PAPER Memo 16: Selfcalibration and Imaging of PAPER data using standard tools

Chris Carilli, October 14, 2011

1 Introduction

I have performed standard wide field image processing (flagging, imaging, selfcalibration) of test data from the PAPER array in South Africa using the tools in AIPS.

The data involved six snapshots of 10min each from the 64 dipole element arrays in SA (Parsons ea. 2010, AJ, 139, 1468), observed in July 2011. The maximum baseline is around 250m. The data has 2048 channels of 48.8kHz from 100 to 200 MHz. The snapshots included Cen A in 4, and the Galactic plane in 2. The record length is 10.73sec.

Initial editing, calibration, and phase tracking of mean zenith over 10min was done using a script in casapy + AIPY to generate a measurement set. Exportuvfits was then used to generate a FITS file. The data were loaded into AIPS, and header altered to make it work (stokes = -1, remove SU table).

Figure 1 shows the UV coverage for a single channel, then for the full band.

2 Initial flagging

Channels below 250 and above 1850 were flagged at the start.

A scalar averaged spectrum (summing visibility amplitudes) of all the data was generated for 100 channels at a time (POSSM), and bad spectral ranges were flagged using UVFLG. Figure 2 shows an example for the full range, before and after flagging and subsequent calibration. This process removes the strong RFI, such as at 136MHz (channels around 760), and other narrow spectral regions. This is perhaps the most user-intensive part of the reduction, but might be automated as part of the next steps.

The data were then copied to a series of 200 channel (10MHz) files, which were then processed separately. The copied data were clipped at a high level to remove remaining bad data, determined from a standard UVPLOT. Note that the clip level depends on frequency.

A second level of clipping was then done, after subtracting the continuum emission in the visibilities using UVLIN. This is a very effective step for flagging, but again is frequency dependent.

Figure 3 shows plots of the UV amplitude versus baseline length before and after the clipping stages, for a single channel.

3 Self-calibration

The tasks IMAGR and CALIB were used for all imaging and calibration.

First, a wide field image is generated using a large cell size to search for outlier sources. For instance, in one data set Cygnus A is seen fairly strongly, even though it is 73deg from zenith.

The self-calibration entailed making an image of the central field of 1024 x 200" pixels (56deg), plus small outlier fields with strong sources, using the 3D capabilities for the outlier facets in IMAGR.

I experimented with UV weighting via the Robust parameter, and settled on R=1. The FWHM of the synthesized beam for the zenith facet varied from 23' to 15' from 120MHz to 180MHz. I also experimented with UVMIN in the data, and generated images with no UVMIN and an 8 wavelength min. The latter produce better looking maps (less large-scale undulations), but miss some large scale structure.

Initial calibration from AIPY was adequate to generate a starting model image from the data. This was done by CLEANing 1000 iterations. The CC model was then fed into CALIB, using phase-only selfcal with a 1min averaging time. Figure 4 shows a plot of the antenna-based phase solutions versus time over the 10min. There can be offsets of up to 20deg, or smooth gratients, but generally the phases were stable at the few degree level over 10min. The offsets or phase slopes could relate to antenna positions, but that remains open.

For reference, equation 13-8 in Perley (SIRA II) states that the snapshot image dynamic range for random antenna-based amplitude errors is $\sim N/(1.4\phi)$, where N is the number of antennas and ϕ is the rms error in radians. For a 3^o to 5^o rms error, the expected dynamic range is 500 to 900, which is about what we are getting on the images below.

The self-calibrated data were then imaged as above, generating a new clean component model. This model was then used for the second iteration of self-calibration, with the input data being the output from the first iteration. This second iteration involved phase and amplitude calibration with a 3min averaging time. An example of the solutions are shown in Figure 5, with shorter averaging time to see variations. Again, there can be offsets of 20% or so, but the solutions were generally stable to a few percent over 10min.

Antenna-based gain solutions from selfcal allows for important diagnostics on bad data, and subsequent flagging. From the SN tables, the antennas: 32, 40, 55, 56 were seen as either really bad (dead) or more noisy than the rest, and flagged (note: aips starts at antenna 1, so AIPY antenna 0 = AIPS antenna 1). Calibration solutions were then rederived, and a final, clean and calibrated data set generated.

Note that this process can be repeated an nauseam to improve dynamic range as the model gets better, but for these 10min data sets, it is probably not worth the effort – there are other factors that may be dominating already, such as bandpass/MFS, multiscale, etc...

4 Final imaging

The task SETFC was used in aips to determine the facets required to generate a 3D image of the primary beam. This was run using a cell size of 200" and 256x256 pixel facets, over an area of 46^o radius, with allowed overlap of 30 pixels, resulting in 61 facets. Initial iterations showed significant problems at facet edges, depending on input parameters, with facet edges either not covering the full area, or being visible in the final image. This is a tricky business, and using EDGSKP = 3 in the subsequent FLATN step to add all the facets into a final image mitigated this effect (but not completely).

IMAGR was then run, including any far-out facets as well, using 2000 clean components and a loop gain of 0.1, and overlap = 2 for the facets. All the channels were averaged for each sub-dataset (200 channels = 10MHz).

I should note that the imaging was slow on a high-end workstation (dual-quad core processor). A single 10min dataset, for the full frequency range (8 sets of 200 channels) takes about 2 hours for the full imaging and deconvolution. The process is highly parallelizable (channels could all be done in parallel!), so CASA might go 8x faster, once it is parallelized.

IMAGR generates 61 independent facet images, and then FLATN is used to add all the facets together onto a single image with the correct geometry and weighting in overlap regions.

The final 8 images for a given snapshot (ie. made from 200 channels each) were then added with equal weights. The noise gets higher with decreasing frequency, but the sources get stronger so the signal to noise is about the same.

5 Results

The final widefield images are shown in Figure 6-12. These are screenshots of the images with no UVMIN. The typical dynamic range (peak to rms) is 800. Note that the absolute flux scale is not set. Sources very far from the zenith remain difficult to image, and artefacts from Virgo and Cygnus can be seen.

Figure 13 shows a summed image of Cen A, plus a comparison to single dish imaging. There may be extended structure to the southeast well beyond the classical boundaries of the source. Cen A is huge, and the large negative bowl in which it sits indicates that our interferometry is still not dealing with the very large scale ($>$ 10deg) structure correctly.

Figure 14 shows an image of the Galactic plane, plus the central regions from the VLA 74MHz study of La Rosa ea. (2005, ApJ, 626, L23). Figure 15 shows a blow up of the Galactic center region at 120MHz and 170MHz, plus the IRAS dust image.

Diffuse emission is clear, plus the Galactic center and numerous SNR. We see what appear to be regions of absorption along the Galactic plane. This could be, again, short spacing issues, but it appears to be frequency dependent, ie. deeper at low frequency. These features could be Free-Free absorption by Galactic HII regions. Some of these regions have already been identied in VLA observations of this region at 74MHz. Note that negative regions can occur in interferometric images, if the background continuum is much larger than the absorption regions.

For reference, the Free-Free optical depth is given by:

$$\tau = 8.2 \times 10^{-2} \nu^{-2.1} \mathrm{T_e^{-1.35} EM}$$

where T is in K (assumed 10^4K), ν in GHz, and the emission measure, EM, in pc cm^{-6}. For $\tau > 1$ at 160MHz implies EM $= 6.5 \times 10^4$ pc cm^{-6}. Our spatial resolution of 15' implies a physical scale at the Galactic center of 40pc, comparable to a typical GMC size. An electron density of just 40 cm^{-3} in such a region would be adequate to be optically thick at 160MHz.

Figure 16 shows the final spectrum of Centaurus A over the full band, and a select 10MHz band. There are clear undulations in the band at the 10% level on scales of a few MHz. Note that I have not done any frequency dependent (channel to channel) calibration. The issue of bandpass calibration (gain vs frequency) is hyper-critical.

I have tried to get a rough estimate of the flux scale from CenA, using the integrated emission. From the literature, at 160MHz we expect about

5200Jy. In this case, the rms on the images is about 2Jy to 3Jy. For reference, the flux density of Cen A is collected in:

http://ned.ipac.caltech.edu/level5/March01/Israel/Israel2.html

Adopting this calibration scale, I perform channel differencing at 48kHz resolution to get a measure of the true noise (ie. not dynamic range limited). The difference images do look very much like noise. Adopting a collecting area of 73000 cm^2, I derive a Tsys of about 1000 K. This may be expected in these regions of the sky, not well away from the Galactic plane.

I note that the geometry over these very wide fields gets strange, and source positions aren't quite right from one field to the next. I am not sure what is going on, but am investigating.

6 What next?

Just these few datasets indicate both the richness of the observations, and the extensive task ahead to achieve high dynamic range imaging. I have done the basic wide field, low frequency processing that has been developed in AIPS over the last decade or two, and get to a dynamic range of close to 1000. Note that for the northern fields with Cygnus A, I was getting a factor 10 higher dynamic range with PGB32, probably because the calibration could be done using Cygnus A, which is unresolved and dominates the visibilities. I have not exercised all the options, and following is a list of areas that need to be explored.

The processing has also shown that our imaging data has tons of anscillary science that can be explored, almost immediately.

Technical:

- Flux scale: we need to determine the absolute flux scale versus frequency. This is not trivial, and couples to the primary beam model, unless a source is at zenith. We have a program at the GMRT to determine some absolute flux calibrators over the relevant frequency range in the Southern Sky. We would like to get to a few percent flux calibration, but that remains to be seen.
- Bandpass calibration: it remains unclear how to calibrate the bandpass on the sky (gain vs. frequency). This also couples to the flux scale calibration and the primary beam model. We need to be able to 'flatten' spectra over at least 10MHz to a high level of accuracy. Any systematics could affect the spectral statistics. I currently have

no idea how to proceed here. I have tried using Cen A as a start, but BPASS bombed.

- Multifrequency synthesis: the imaging might be much more effective if we could employ multifrequency synthesis to correct for at least slopes over a 10MHz band, but better yet to be able to properly image the full 60MHz band in one go. Right now the imaging simply averages up the 200 channels in a sub-dataset. Tools exist in AIPS and CASA to invoke at least a powerlaw spectrum for sources over the band, but it is slow.

- Multiscale clean: tools also exist to perform multiscale clean. I think this could be a big gain, given that we have structures on scales of unresolved to 10 degrees (or more). I have tried this once, and the book-keeping became painful. In general, exploring parameter space in terms of data weighting and deconvolution (eg. maxEntropy) is fairly wide-open.

- Direction dependent gains: I have tried PEELing on previous datasets, and it did not perform much better or worse than standard selfcal (ie. position independent gains). I think we are dominated by other effects well before the ionosphere kicks-in with these data. However, we could investigate whether sources at the very edge of the beam (eg. Cygnus A) scintillate or do other funny things.

- I have run into a problem summing images, in terms of getting the geometry correct using HGEOM. The sources appear at different RA and Dec depending on the LST of the observation. I have a query into Greisen about this.

- Primary beam: this couples to many of the phenomena above.

- Polarization: not clear where to begin.

- Pipelining: right now the processing was intensively interactive, until the big imaging step at the end. I can see a way to doing much of this as a pipeline, with a few break points to check eg. flagging parameters. AIPS is unweildy in this regard, but possible. CASA may be better.

- Processing time: again, this seems to be many times longer than real-time. Parallelization might help, and I have some ideas for tests along these lines, but this may be problematic right now.

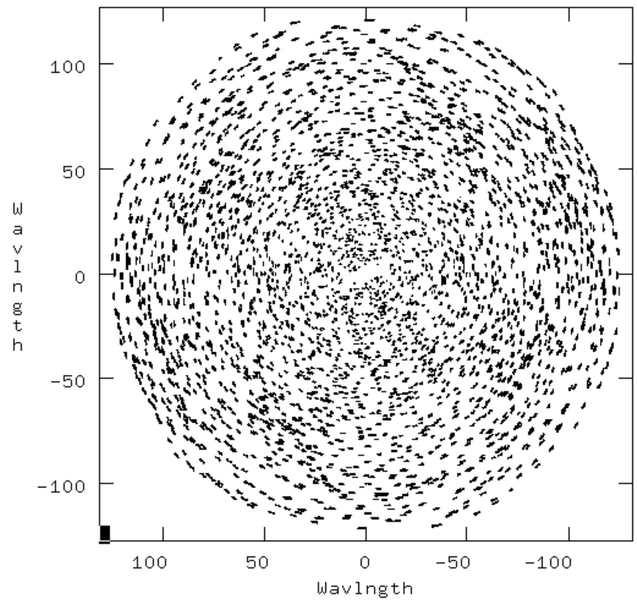
Science:

- Galactic: We explore a unique parameter space in terms of areal coverage, frequency coverage, spatial resolution, and sensitivity. The PAPER data are excellent for supernova remnant studies, in terms of structure, spectra, and finding new, big SNR (Brogan ea. 2006, ApJ, 639, L25; see the D. Green catalog): http://www.mrao.cam.ac.uk/surveys/snrs/

- Galactic: There is already evidence for Free-Free absorption by Galactic HII regions. Calibrated spectra could provide good measures of the opacity, and hence emission measure. Comparison to the non-thermal emission also relates to the 3D geometry (La Rosa ea. 2005, ApJ, 626, L23).

- Centaurus A: PAPER yields a unique view of the very large scale structure of the nearest radio galaxy, Centaurus A. We may be seeing even larger structure than has been seen prior, and our spectral study is unique (Feain, I. ea. 2011, ApJ, 740, 17).

- Transients: we observe the full sky contiuously. This is a truly unique data set for studies of transients at the mJy level at low frequency on timescales of minutes to months.



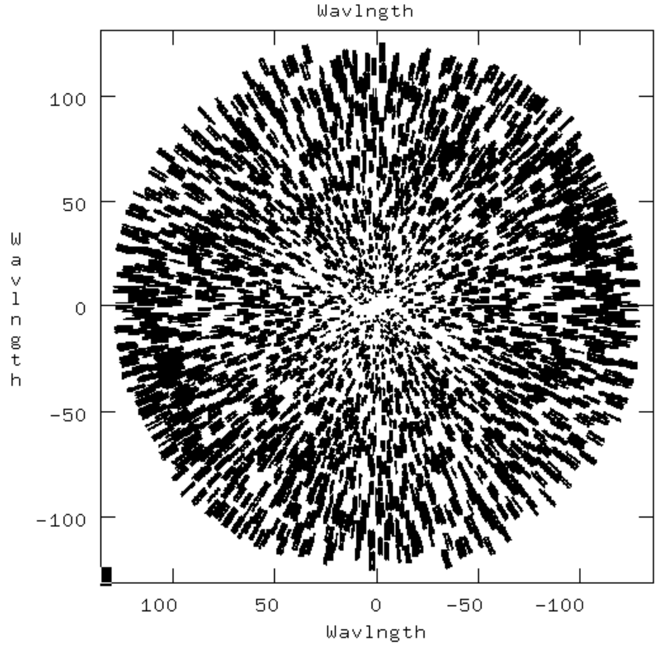


Figure 1: The UV plane coverage of PAPER SA 64 in 10min. Upper plot is for a single channel, lower plot is all channels (bandwidth synthesis)

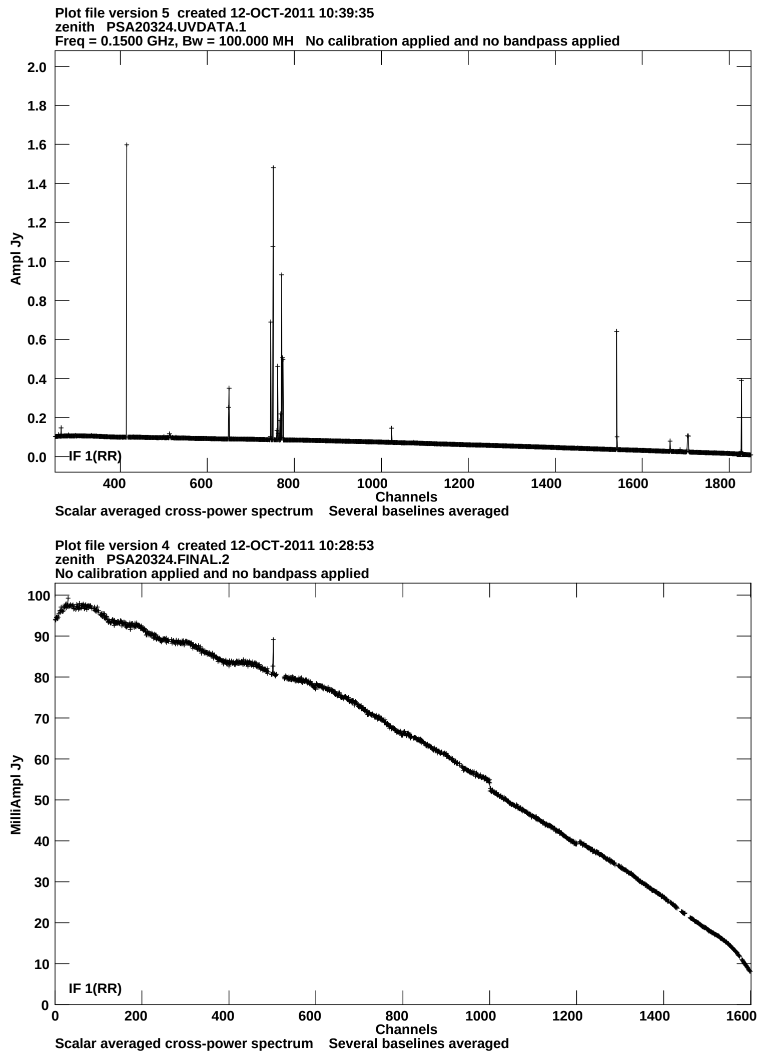


Figure 2: Spectra from PSA64 using a scalar average of all the visibilities (ie. amplitudes only). The upper plot is before data flagging, and the lower after flagging and selfcalibration.

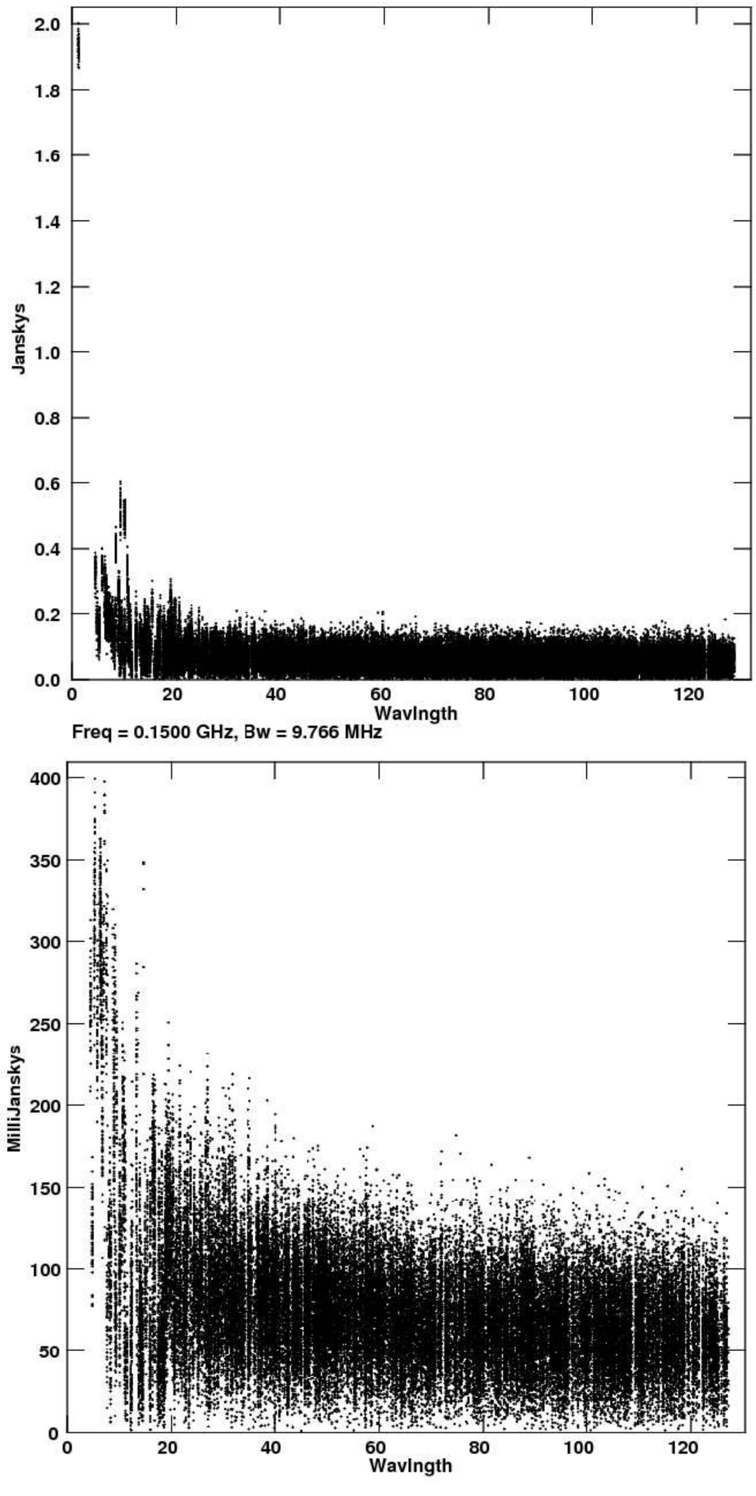


Figure 3: UV plots (amplitude vs. baseline length) for PSA64. The upper plot is before data flagging, and the lower after flagging and selfcalibration.

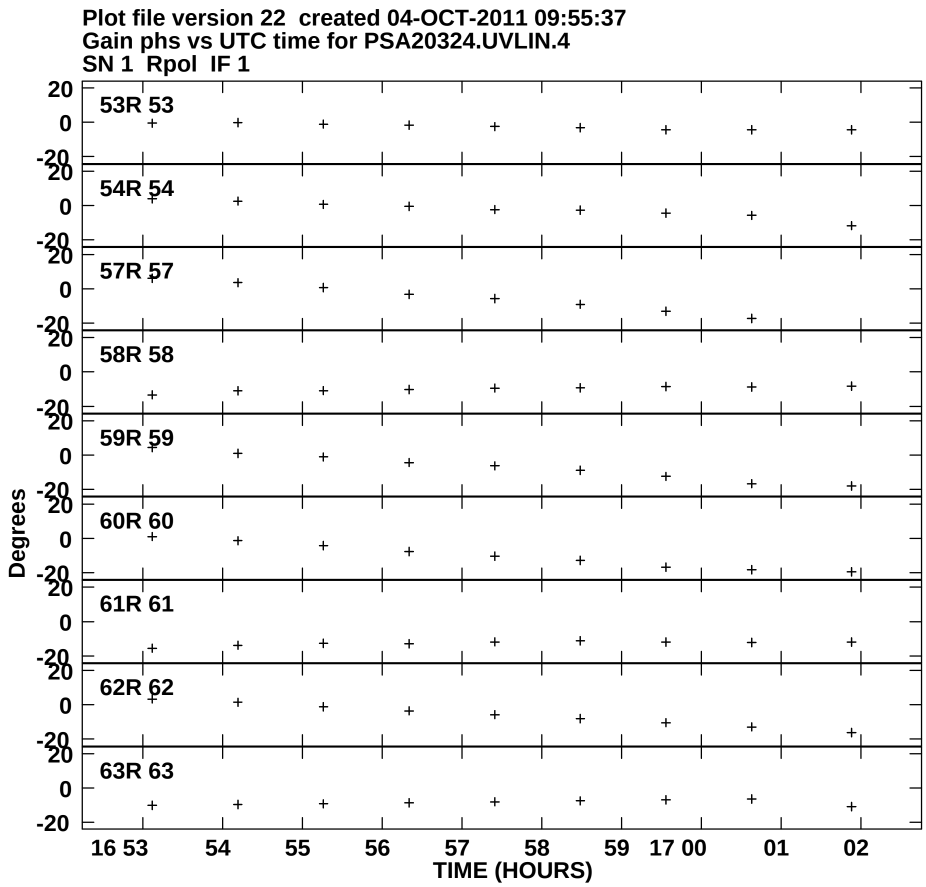


Figure 4: A plot of the antenna-based phase selfcal solutions for 10min of PSA64 data.

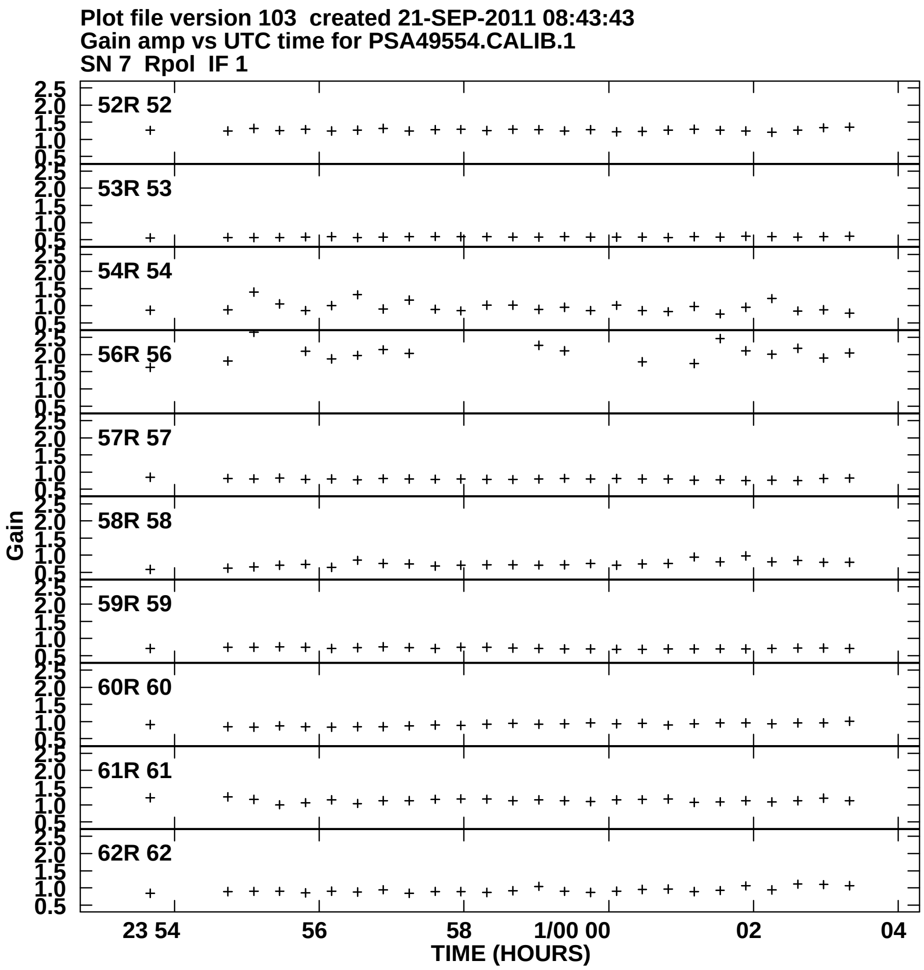


Figure 5: A plot of the antenna-based amplitude selfcal solutions for 10min of PSA64 data.

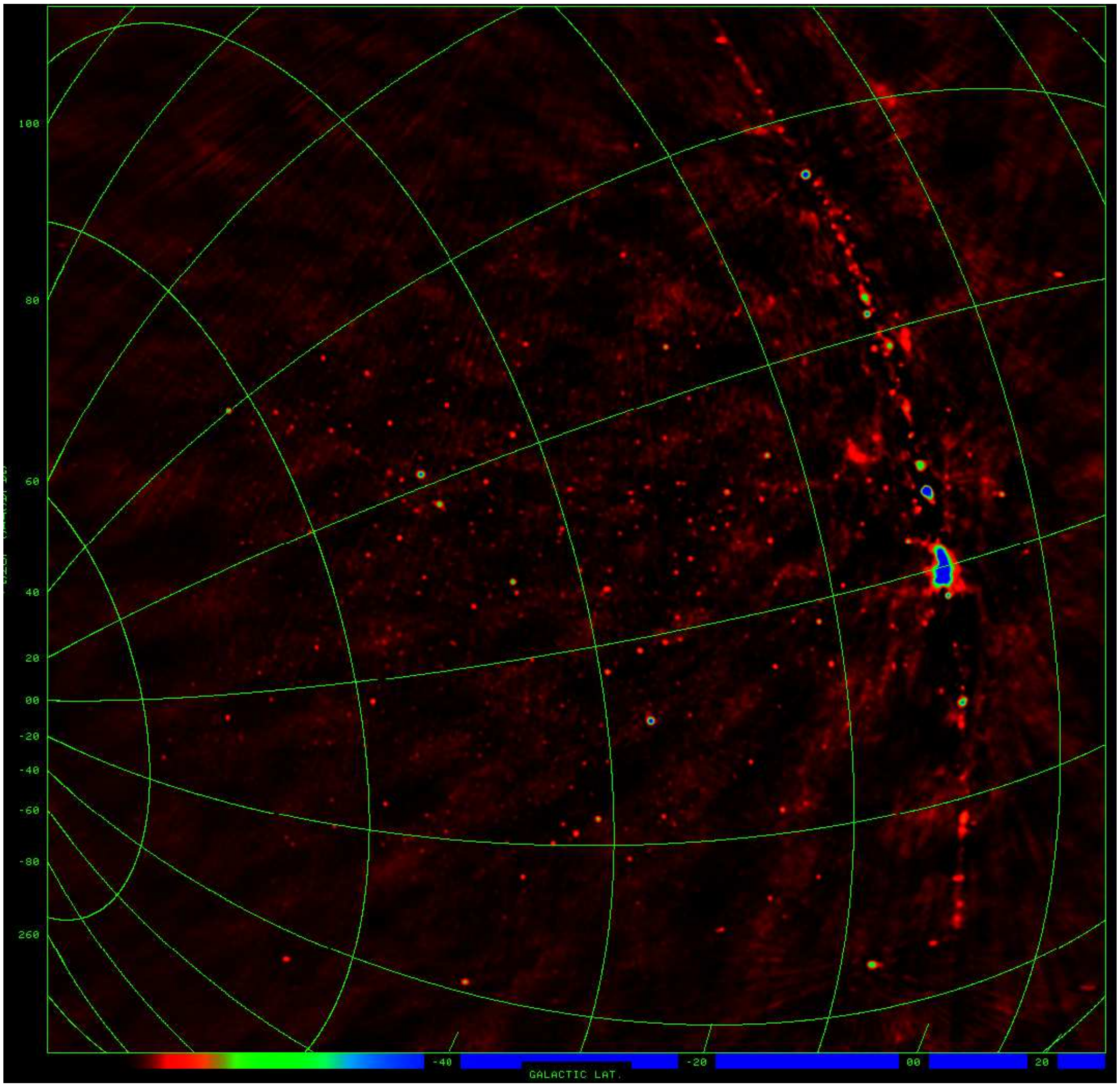


Figure 6: A color version of the PSA64 snapshot image (10min) of dataset 49554.

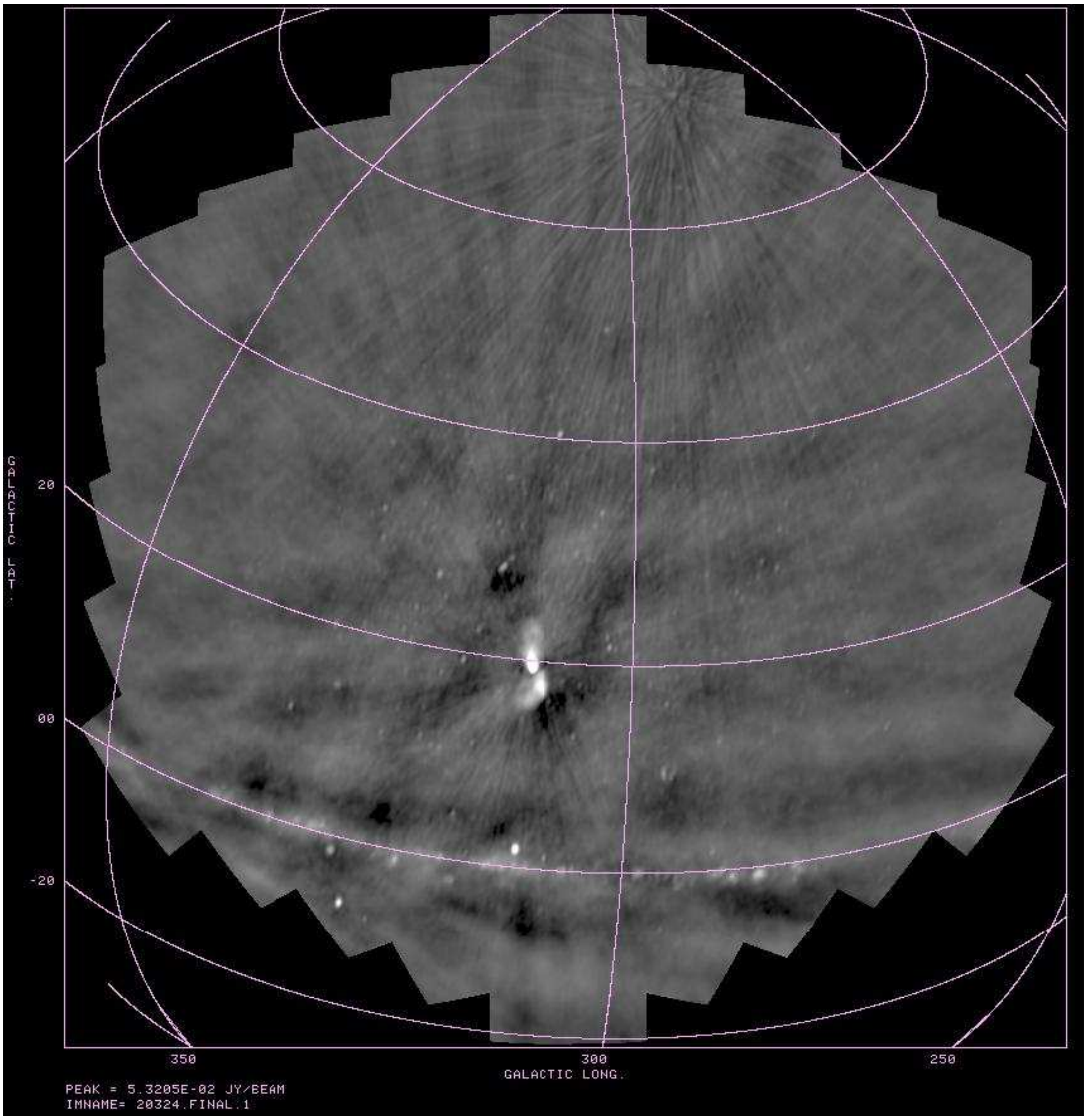


Figure 7: The PSA64 snapshot image (10min) of dataset 20324.

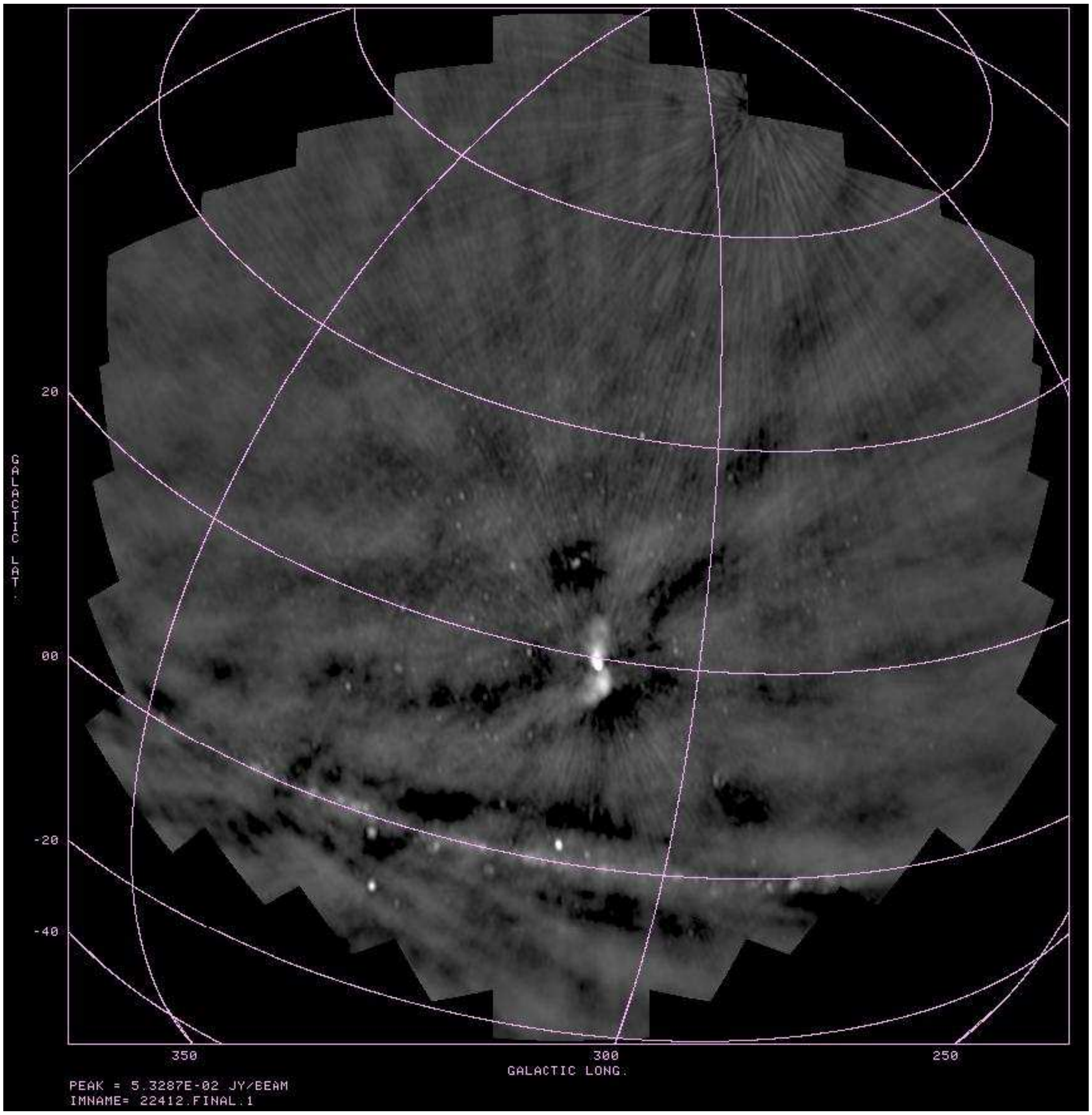


Figure 8: The PSA64 snapshot image (10min) of dataset 22412.

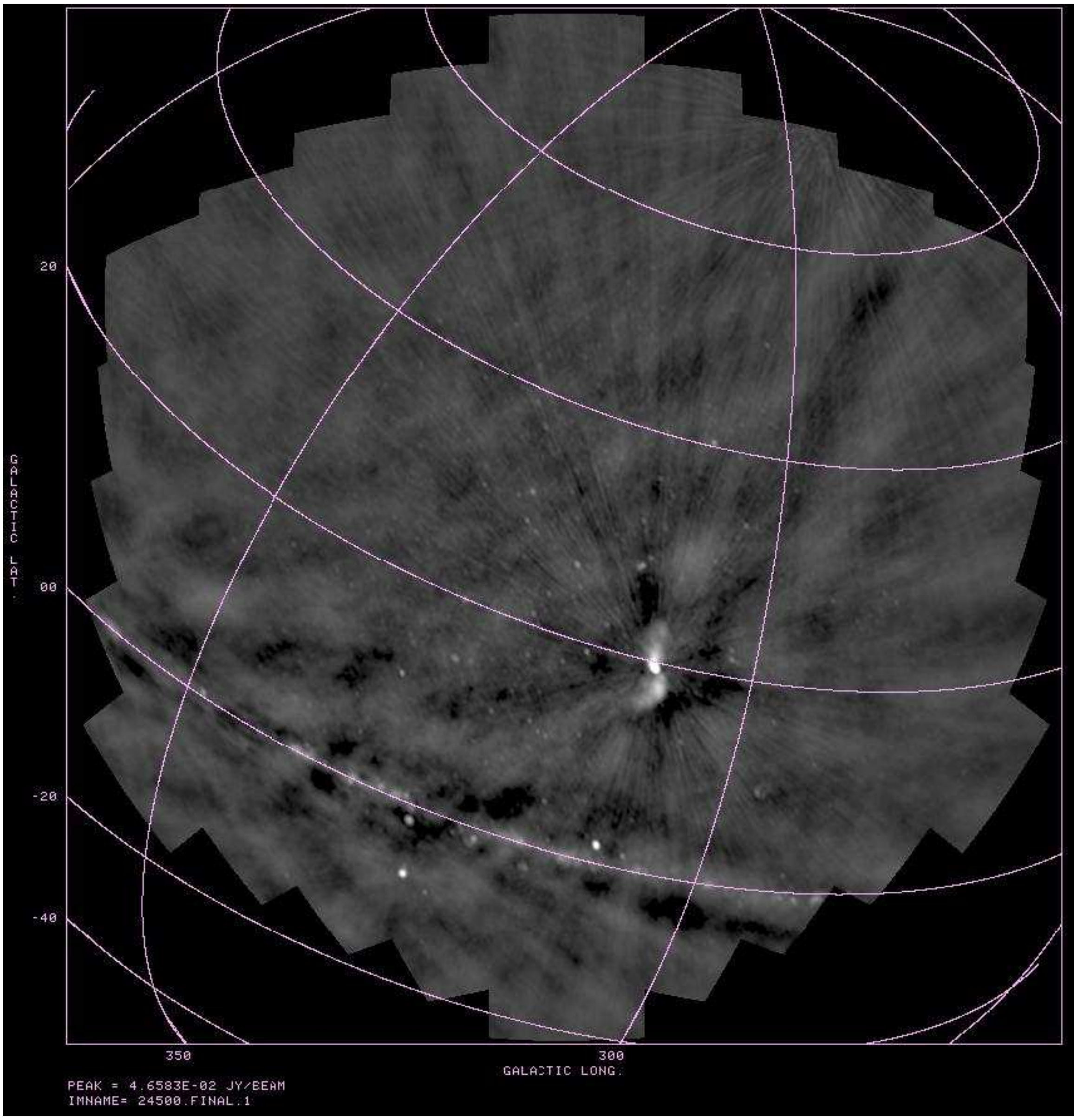


Figure 9: The PSA64 snapshot image (10min) of dataset 24500.

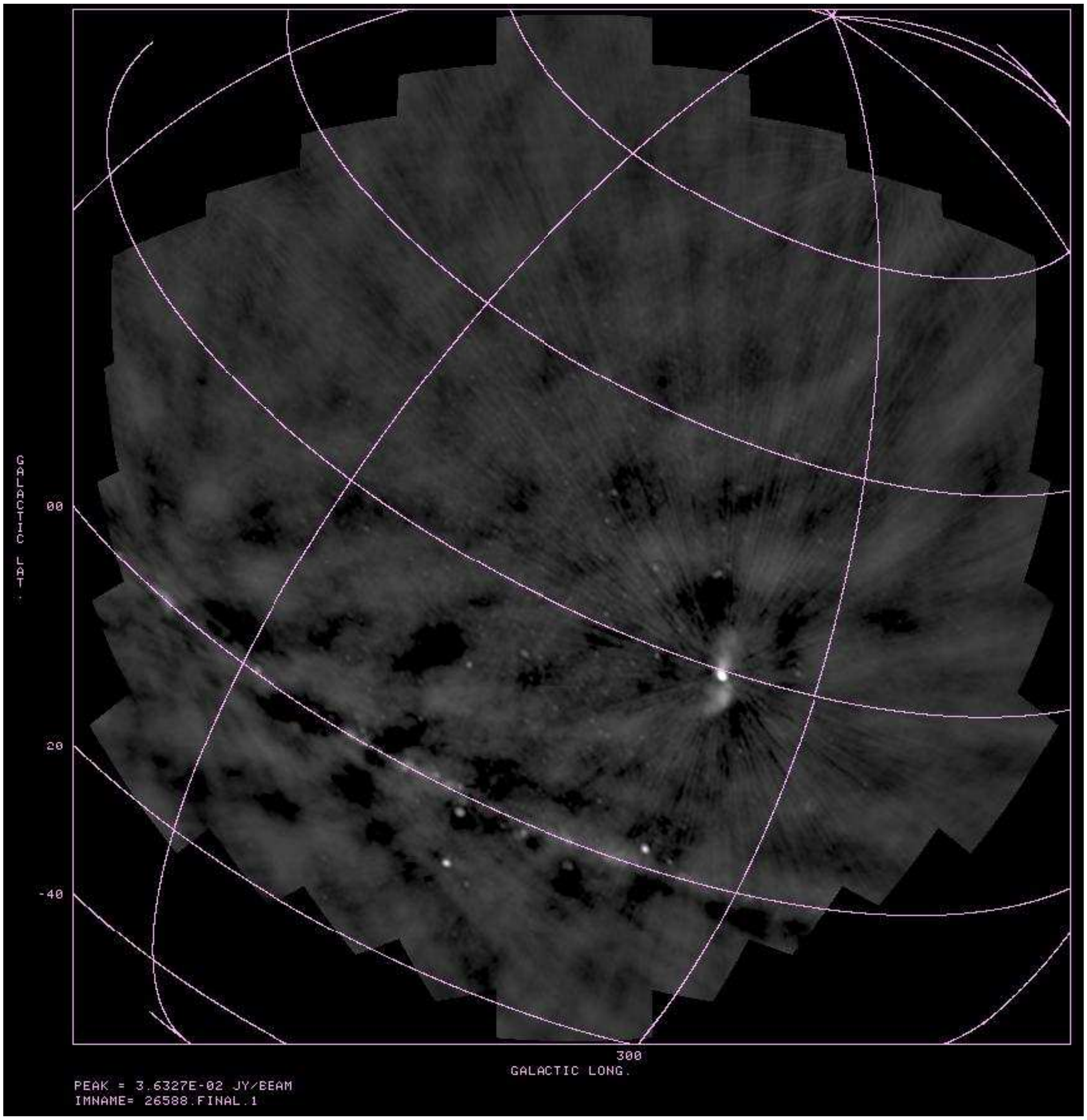


Figure 10: The PSA64 snapshot image (10min) of dataset 26588.

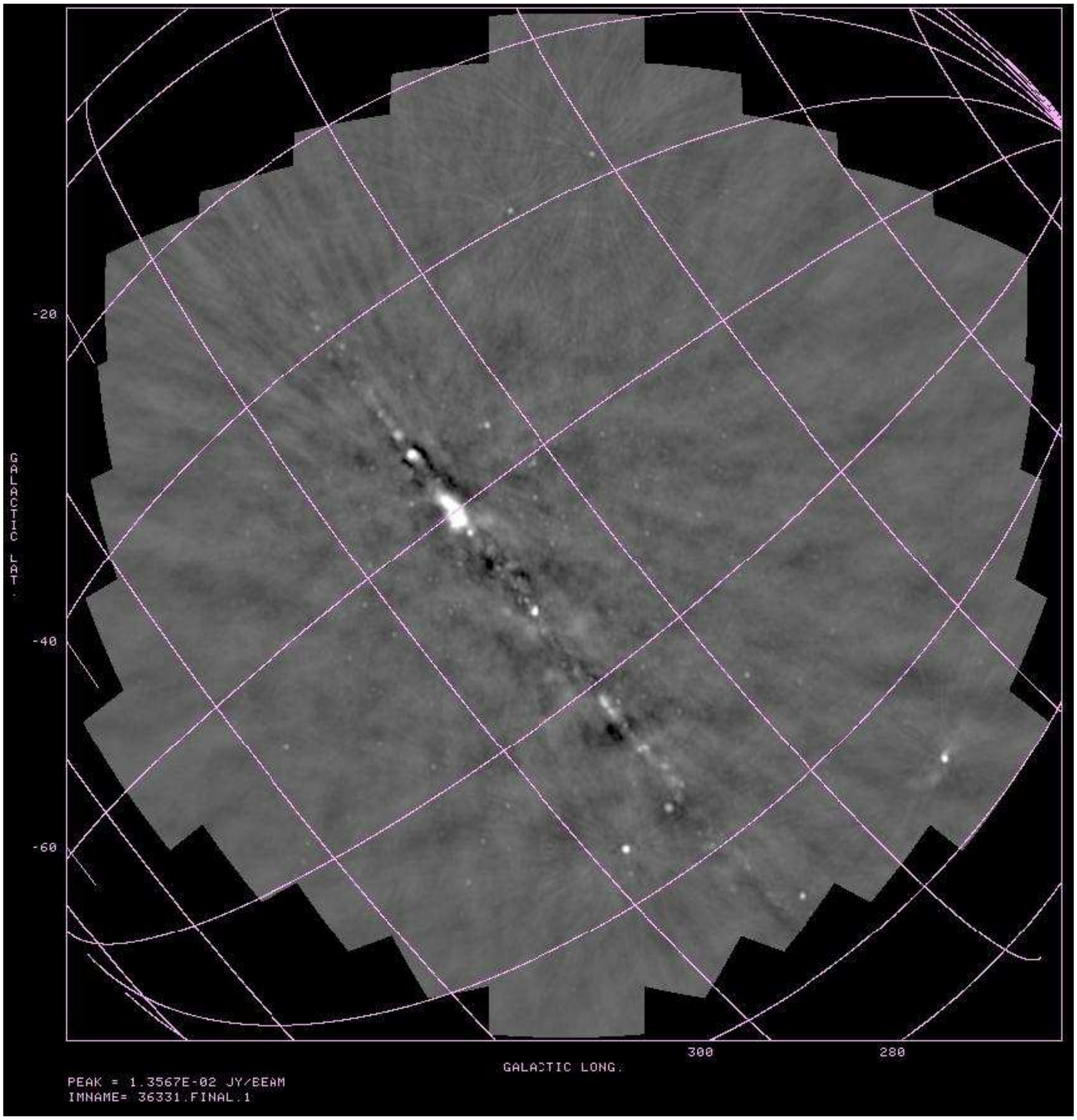


Figure 11: The PSA64 snapshot image (10min) of dataset 36331.

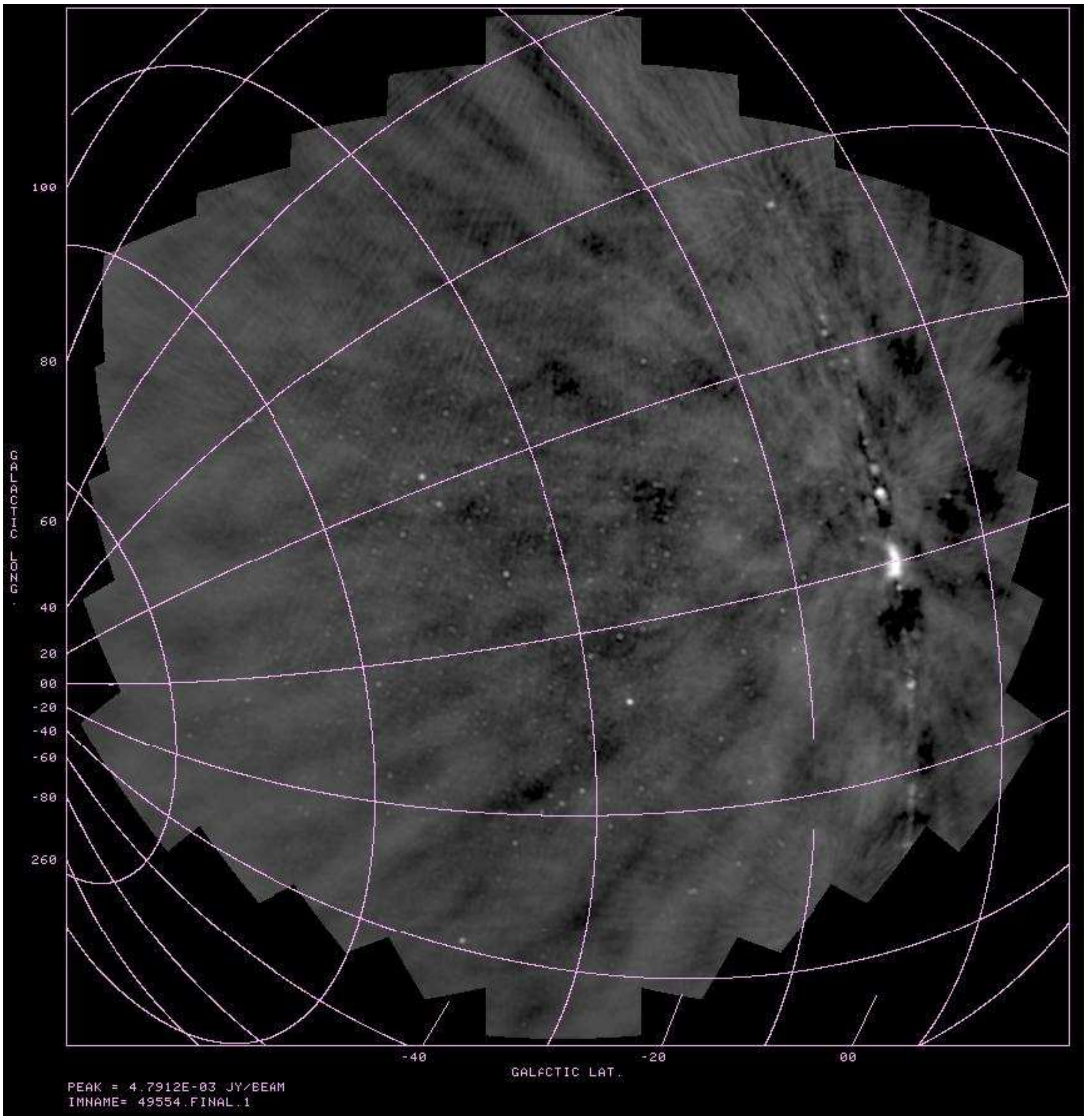
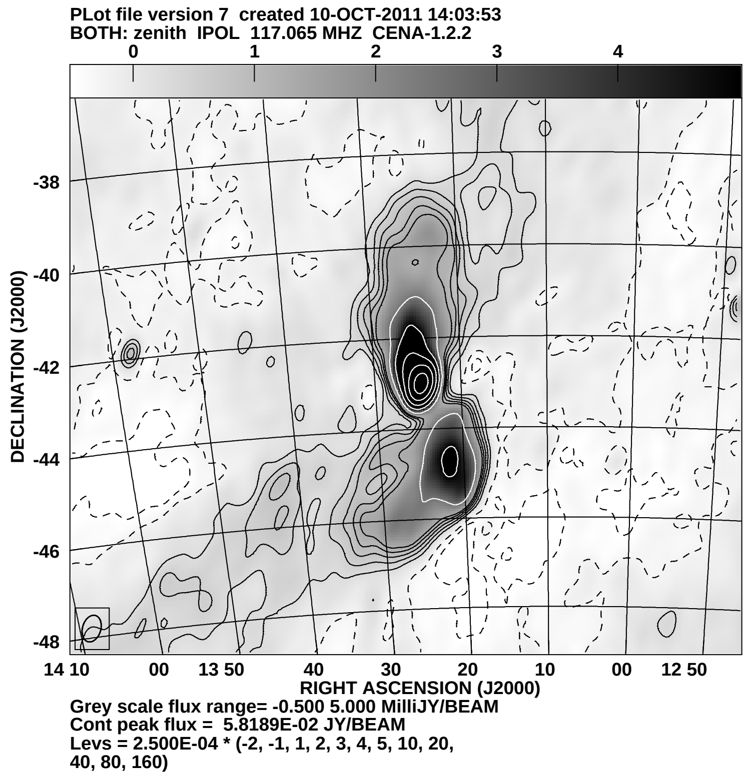


Figure 12: The PSA64 snapshot image (10min) of dataset 49554.



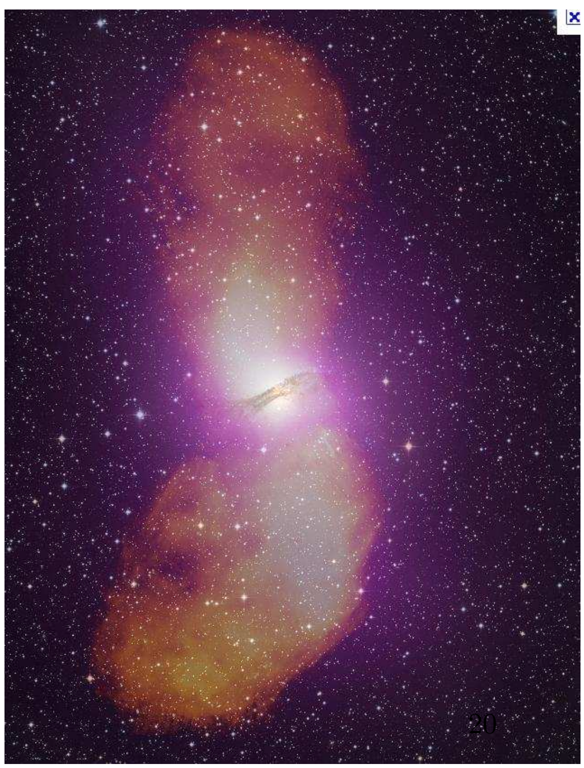


Figure 13: Centaurus A with PAPER and with Parkes.

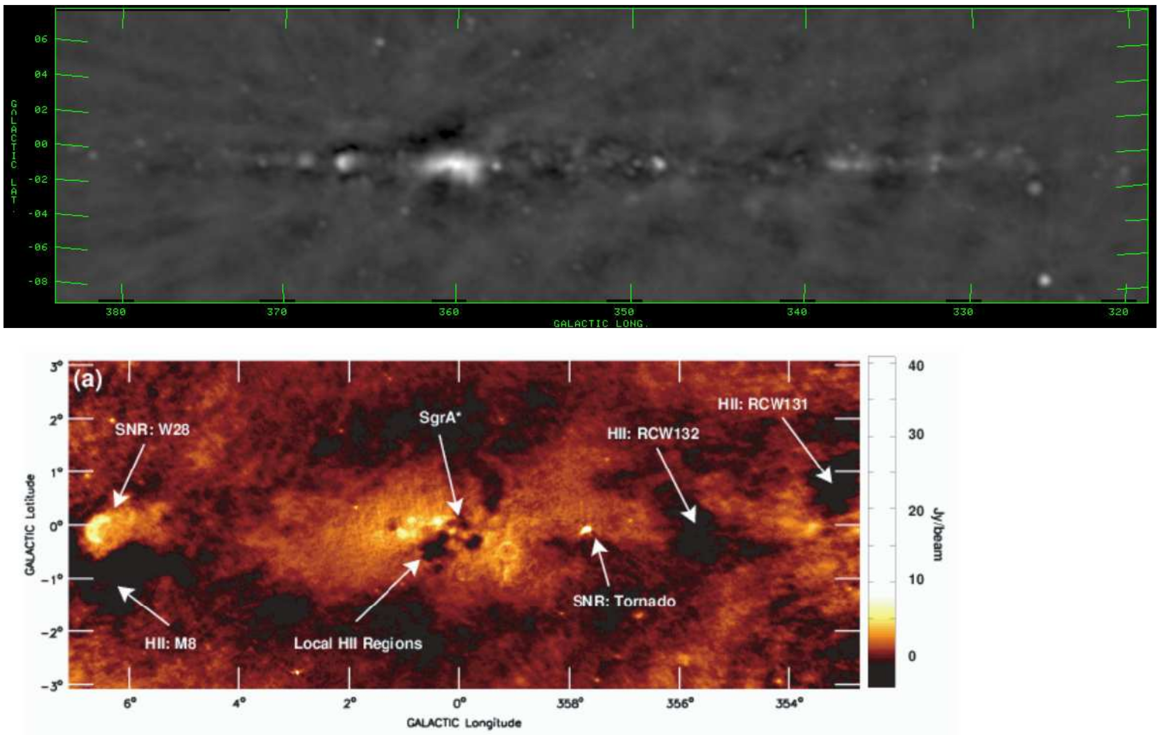


Figure 14: The PSA64 image of the Galactic plane, plus the VLA 74MHz image from La Rosa ea. Note that the scales are different, with the PSA image covering a 60^o and the 74MHz image covering 15^o.

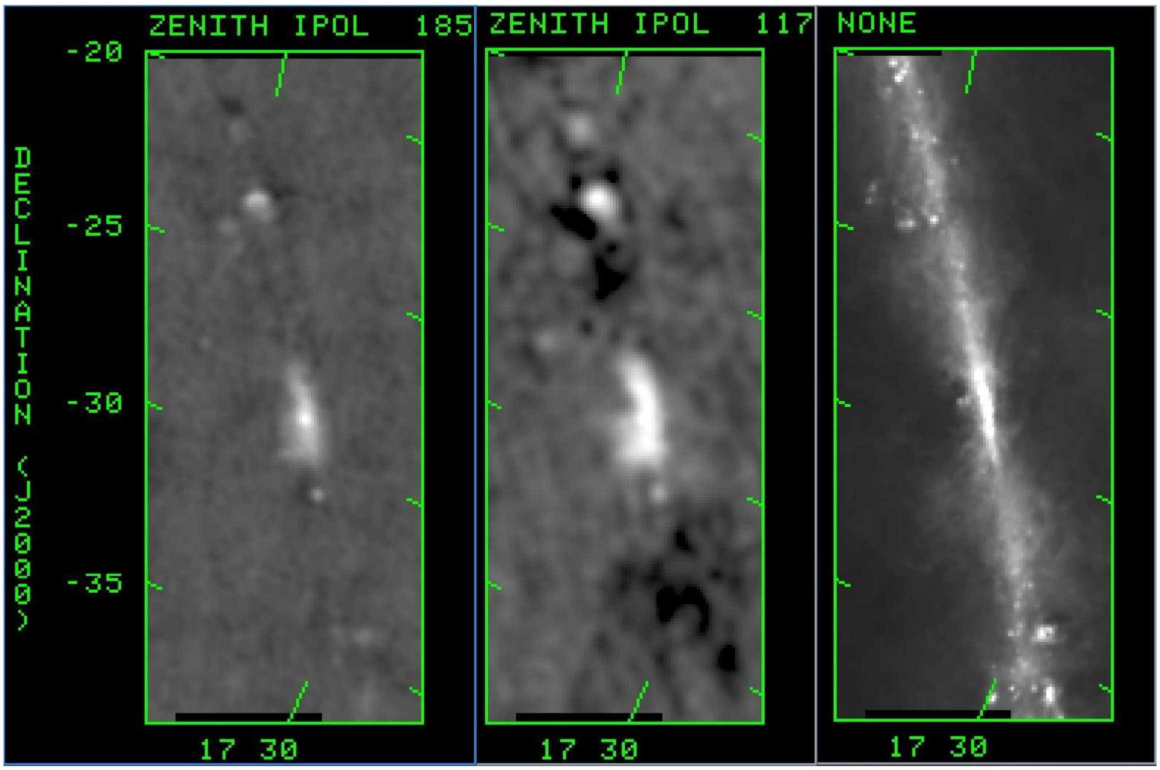


Figure 15: The Galactic center with PSA64 at 120MHz and 180MHz, plus the IRAS 100um image of the dust.

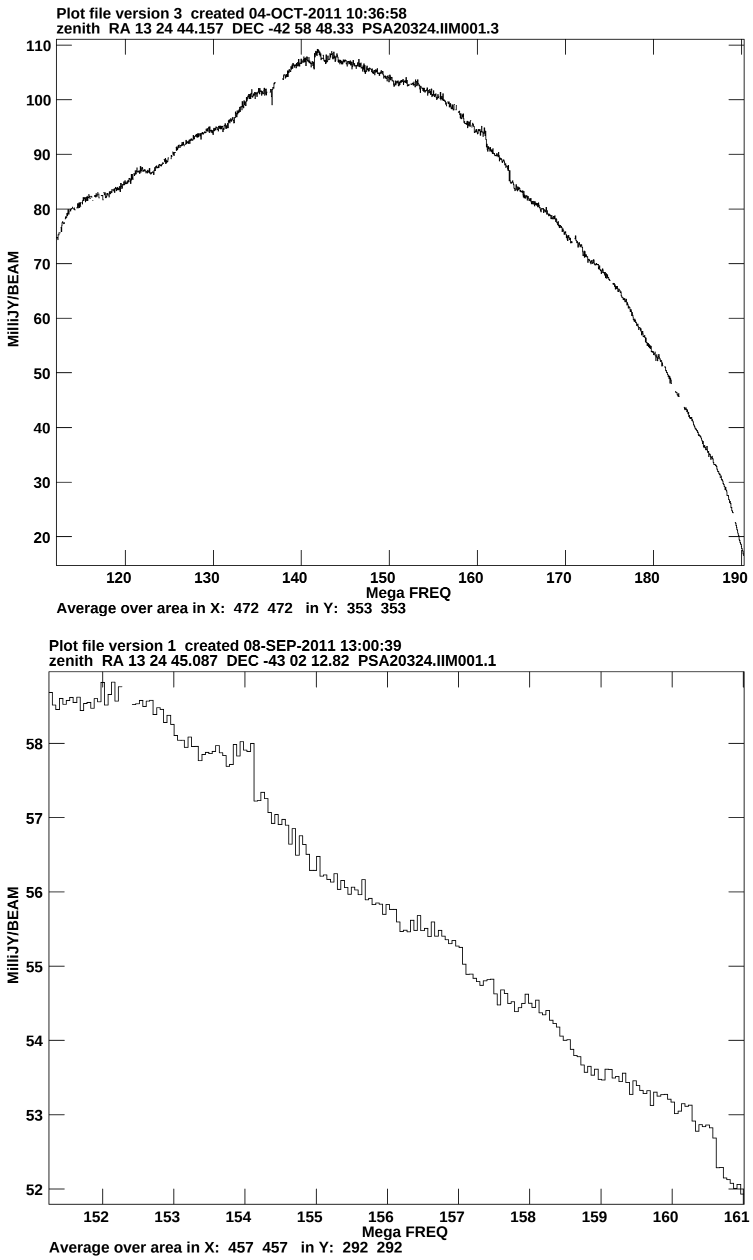


Figure 16: Spectra of Cen A over the full frequency range, and over a 10MHz band. Note that the flux scale is arbitrary.