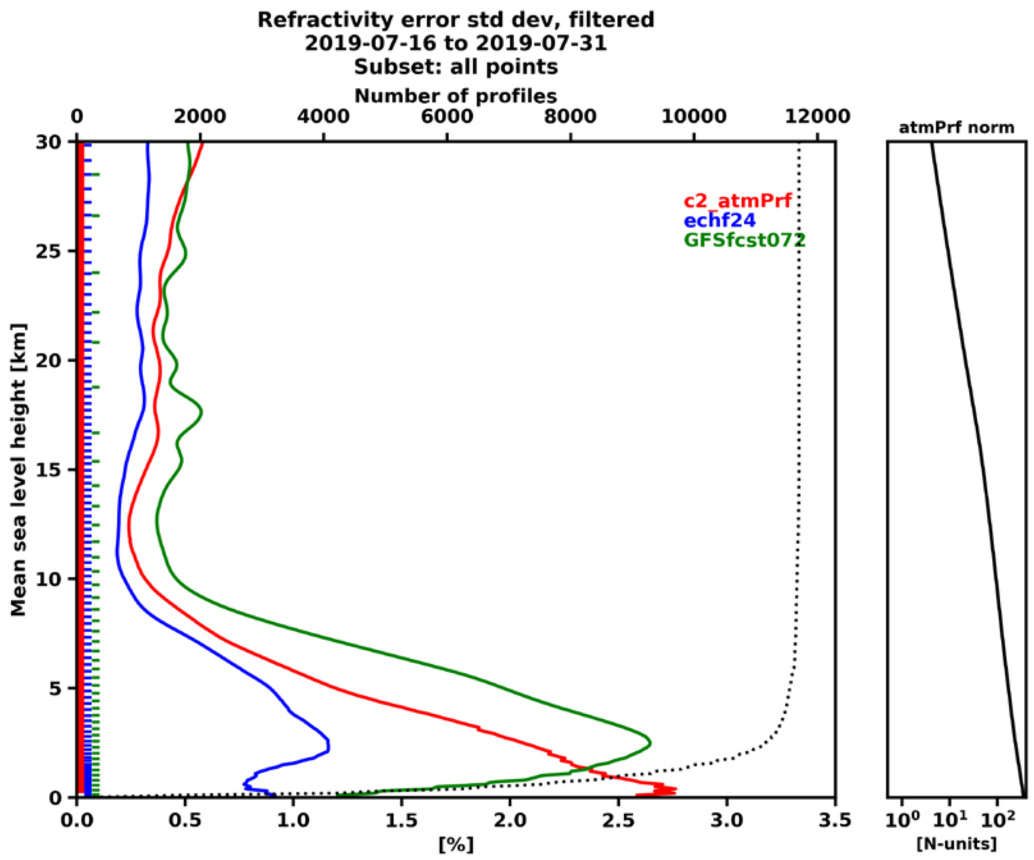
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 72 hr fcst

Std dev of differences



3CH estimates

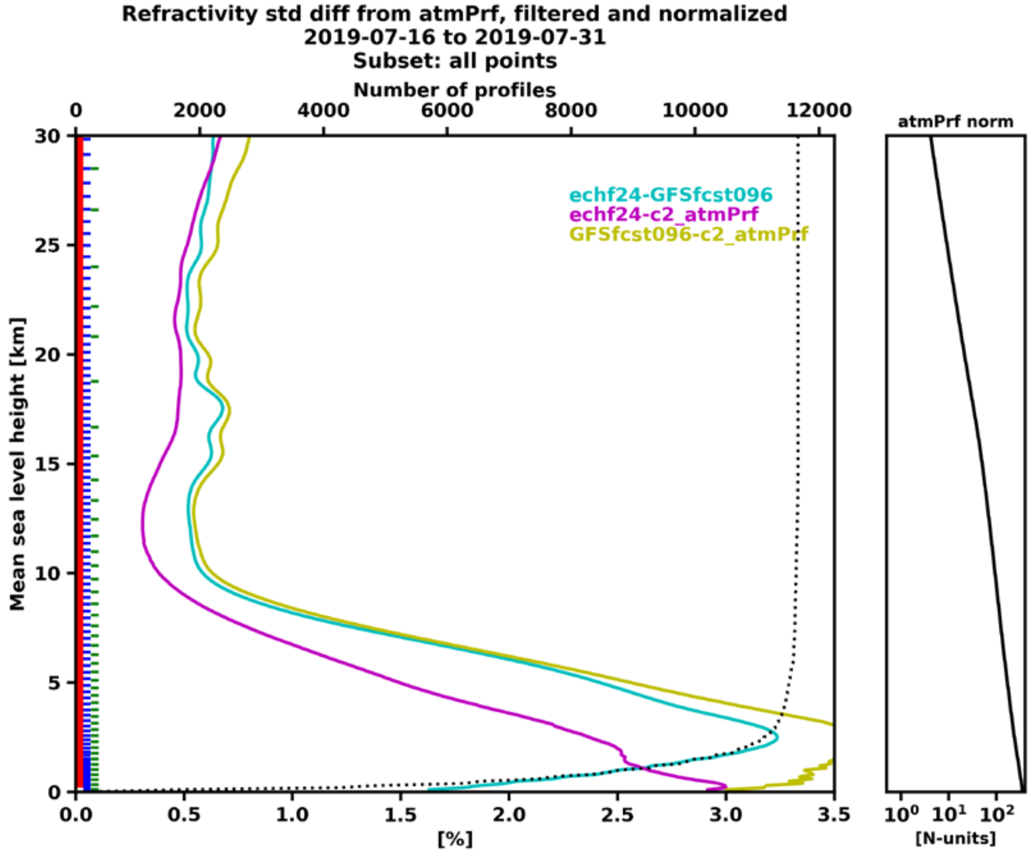


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

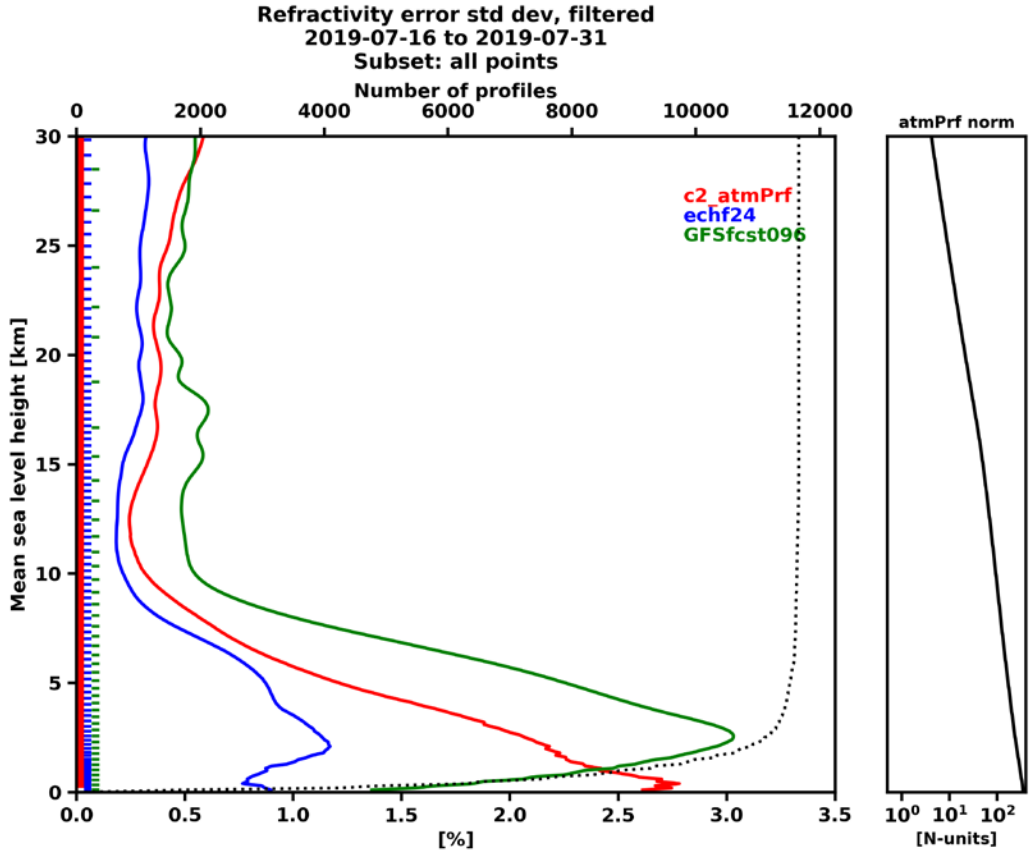
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 96 hr fcst

Std dev of differences



3CH estimates

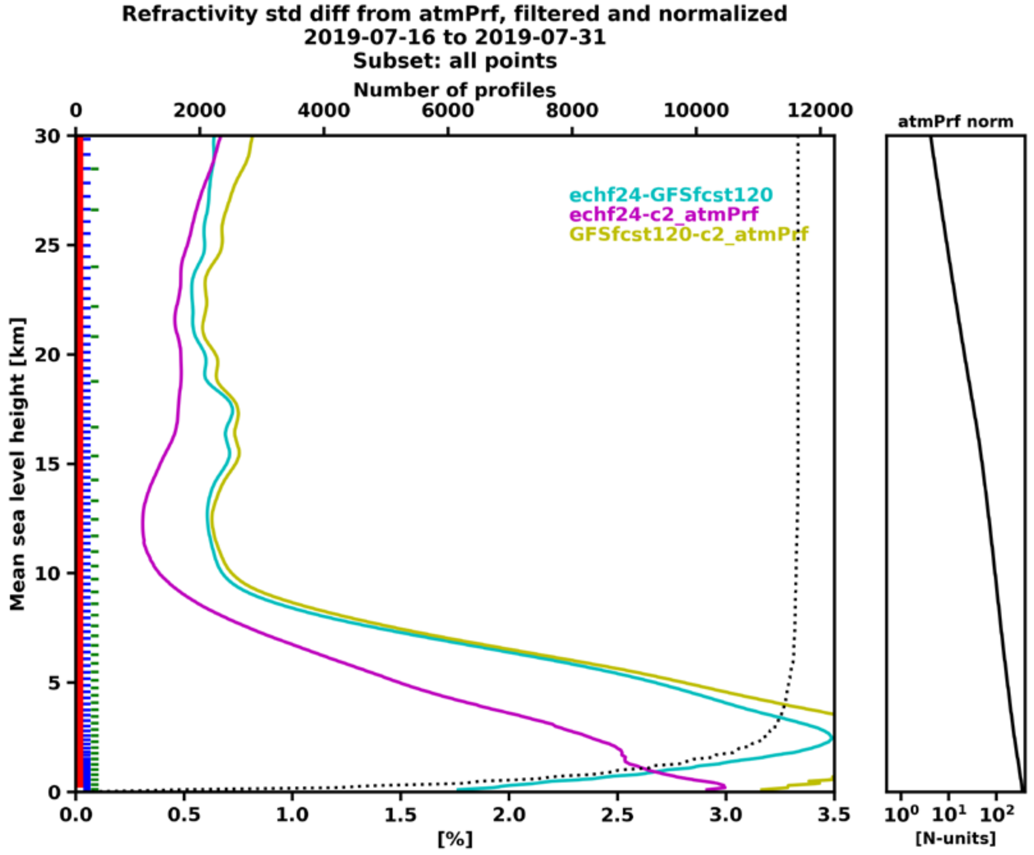


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

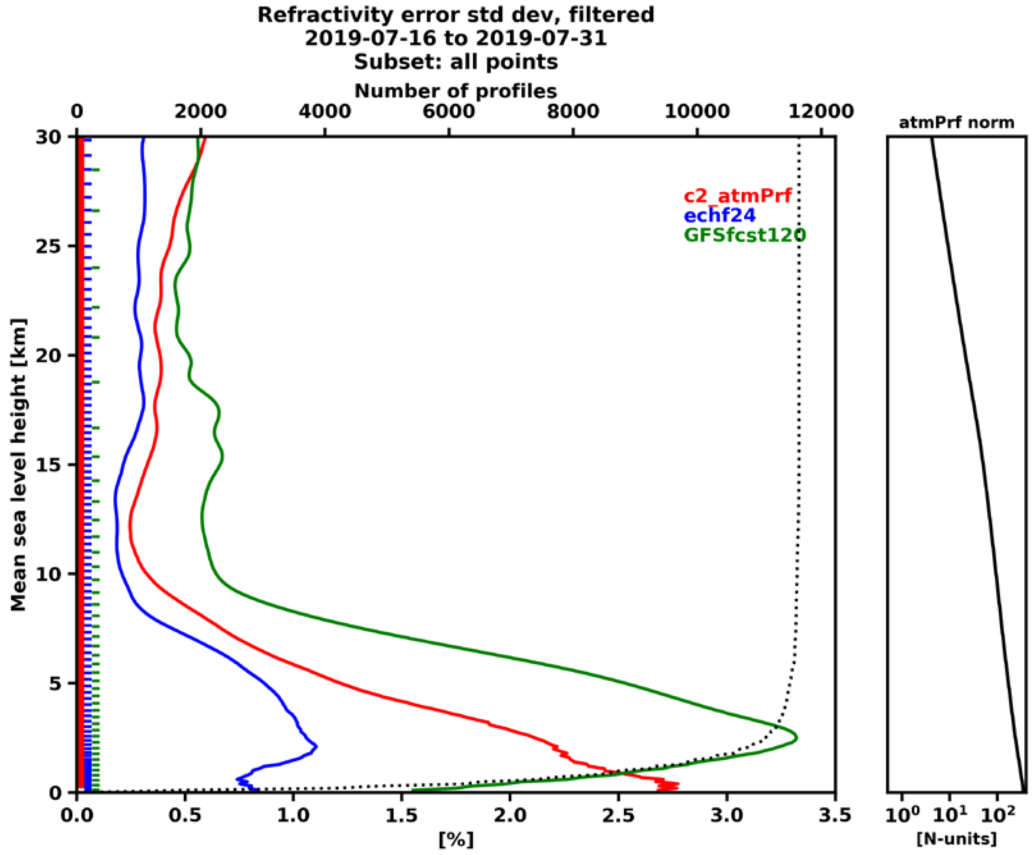
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 120 hr fcst

Std dev of differences



3CH estimates

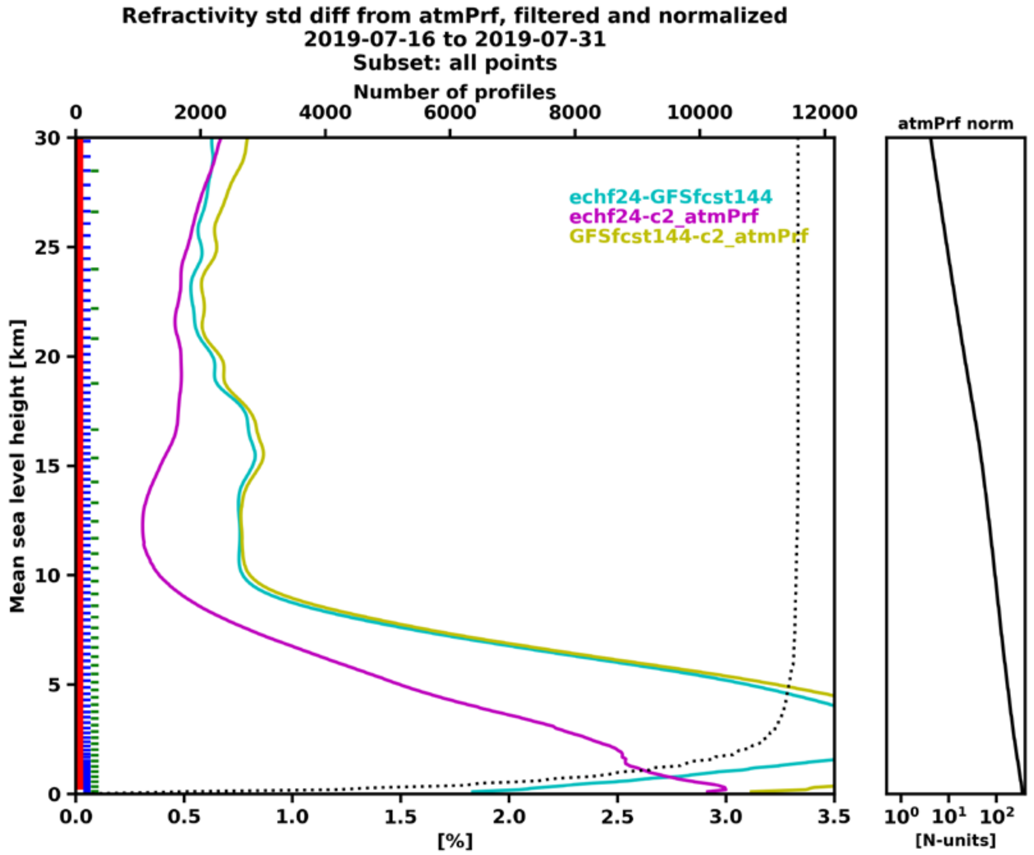


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

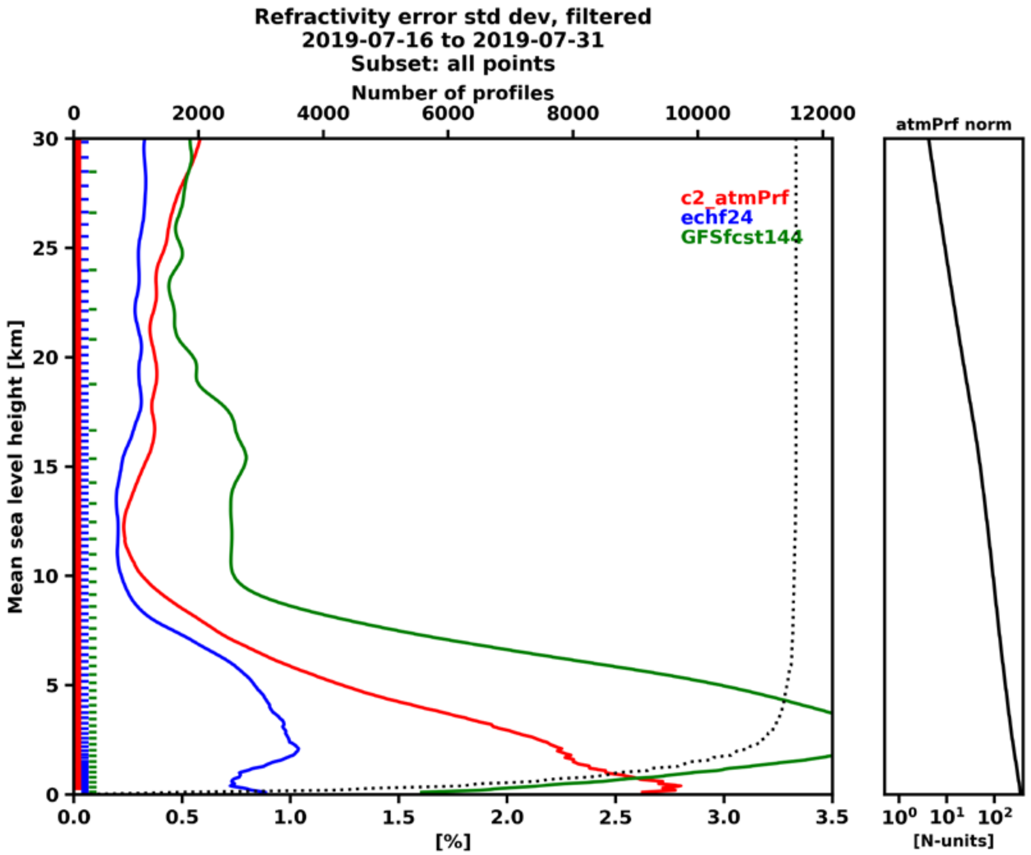
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 144 hr fcst

Std dev of differences



3CH estimates

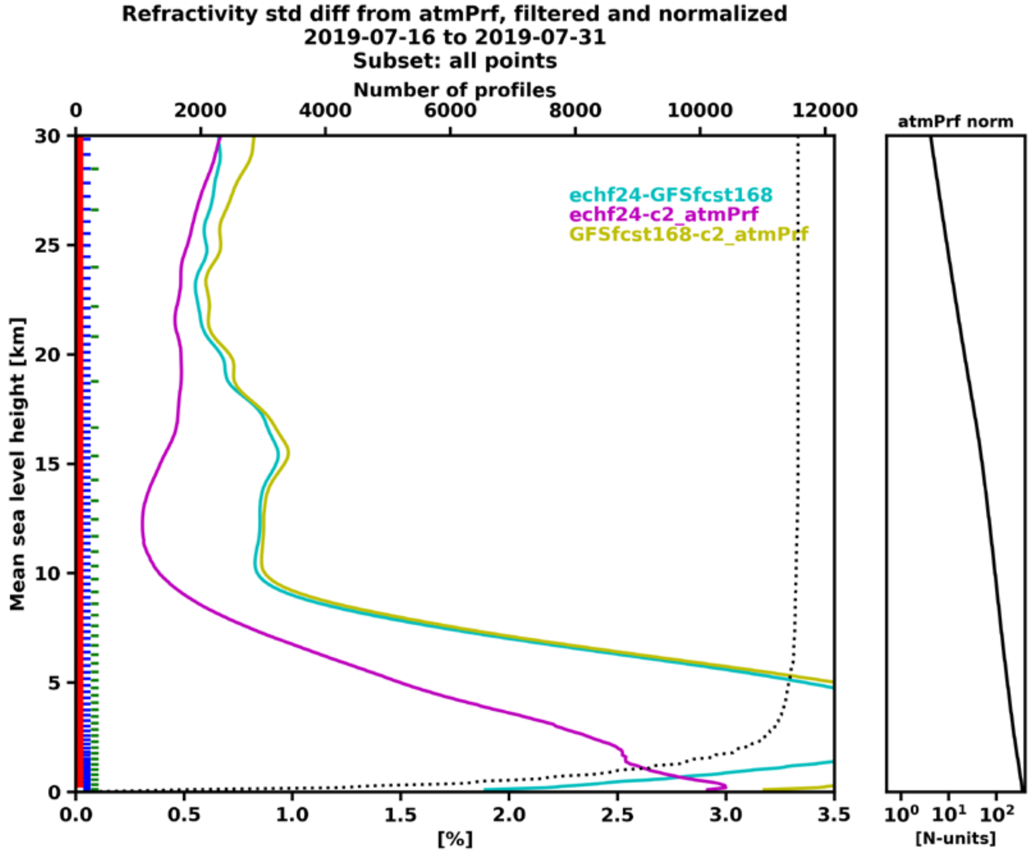


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

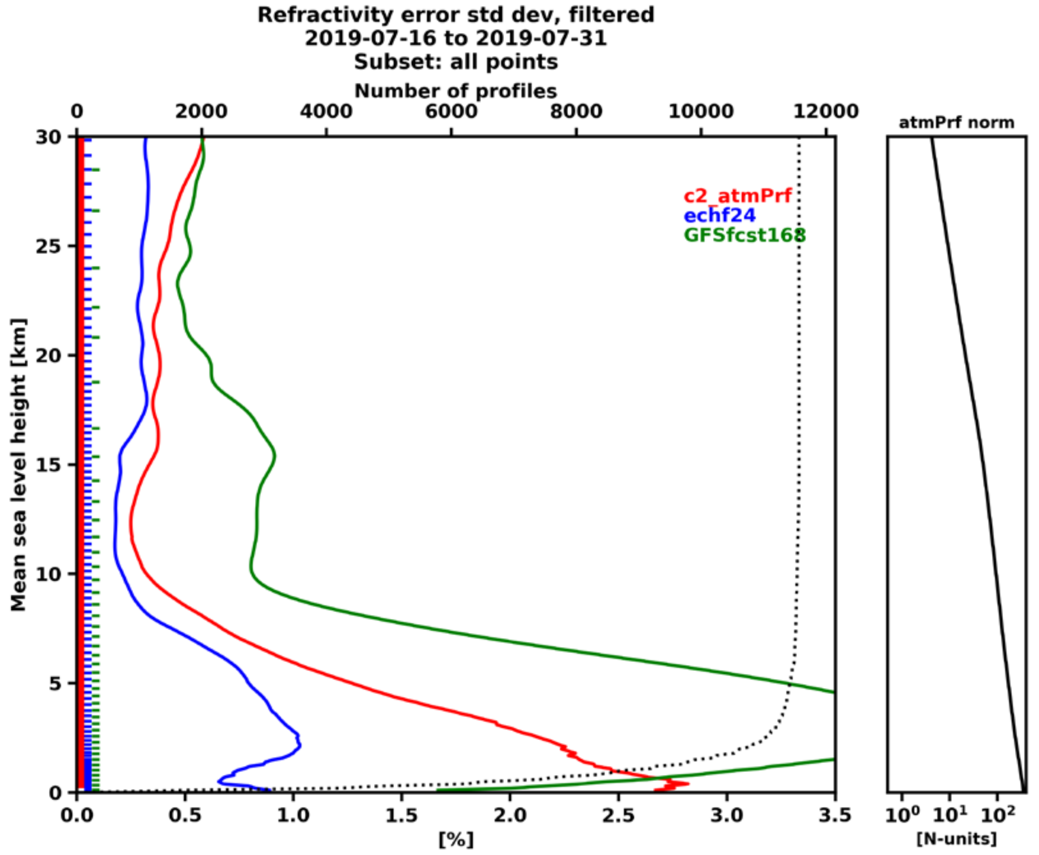
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 168 hr fcst

Std dev of differences



3CH estimates

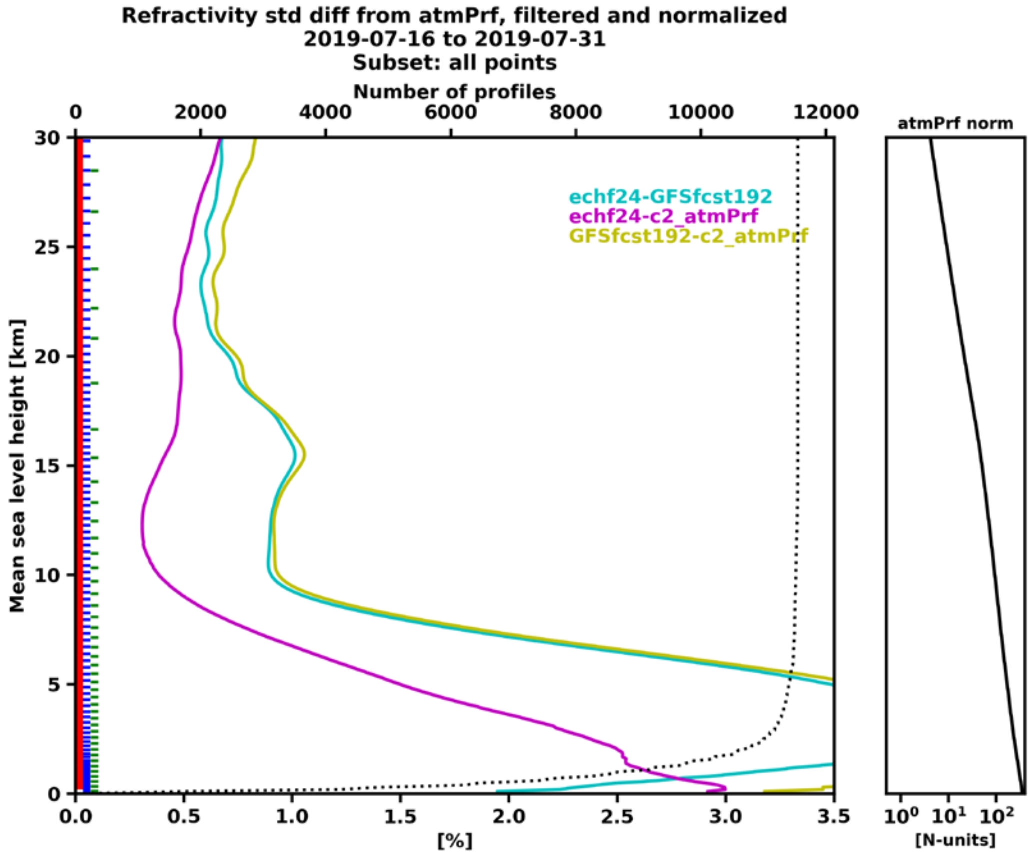


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

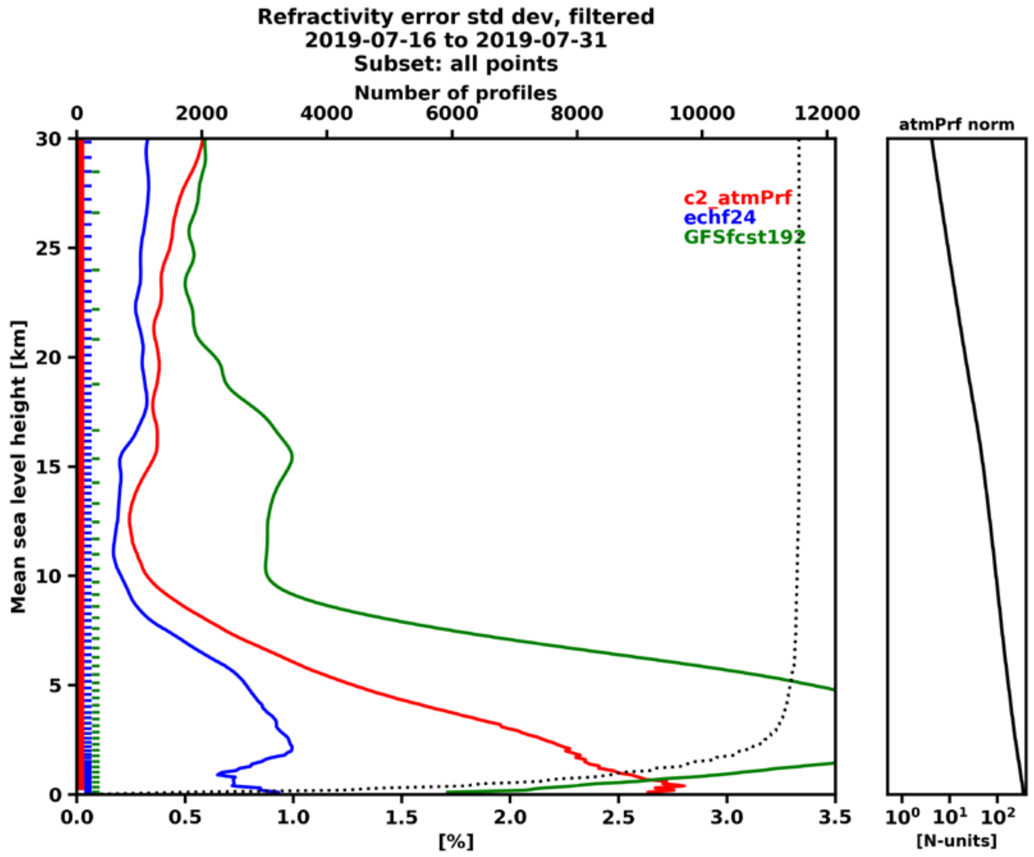
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 192 hr fcst

Std dev of differences



3CH estimates

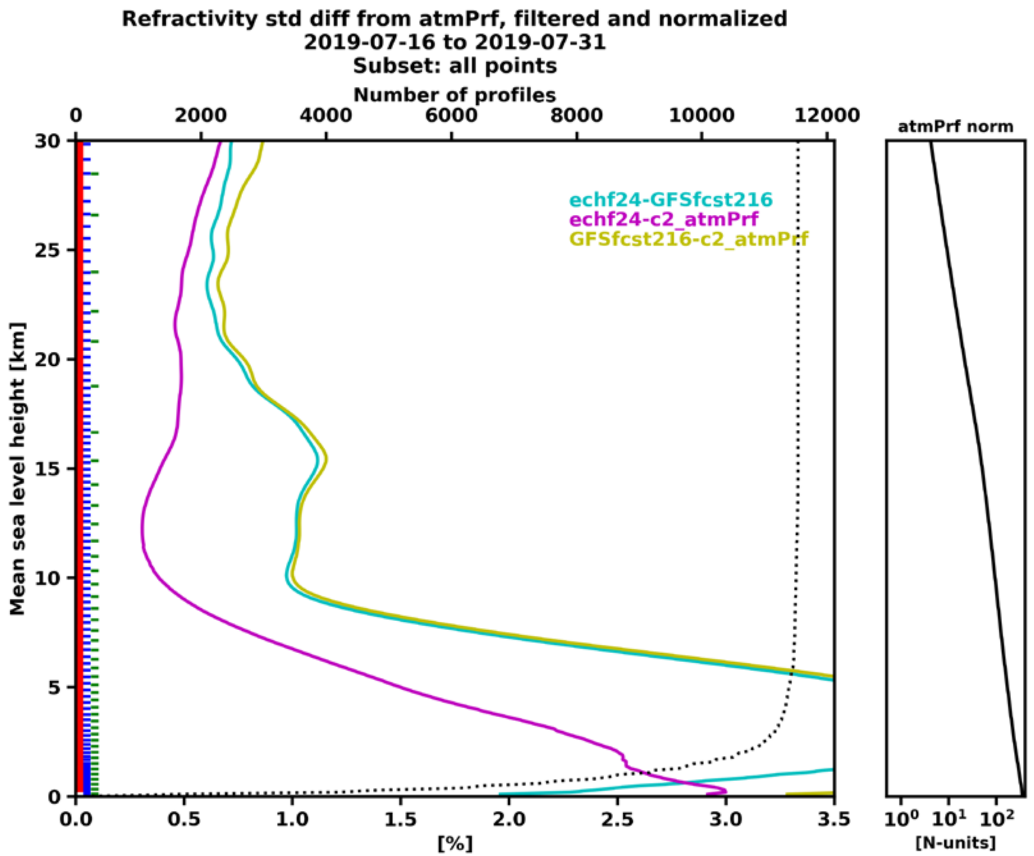


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

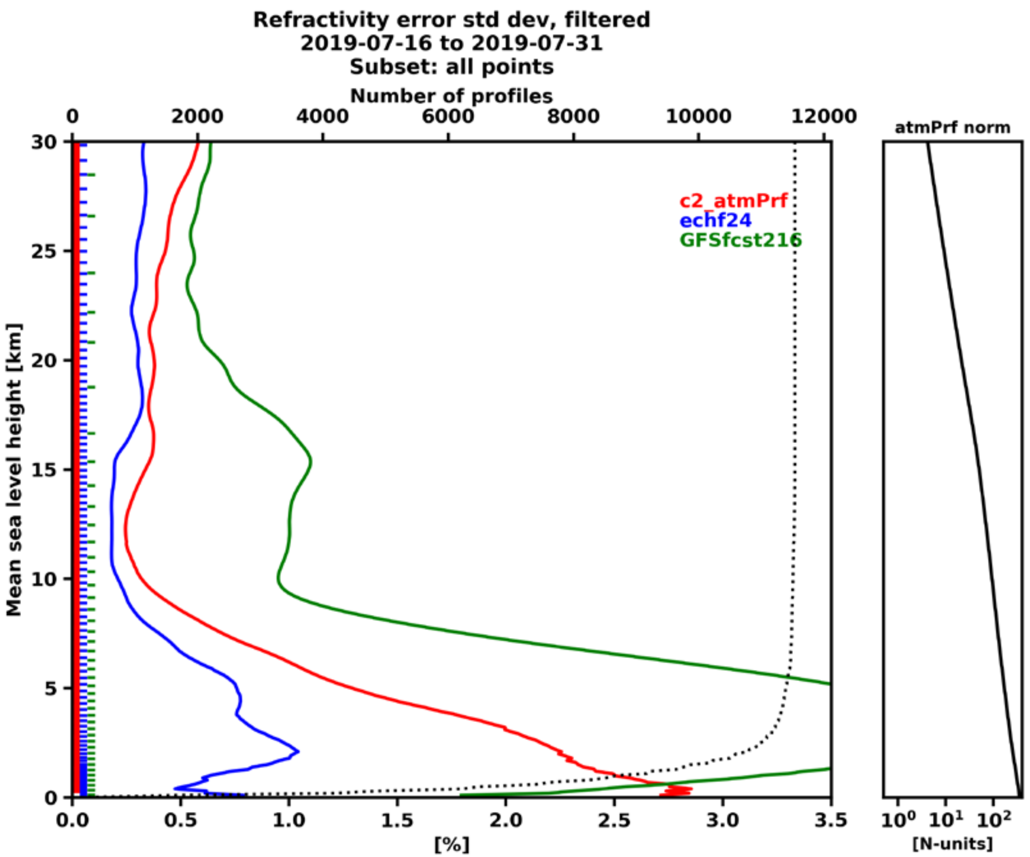
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 216 hr fcst

Std dev of differences



3CH estimates

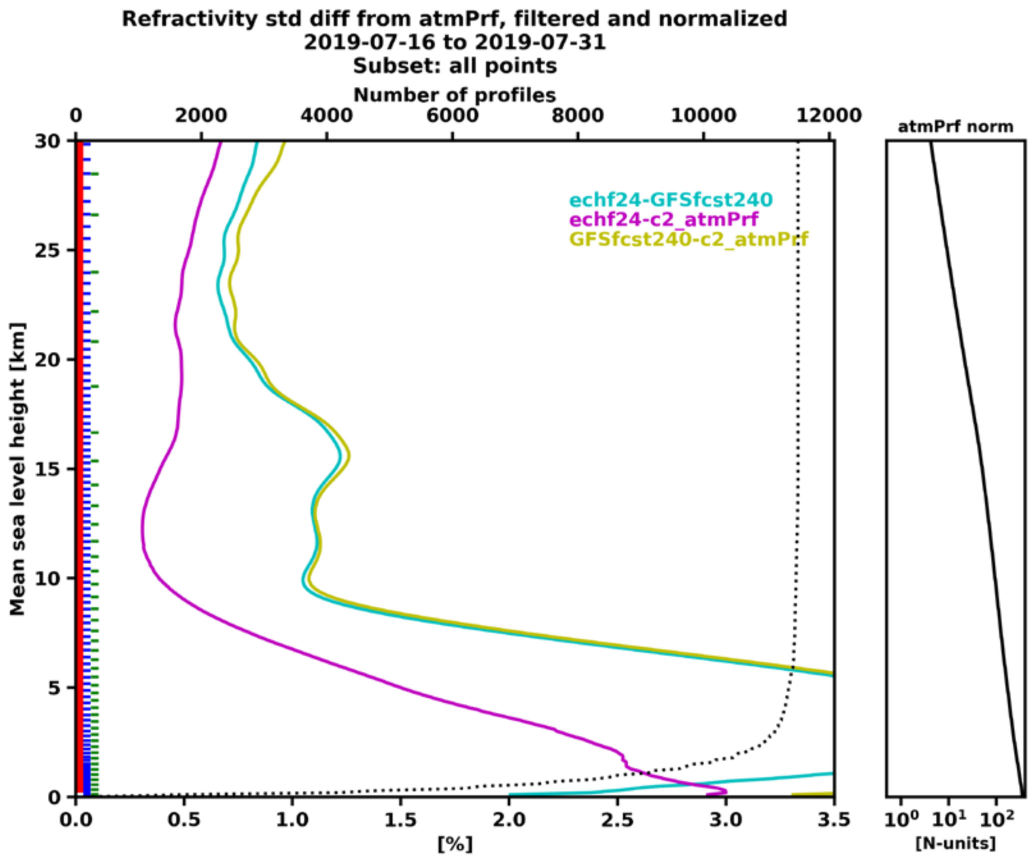


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

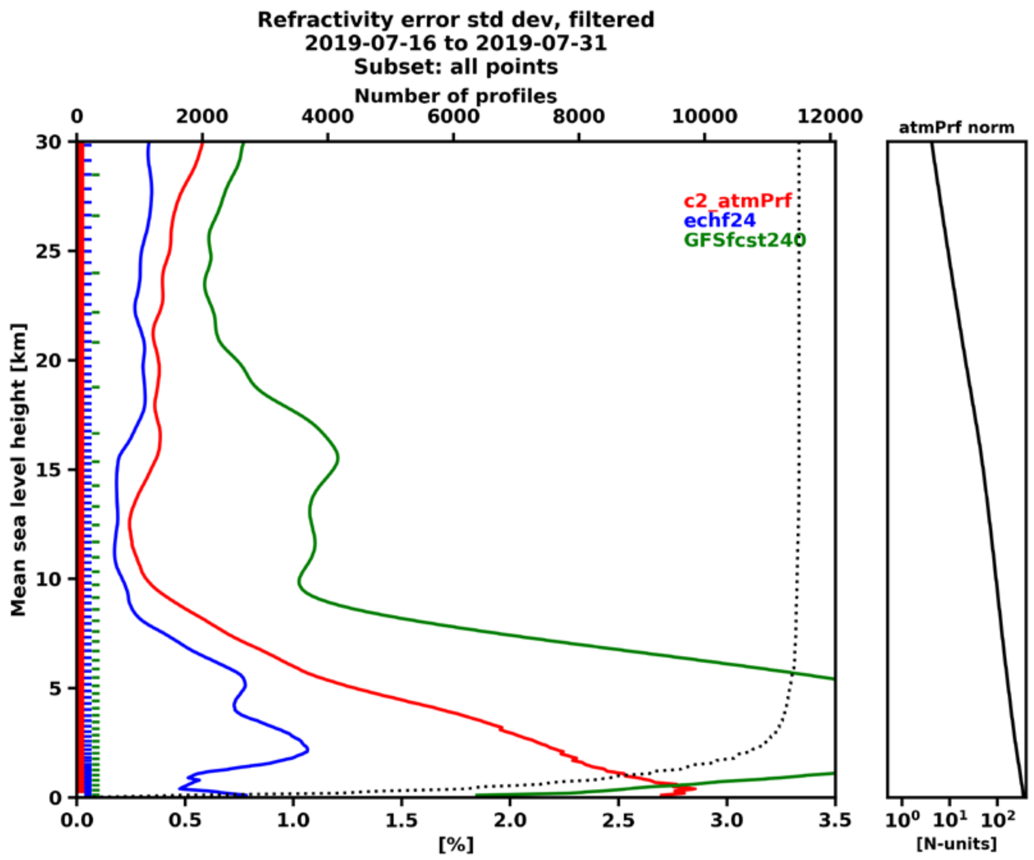
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 240 hr fcst

Std dev of differences



3CH estimates

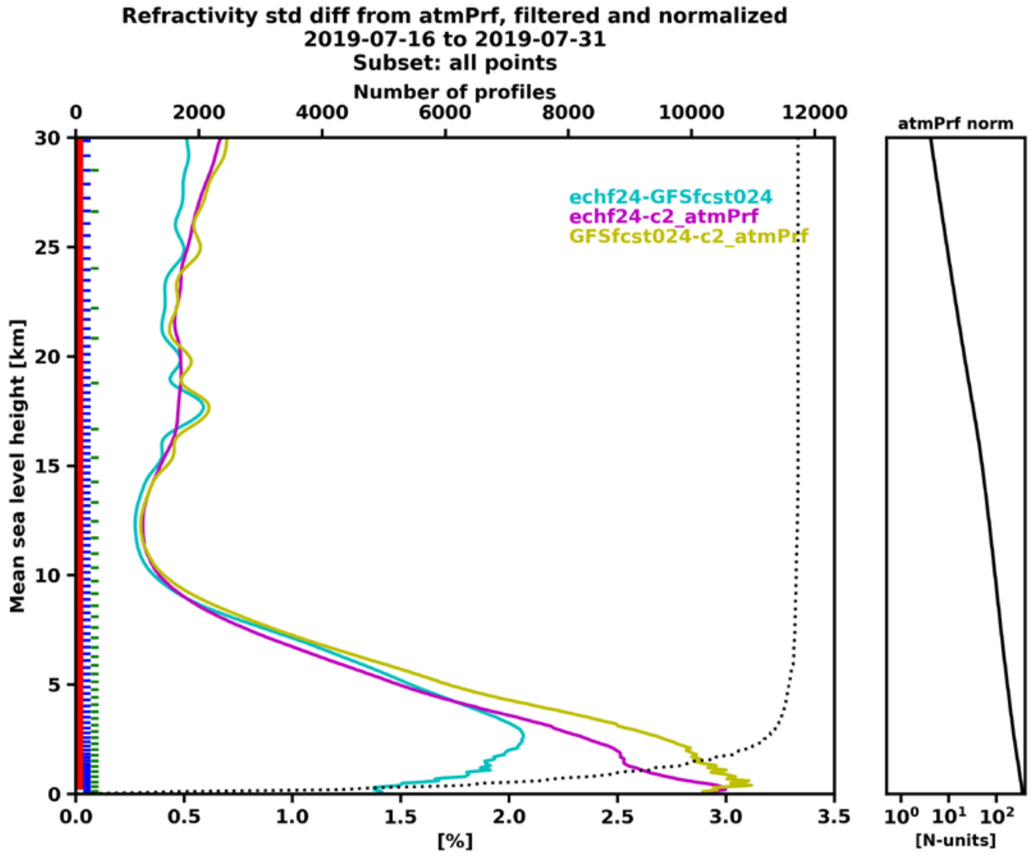


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

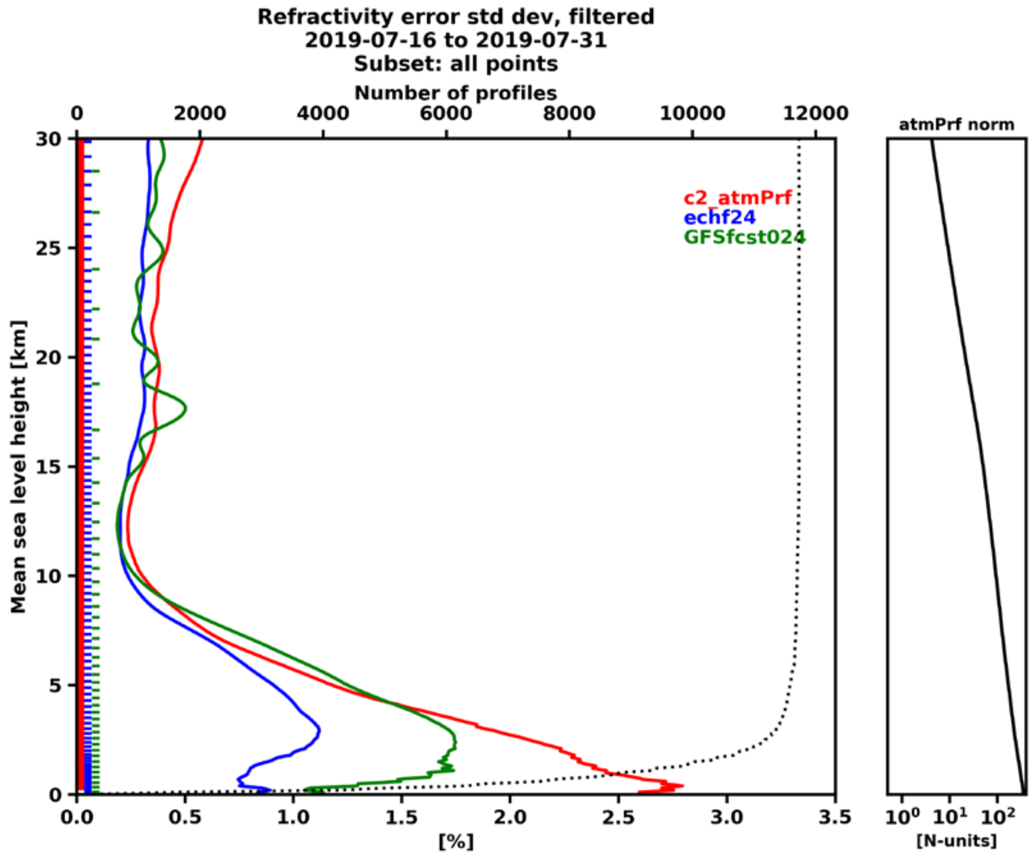
Initial COSMIC-2 sensitivity analysis: refractivity

C2 atmPrf + EC 24 hr fcst + GFS 24 hr fcst

Std dev of differences



3CH estimates

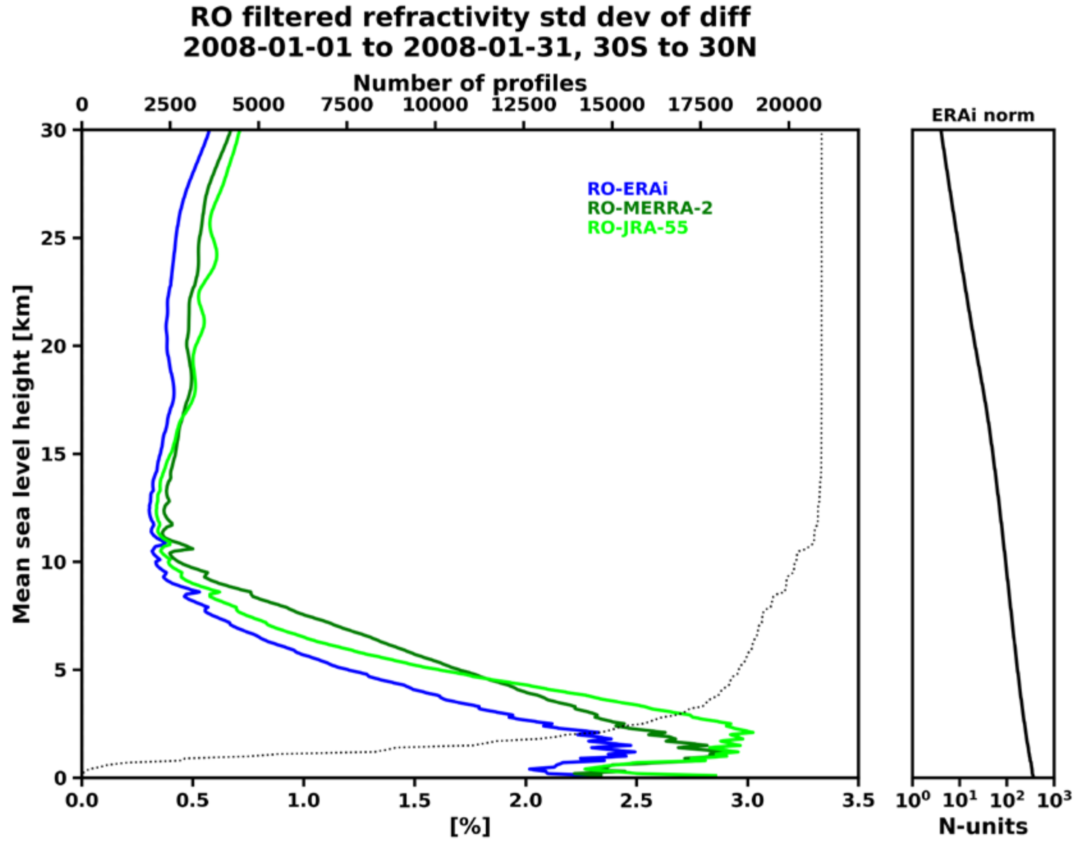


Forecast error of GFS grows with increasing forecast time;
C2, EC errors remain steady

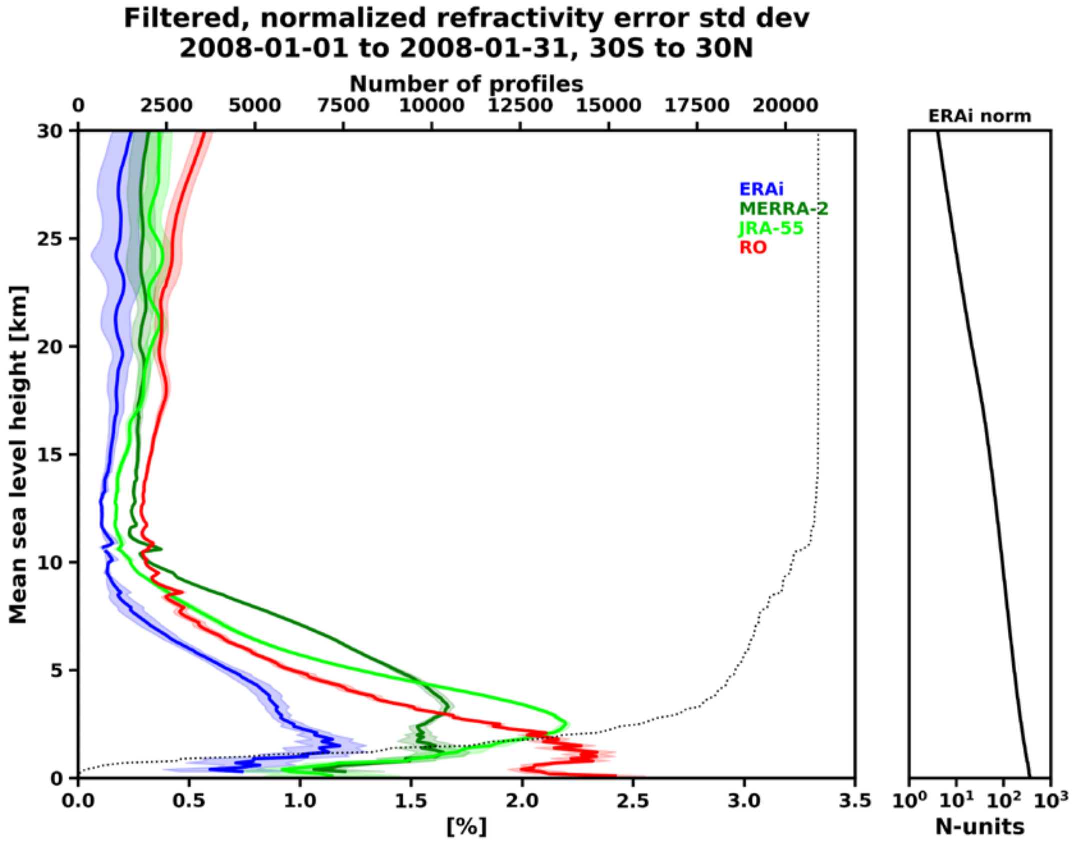
Post-processed COSMIC-1 results: refractivity

C1 atmPrf + ERAi + MERRA-2 + JRA-55

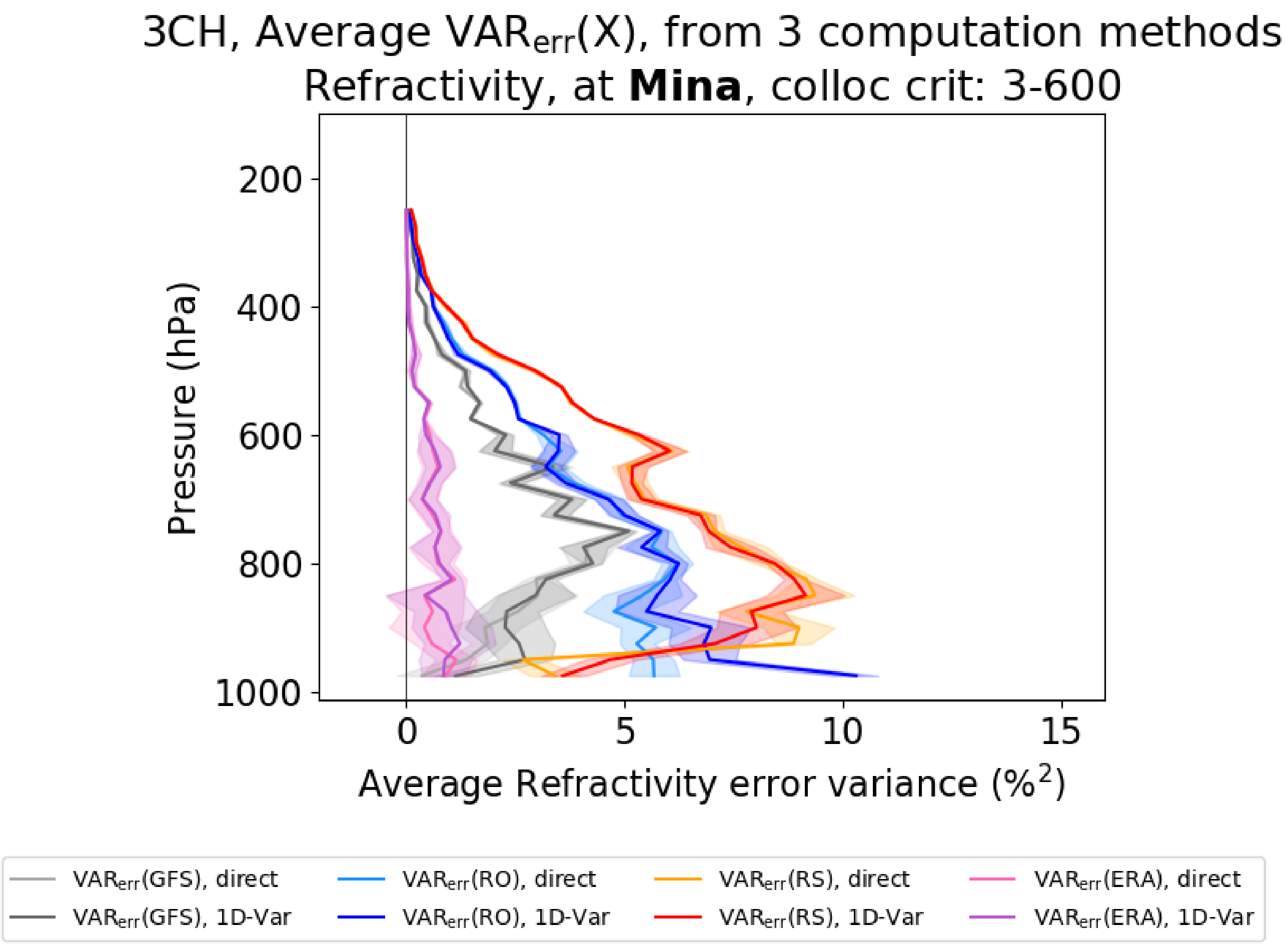
Std dev of differences

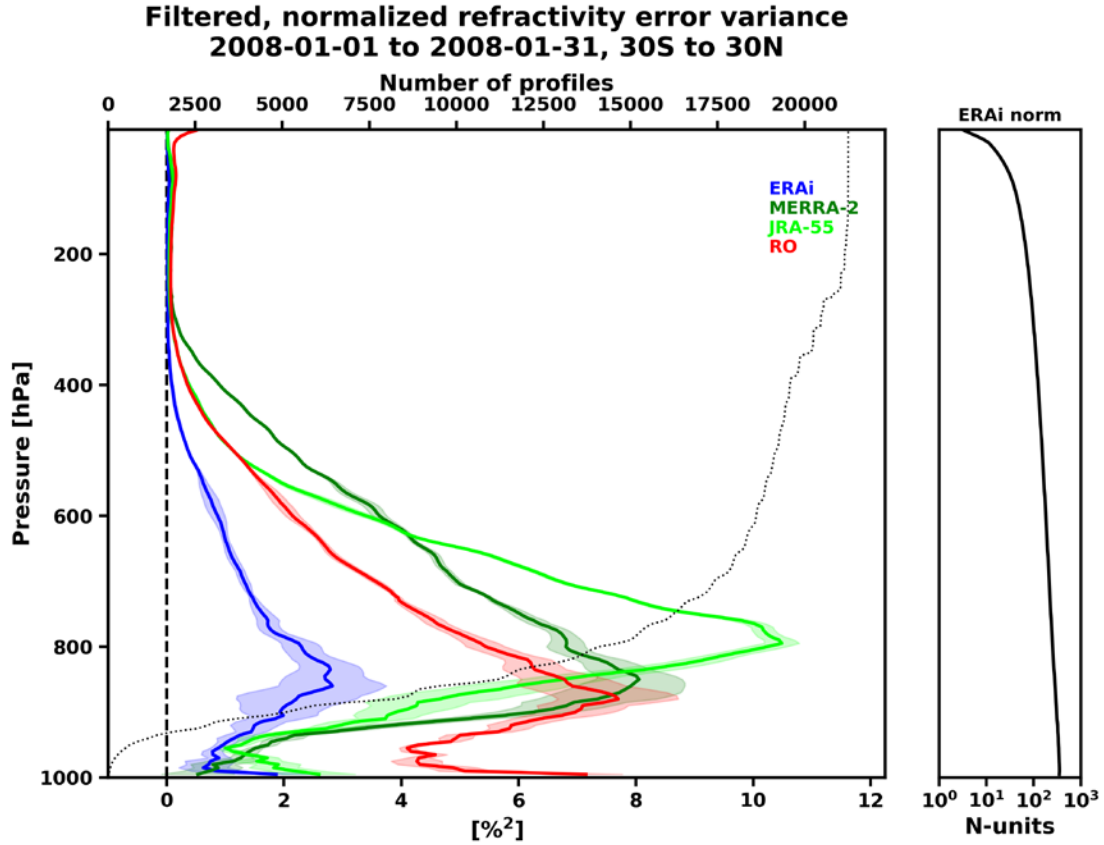


3CH estimates



COSMIC-1 results

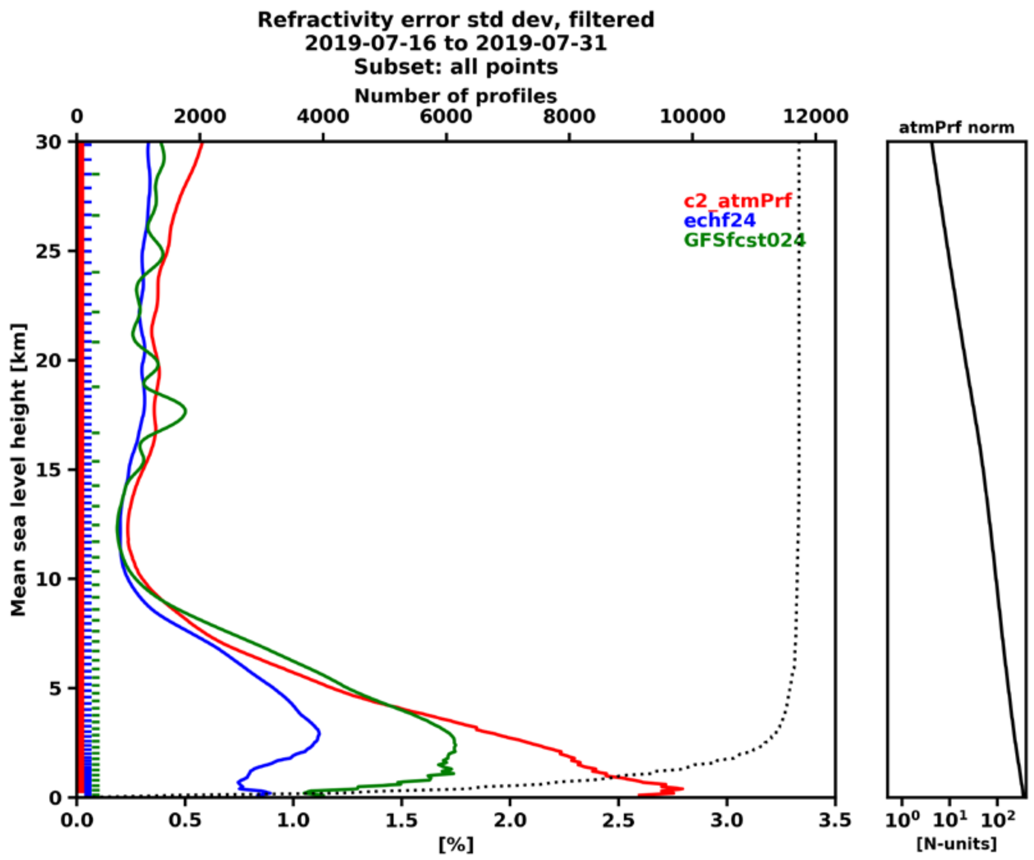




Anthes and Rieckh (2018),
DOI:10.5194/amt-11-4239-2018

Summary

- We apply the 3CH method for estimating error variance (std dev) with RO+model data
- Substituting/including more data sets allows us to identify triplets with no (minimal) error covariance
- We find…
  - consistent, smaller estimates of error statistics when compared with traditional methods
  - estimates of the error variance in models, reanalyses



For more, chat with Rick Anthes and Jeremiah Sjoberg at their posters

Thank you!

Theory

- Aside: relation to error variance estimates with two datasets
  - "Two-cornered hat"
    - Combining the above with

$$\mathrm{Var}\left[X_n\right] - \mathrm{Var}\left[Y_n\right] = \mathrm{Var}\left[\varepsilon_{X,n}\right] - \mathrm{Var}\left[\varepsilon_{Y,n}\right] + 2E\left[(\varepsilon_{X,n} - \varepsilon_{Y,n})T'_n\right]$$

we get the two-cornered hat relation

$$\mathrm{Var}\left[\varepsilon_{X,n}\right] = \frac{1}{2}\left(\mathrm{Var}\left[X_n - Y_n\right] + \mathrm{Var}\left[X_n\right] - \mathrm{Var}\left[Y_n\right]\right) + \mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right] - E\left[(\varepsilon_{X,n} - \varepsilon_{Y,n})T'_n\right]$$

    - This has been used by previous studies (Stoffelen 1998, Vogelzang et al., 2011); also called "triple collocation"
  - (Root) mean square difference with biases removed

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

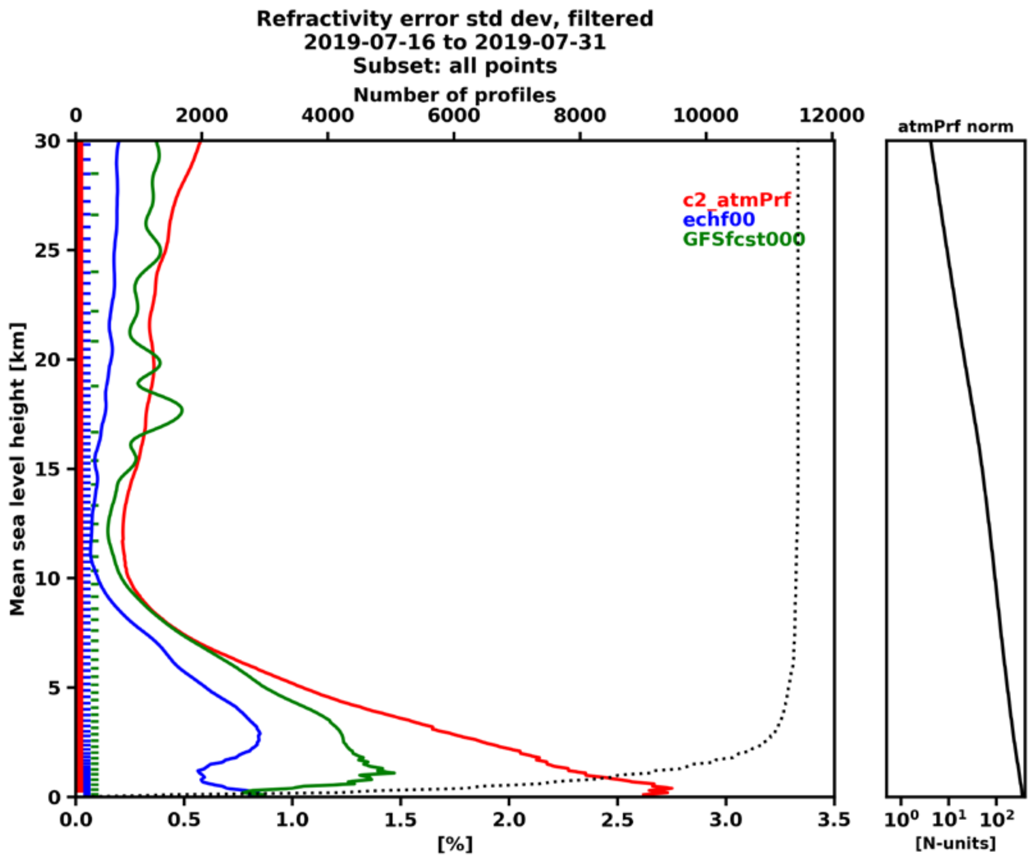
  - Mean square deviation

$$E\left[(X_n - Y_n)^2\right] = (b_X - b_Y)^2 + \mathrm{Var}\left[\varepsilon_{X,n}\right] + \mathrm{Var}\left[\varepsilon_{Y,n}\right] - 2\mathrm{Cov}\left[\varepsilon_{X,n}, \varepsilon_{Y,n}\right]$$

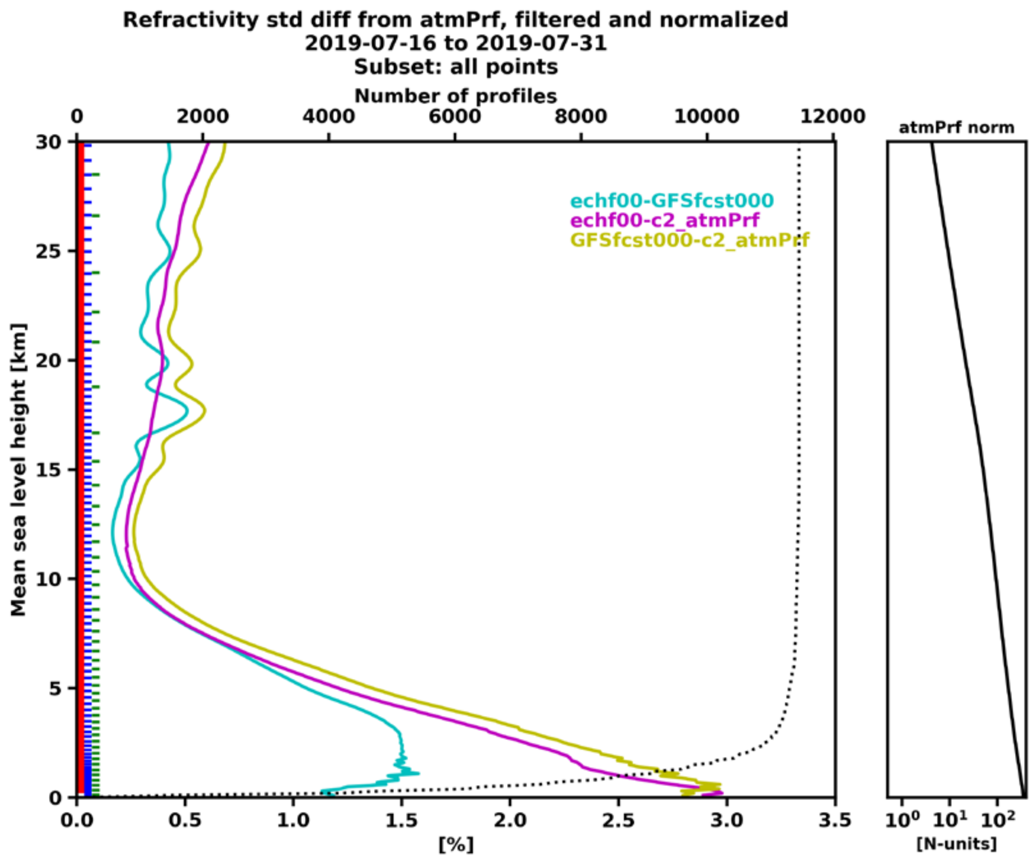
Initial COSMIC-2 analysis: refractivity

C2 atmPrf + EC analysis + GFS analysis

3CH estimates



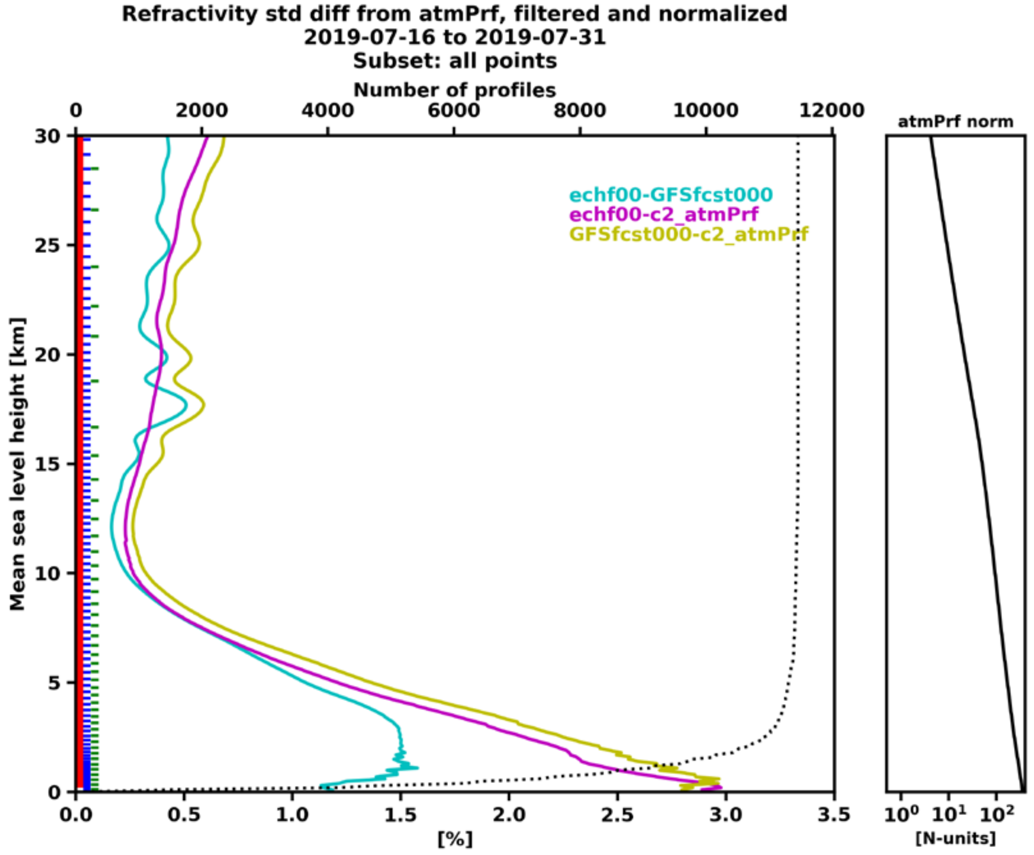
Std dev of differences



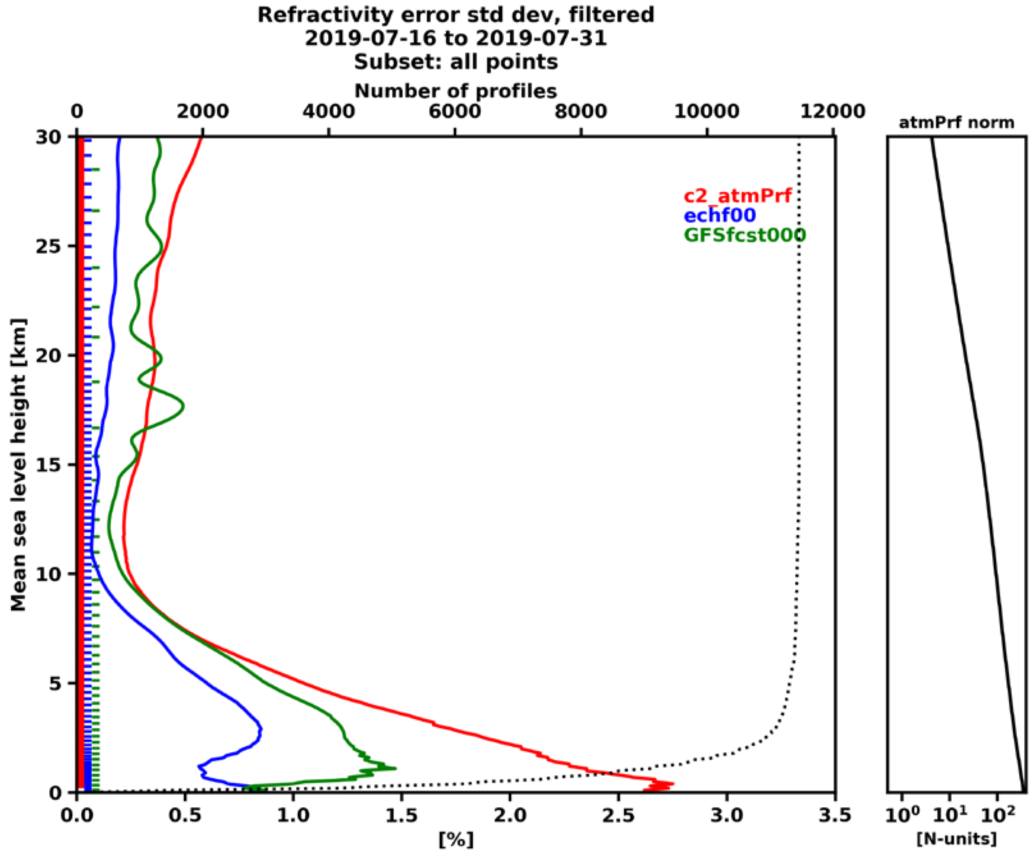
Initial COSMIC-2 analysis: refractivity

C2 atmPrf + EC analysis + GFS analysis

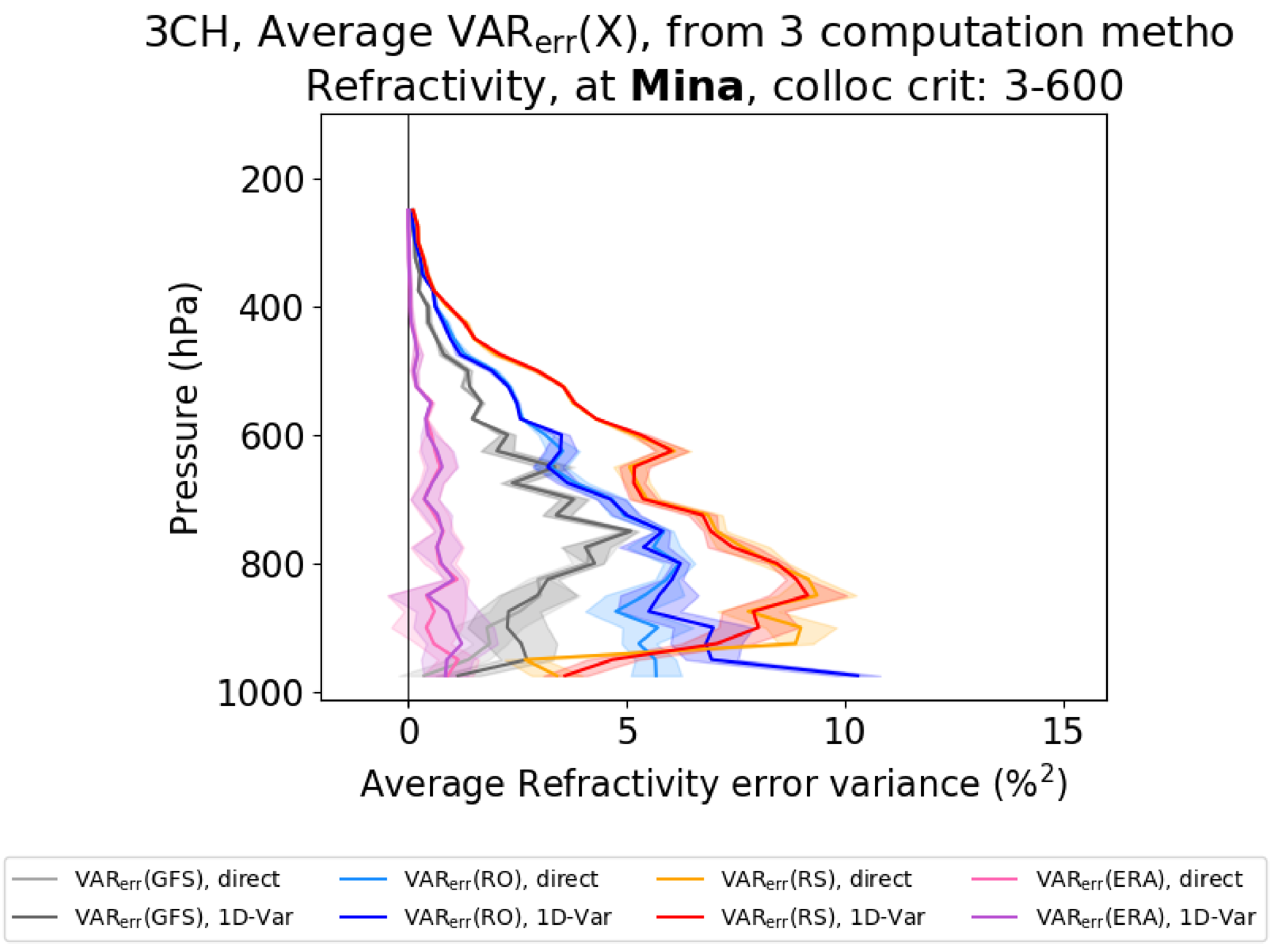
Std dev of differences



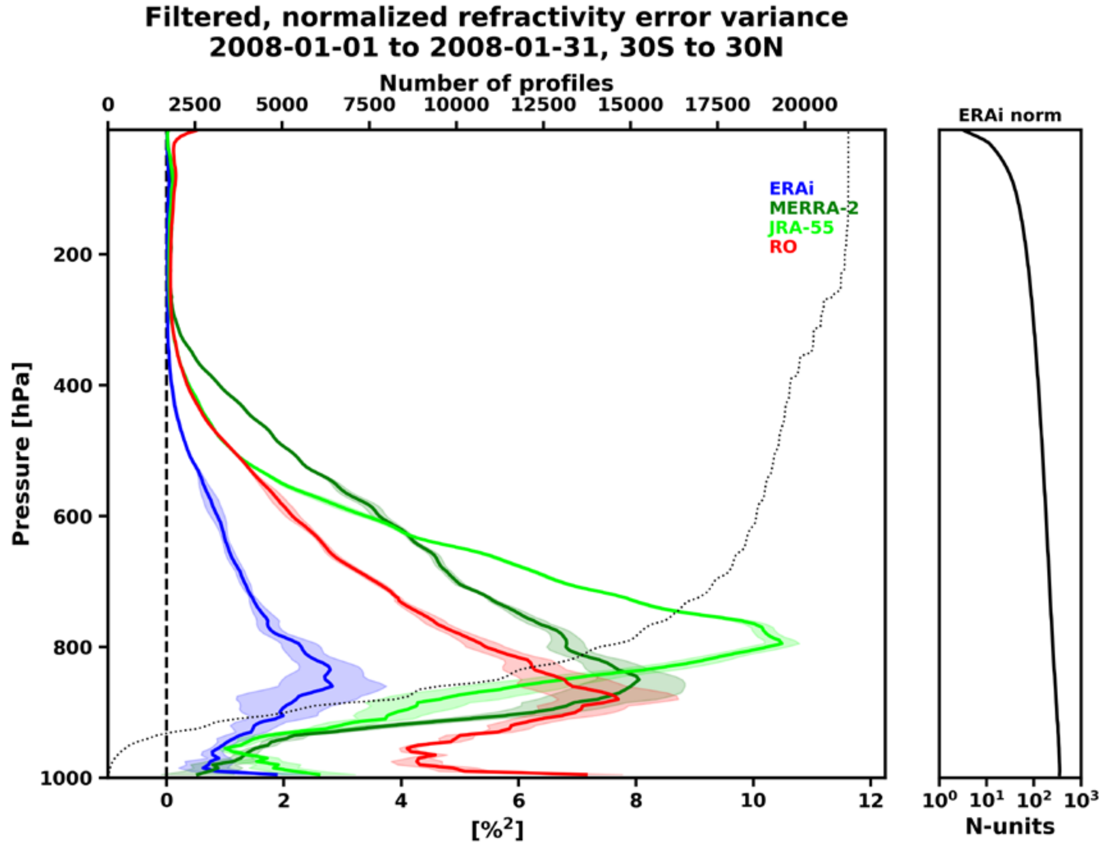
3CH estimates



COSMIC-1 results

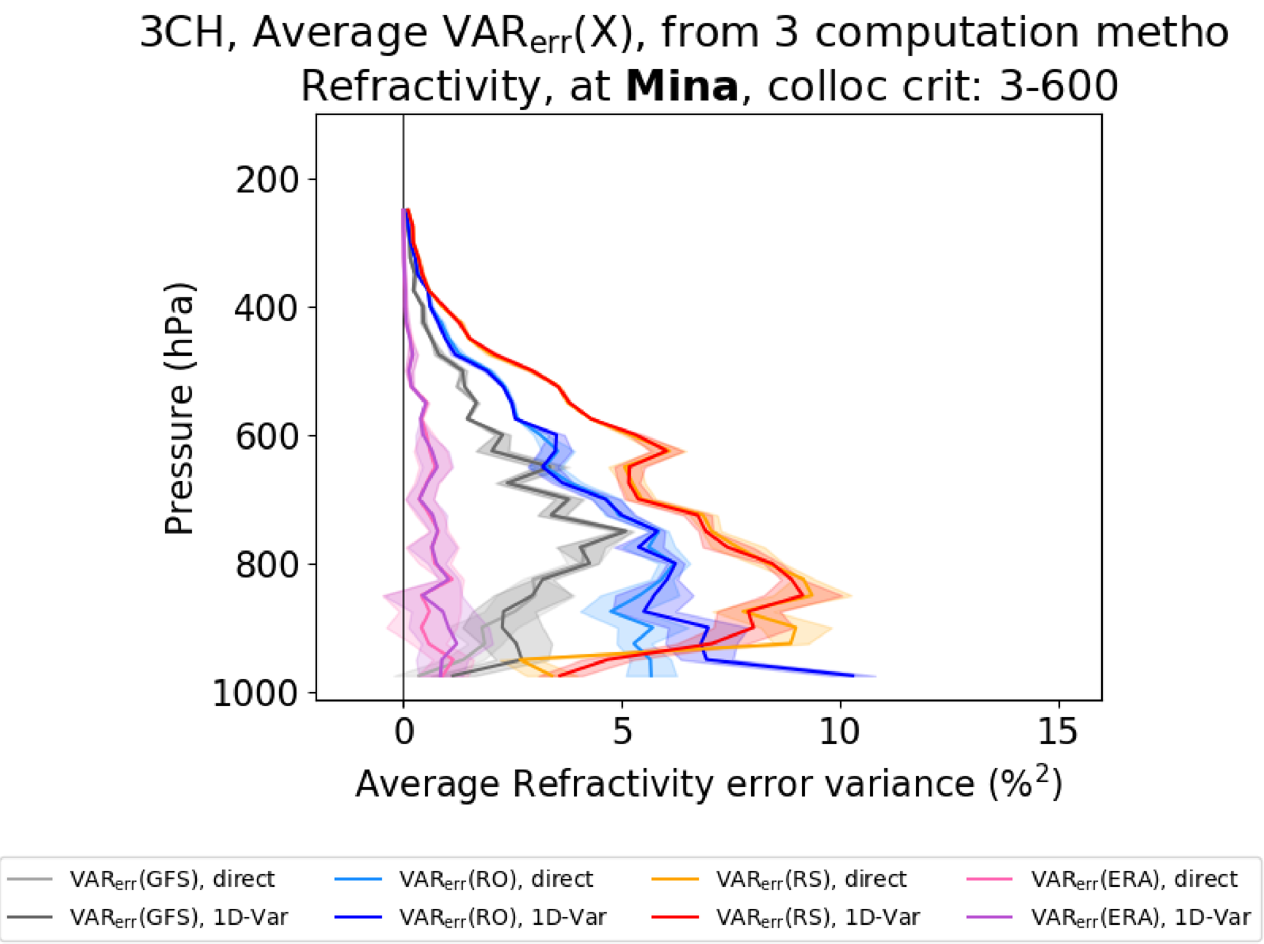


Anthes and Rieckh (2018),
DOI:10.5194/amt-11-4239-2018

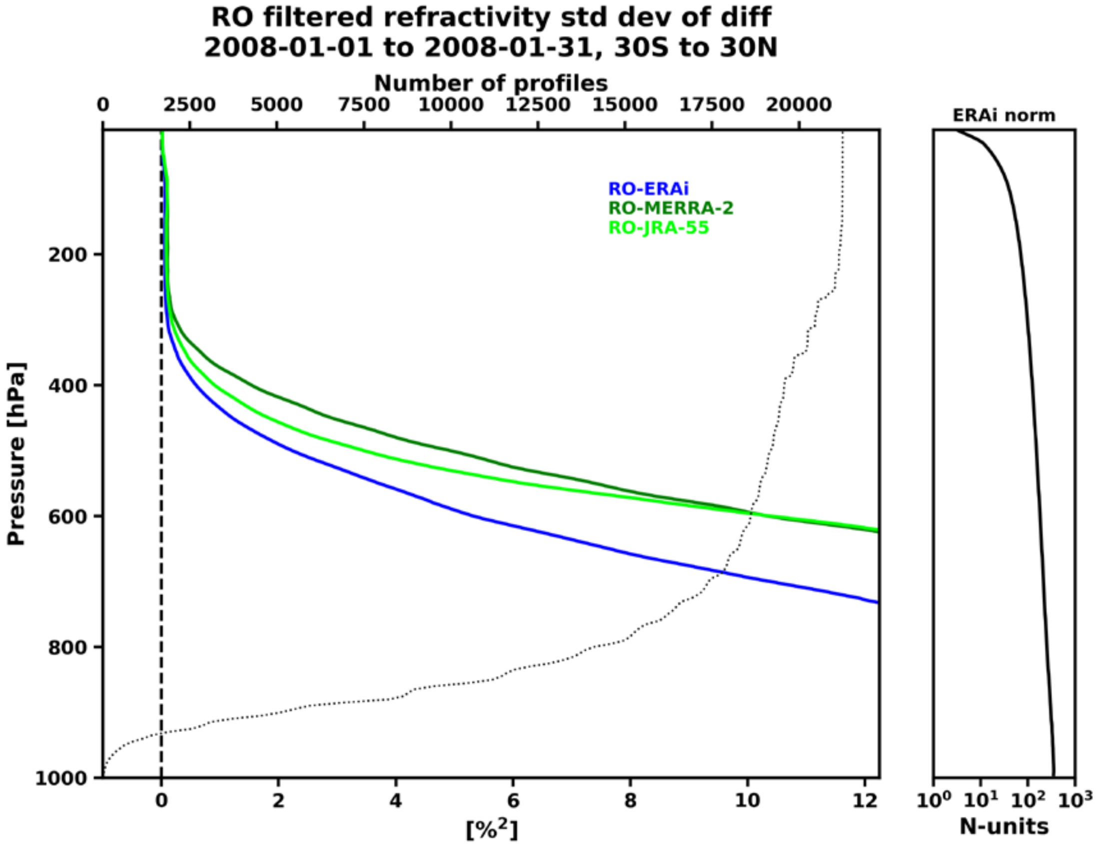


Normalized error standard deviations comparable to (O-B)/B relationship

COSMIC-1 results



Anthes and Rieckh (2018),
DOI:10.5194/amt-11-4239-2018



Normalized error standard deviations comparable to (O-B)/B relationship

COSMIC-1 results

C1 wetPrf (unaltered refractivity) + EC analysis + GFS analysis