EDGES MEMO #211

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886

August 18, 2016

Telephone: 781-981-5414
Fax: 781-981-0590

To: EDGES Group

From: Alan E.E. Rogers

Subject: Tests of fitting, filtering and averaging schemes.

1] Introduction and motivation

While EDGES data is well calibrated there are many reasons for filtering, fitting and averaging the data on different time scales in various ways. In the scheme developed at Haystack there are 3 processing steps:

1] acqplot

2] edges2

3] longav

In the first step the raw data is processed to obtain an uncalibrated spectrum from each 3-position switch cycle. Each cycle can be included or excluded based on its time, deviations of the spectra of the antenna, load and load plus cal from the expected values due to strong RFI or equipment malfunction. In addition, individual spectral points can be flagged for exclusion from the average spectrum in the output of acqplot. The spectral weights in the average uncalibrated spectrum from acqplot are obtained by an iterative fitting using a high order polynomial to allow the most sensitive detection of narrowband RFI. For the best results at least one hour of data is normally used. The model is added back in so that the output is an uncalibrated spectrum.

The second stage processing in edges2 includes the calibration and beam correction. The beam correction is done on a time scale of 30 minutes centered at the average time of the uncalibrated spectrum from acqplot.

The third stage processing in longav performs averaging of data files from edges2 and allows exclusion of these input files based on the rms residuals to an rms fit of a given number of terms of different types (physical or polynomial) and weighting of individual frequency channels. At this point there are many choices. Normally acqplot would smooth the spectrum (of 6.1 kHz resolution) over 8 points (using -smooth -8) to 48.8 kHz resolution and edges2 smooths its output (using –smooth -8) to 391 kHz resolution which is still fine enough that weak digital TV RFI can be detected and filtered at the third stage of processing in longav.

For processing months of data there are a number of choices from a fine time scale in which data from each day is retained in 24 one hour averages from LST or GHA from 0 to 23 hours. This time scale for 300 days of data has 7200 spectra. To get the best “Galaxy down” spectra one could

a) Obtain 1 hour spectra from each day for GHA 7, 8, 9, 10, 11, 12, 13 from acqplot then process each in edges2 and average the 2100 spectra in longav
b) Obtain 7 hour spectra from acqplot and process through edges2 and average 300 spectra in longav.

These 2 schemes should give identical results if there is no RFI excision or down weighting of frequency channels but in practice they will give slightly different final results.

In averaging it is useful to recognize that the residuals to the average of all the spectra is the same as the average of residuals to the individual spectra. This is mathematically true as long fitting functions and weights are the same in each case. In the presence of RFI or bad data due to weather or other conditions the best find calibrated spectrum for a long period needs to be filtered on all time scales. A very short time scale is needed to remove transient RFI and longer time scales are needed to detect RFI which is very weak but is present in the data on longer time scales. For a constant RFI signal present for a fixed time scale the ability to detect the RFI is to average at the same time scale. While averaging for 7 hours each day in the first stage and then averaging the spectra from each day it fails to cover time scales of a few hours and time scales of serval days.

Comparison tests:

Test	GHA each day	Time hrs	# spectra	
A	7, 8, 9, 10, 11, 12, 13	1	7	1513
B	10	7	1	228

Table 1. Nominal day range 2015:286 to 2016:199 Beam correction GF 3.5 2e-2

The first comparison is between “Case A” in which separate spectra are produced for each 1 hour integration of each day and “Case B” in which a single spectrum is produced for a 7 hour integration each day. The residual 5 physical terms and difference or residuals are shown in Figure 1 for both day and nighttime data and in Figure 2 for nighttime data only. The differences are small and fairly free of systematics. For comparison figures 3 and 4 show the difference for a small change in assumed soil dielectric and a change to an infinite ground plane. In both cases A and B the processing time is dominated by acqplot which takes about 2 seconds per hour of realtime data. Second stage processing with beam correction takes about ¼ second per hour of realtime data and the third stage takes a negligible amount of time and doesn’t depend on the amount of realtime because it operates on averages. The second stage only depends on realtime which beam correction is applied. This dependence could be eliminated using tabular storage of beam corrections which repeat with LST.

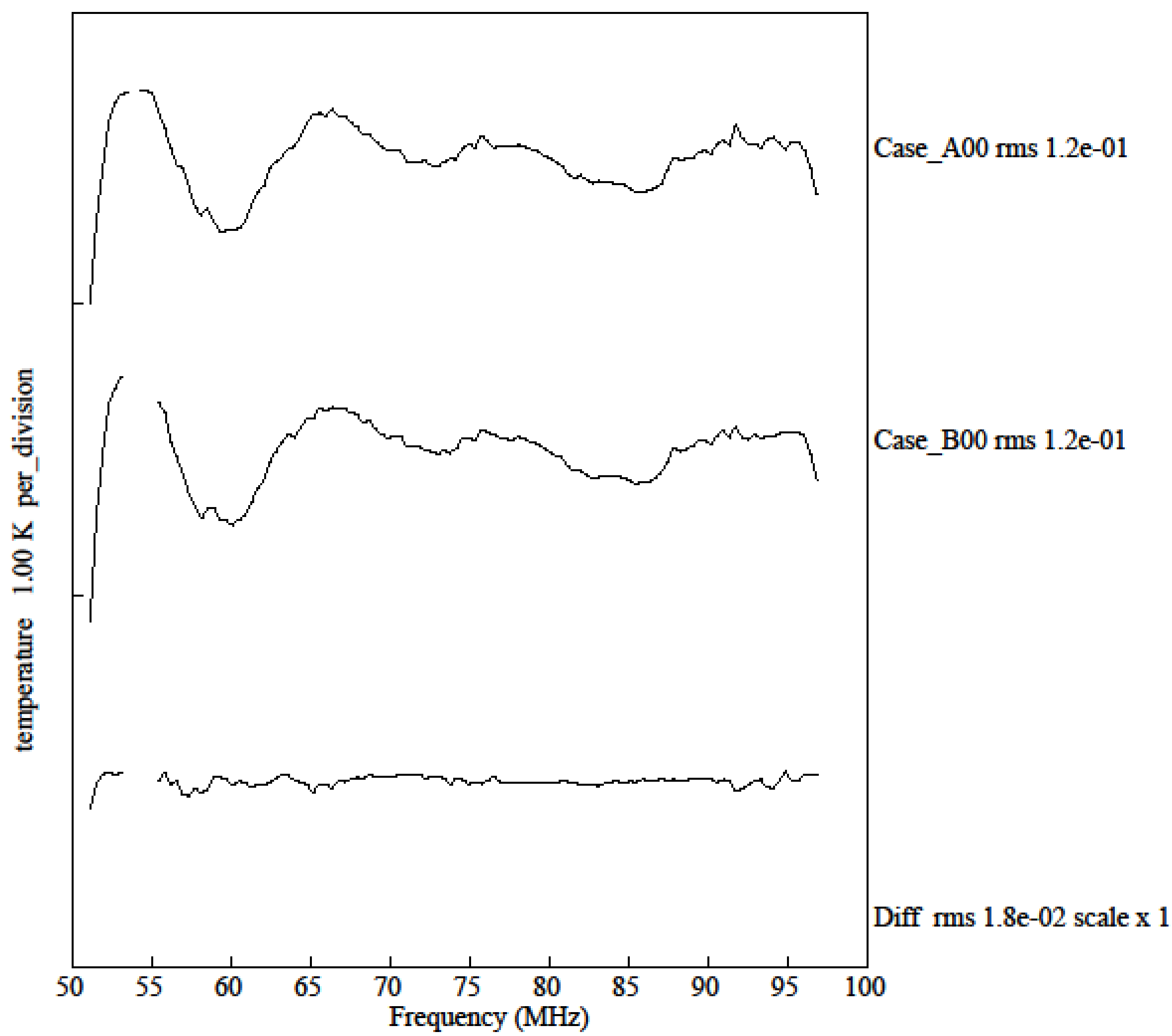

Figure 1. Comparison of Case A and Case B for day and nighttime.

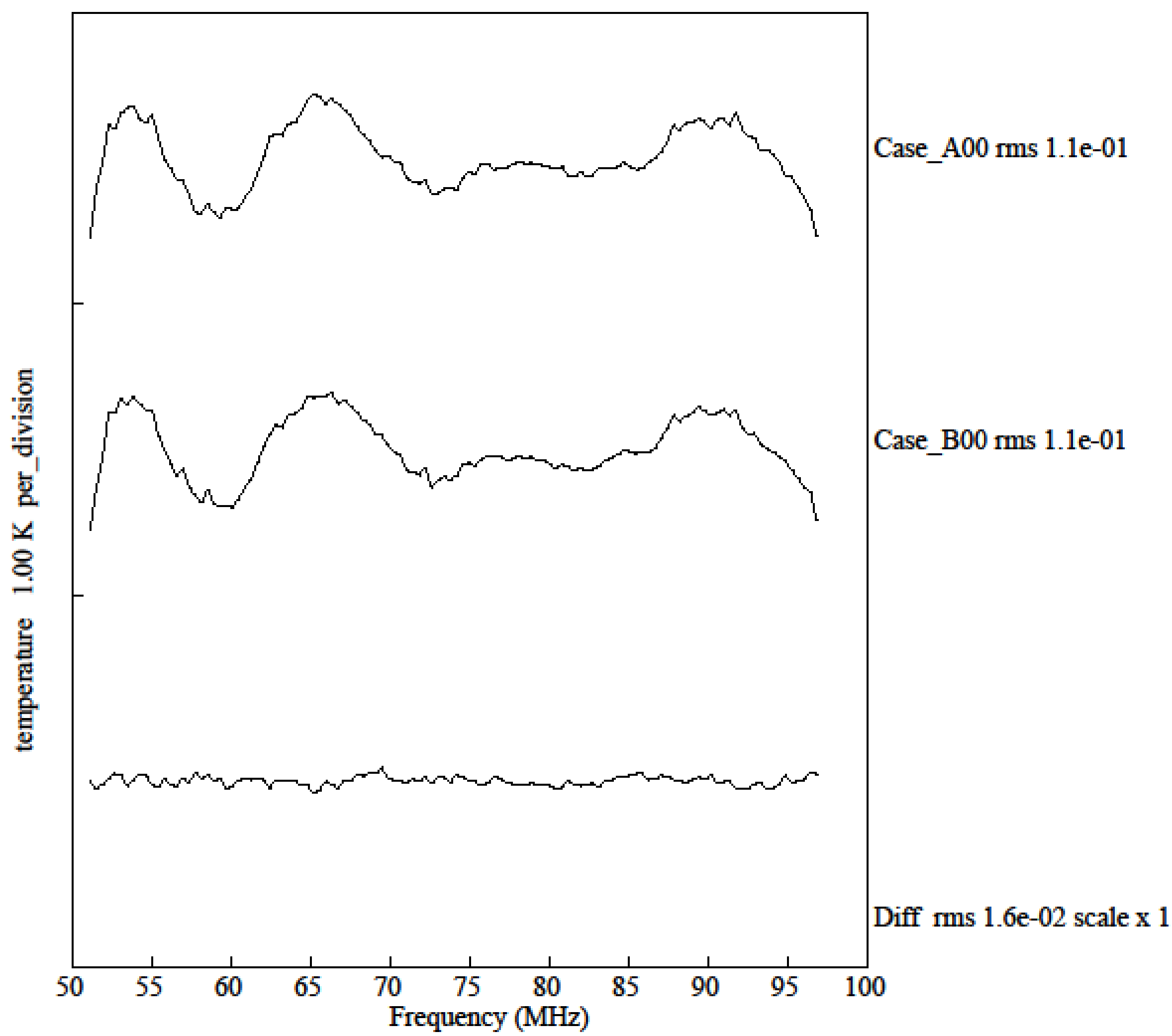

Figure 2. Nighttime only

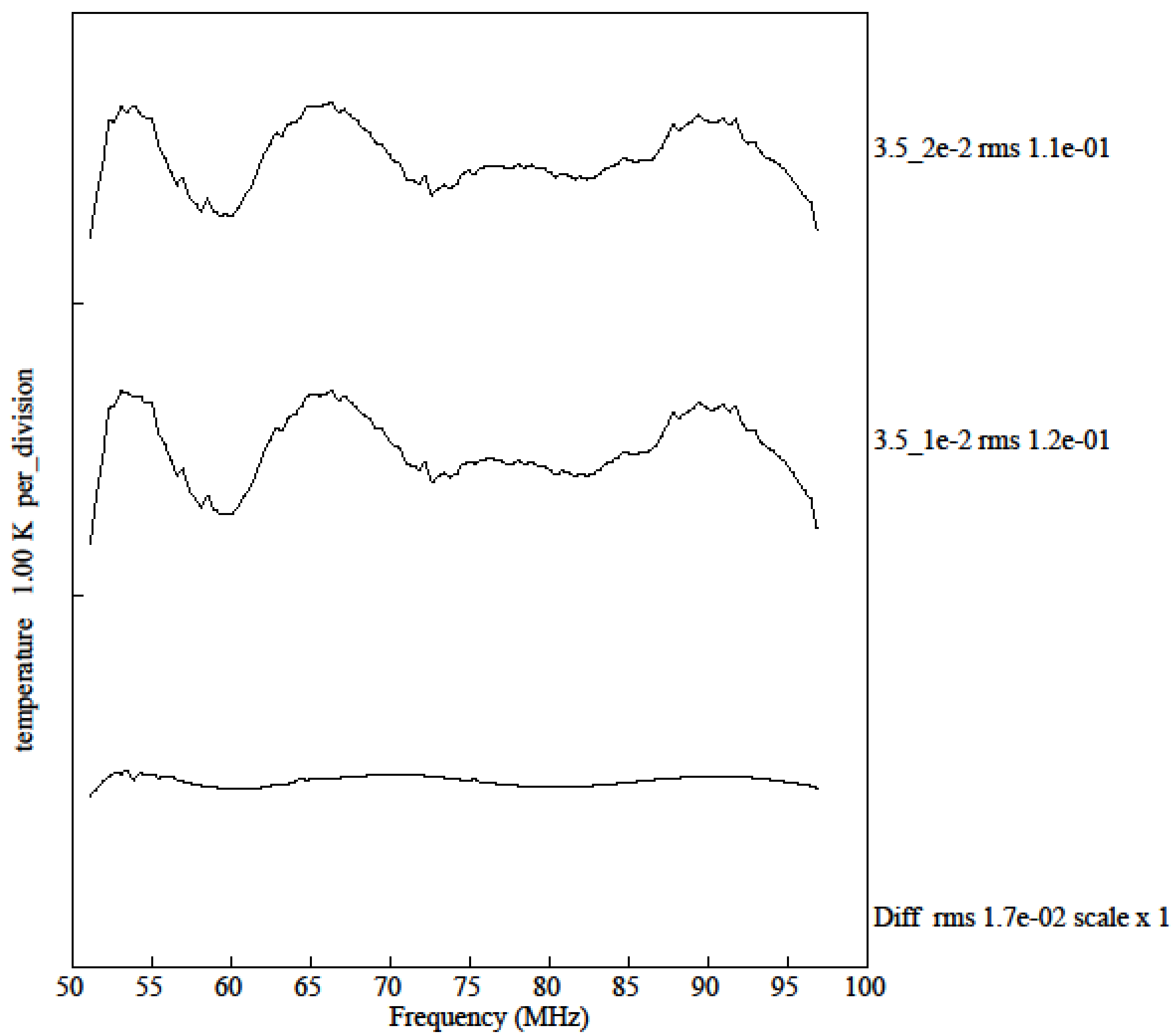

Figure 3. Difference between residuals using beam correction with different soil conductivity.

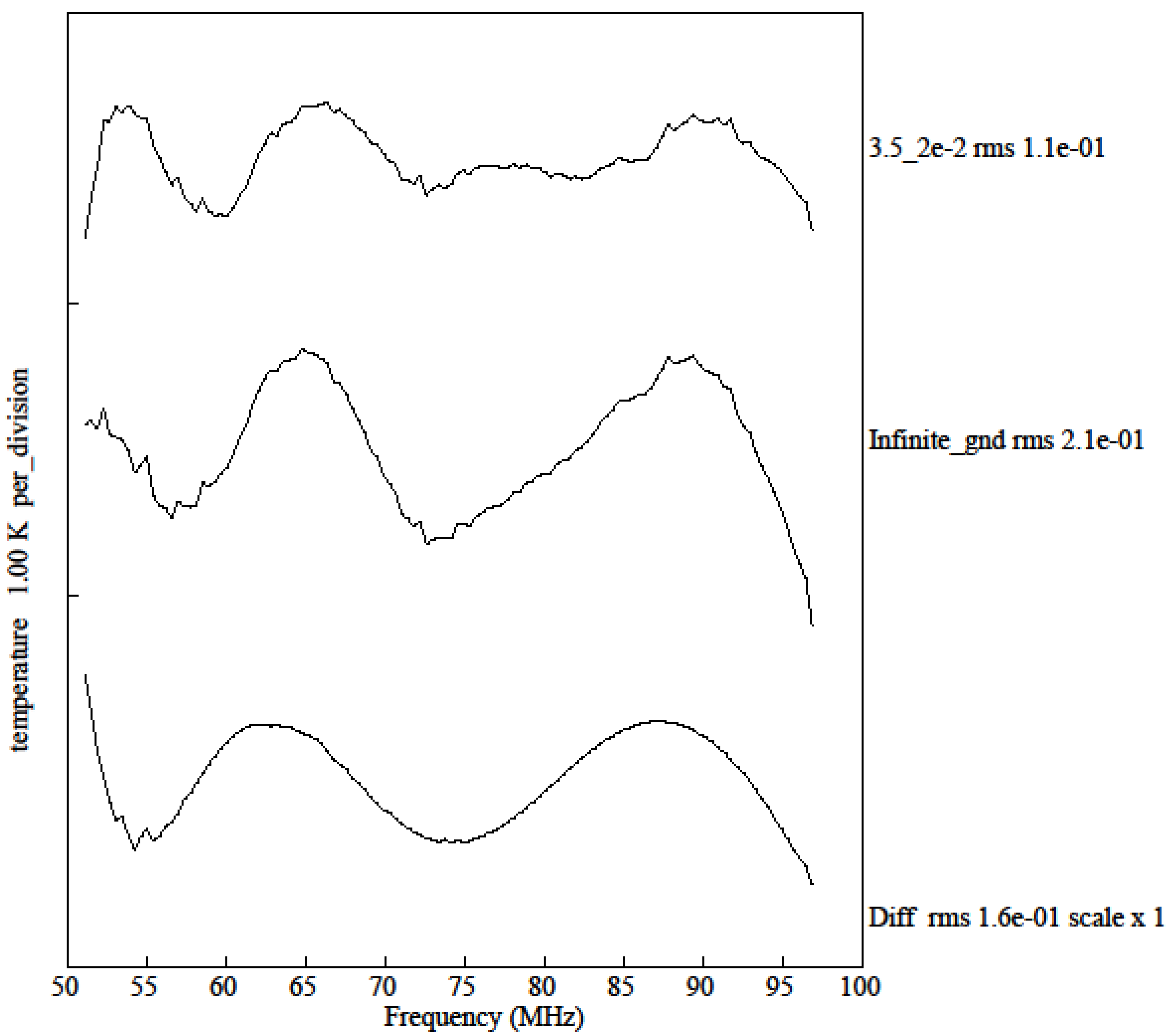

Figure 4. Difference between residuals using finite and infinite ground plane.