Issue 4: Reproducibility of temperature results

The comparison of temperature retrievals from four different centers showed very similar changes over time. This is a good sign that the key measure of temperature was useful for climate change purposes. But the differences suggest that retrieved temperature should not be considered a benchmark observation. Analysis quantifying the differences (both average difference (bias) and absolute value of the differences) by latitude bands and within latitude bands over land and ocean separately would help clarify the appropriate role of GPS-RO temperature retrievals in climate analyses.

a. Define the question

Requests raised from this issue are based on slide 9 of Ho et al. (AMS COSMIC presentation, 2008), which is re-plotted in **Fig. 1**. This plot **(Fig. 1)** is actually for fractional refractivity comparison (not comparison of temperature retrievals) from four centers. Again, Since an operational error was found in GPS RO data from Wegener Center of the University of Graz (WegC), Graz, Austria (it causes time constant systematic error and it will not be seen in a trend analysis), here only RO data from UCAR, JPL and GFZ are plotted in **Fig. 1**. Because pixel level RO profiles from WegC and JPL will not be available until early March, only GFZ and UCAR RO data are compared here. We leave the comparison of pixel level RO data from all four centers in a future study when all data are available. To make consistent comparisons as slide 9 (and other related slides) of Ho et al. (AMS COSMIC presentation, 2008), here we conduct pixel level fractional refractivity comparisons. Temperature difference is about 1.5 to 2 times larger (dependent on different height) than that for fractional refractivity difference (0.01% for fractional refractivity difference is about 0.015-0.02 K temperature difference), which is also leaving for a future study.

b. GFZ and UCAR pixel-level fractional refractivity comparison

GFZ refractivity profiles from Jan. 2006 to Dec. 2006 are paired with those of UCAR. More than 20000 pairs of GFZ and UCAR refractivity profiles are compared (**Fig. 2**). Mean bias of 100*(GFZ N-UCAR N)/UCAR N (%) from 8 km to 30 km is equal to 0.03%, where their mean MAD from 8 km to 30 km is about 0.16 % (**Fig. 2**).

c. Mean fractional refractivity difference between GFZ and UCAR at different latitudinal bands

Contour plots for mean fractional refractivity difference, sample numbers, and MAD for 10 degree latitude bands are generated in **Figs. 3a, b, and c**, respectively. It is shown in **Fig. 3** that the mean GFZ is very consistent with that of UCAR (within 0.1%, see **Fig. 3a**; with 0.2% MAD, see **Fig. 3c**). Larger mean biases and MAD are found above 25 km (mean bias is from 0.1% to 0.3% and

MAD is from 0.2% to 0.4%) and below 10 km (mean bias is from 0.1 % to 0.2% and MAD is from 0.2% to 0.3%).

Similar conclusions can also be found for land-only contour plots (**Fig. 4**) and ocean-only contour plots (**Fig. 5**).

d. Time series of fractional refractivity difference between GFZ and UCAR at different vertical layers for different latitudinal zones

To quantify if RO data is suited for climate study, here we examine the "reproducibility (consistency)" of GFZ and UCAR refractivity by generating the time series of monthly mean of the mean fractional refractivity differences for GFZ-UCAR pairs for different latitudinal zones (90°N-90°S, 60°N-90°N, 20°N-60°N, 20°N-20°S, 20°S-60°S, 60°S-90°S) with different vertical layers (8-30 km, 8-12 km, 12-20 km and 20-30 km) in **Figs. 6, 7, 8 and 9**, respectively. These plots are binned similarly with those plots shown in response to issue 3 (**Figs. 2, 4, 6 and 8**), except that sampling errors shall be largely eliminated here.

Figs. 6, 7, 8 and 9 show that time series of GFZ refractivity profiles are highly consistent with those of UCAR. The mean fractional refractivity differences for GFZ-UCAR pairs are almost constant for different months for 2006 for different height and for different latitudinal zones (**Figs. 6-9**). The mean difference and its standard difference for each region at different layers are summarized in **Table 1**. In general, the mean global bias is about 0.11 % for 8-12 km layer and about 0.053 % for 20-20 km layer (see **Table 1** and **Fig. 6**), yet the global variations of the mean (standard deviation) for different layers are all within 0. 02 %.

e. Trend of the best fit of GFZ-UCAR fractional refractivity difference for 2006

The trends of the best fit for GFZ-UCAR fractional refractivity anomalies are computed (in green line) and are also summarized in **Table 1**, where trends larger than 0.1% are in red. Because CHAMP data for most of July 2006 are not available, we usually have a larger bias in July at different layers for each region. Much smaller sample over polar regions (60°S-90°S and 60°N-90°N) also contribute to larger biases than other months at different layers. These lead to larger trend value especially over polar regions (trend values in red in **Table 1**). Different ionospheric calibration methods may also contribute the larger biases over polar regions in 20-30 km layer (**Figs. 9 b and f**). Note that these trends are computed from time series for only 12 months. Systematic seasonal biases shall be removed when multi-year of data are used. This will be examined for using multi-year of GFZ and UCAR data.

The fact that the variation of the mean fractional refractivity differences in **Figs. 6, 7, 8 and 9** are much smaller than those in **Figs. 2, 4, 6 and 8** for issue 3 demonstrates the reproducibility of GPS RO data when sampling errors are largely eliminated.