Fast Targeted Spectral Line Imaging for the MWA

Natasha Hurley-Walker, Curtin University / International Centre for Radio Astronomy Research
May 26, 2020

***Abstract*—Spectral line data processing with the Murchison Widefield Array (MWA) can be a difficult and time-consuming process. This memo outlines a rapid method of performing spectral line data reduction to produce mJy-level noise in single 10-kHz channels, opening up the scientific potential of MWA spectral line data.**

I. Introduction

This memo is aimed at experienced users of MWA data; thus, this introduction is fairly minimal.

Low-frequency spectral lines are expected from various astrophysical sources, such as radio recombination lines, molecular transitions, mased emission, and absorption of continuum emission from distant sources by intervening HI along our line-of-sight. Typically only a few pixels in the very wide field-of-view of the MWA are expected to be sources of such lines for reasonable integration times ($\approx$ 10–100 hours). However, the wide field-of-view of the MWA yields continuum emission from thousands of sources, which must be removed in order to detect lines. This can be expensive, and even modern supercomputers cannot jointly deconvolve the 30.72-MHz bandwidth of the MWA at 10 kHz resolution over the fields-of-view necessary to deconvolve all sources. Existing pipelines (e.g. [1]) divide the data into 1.28-MHz (coarse) channels, and compromise on imaged field-of-view in order to fit the data into the memory constraints of existing computers. So far, this method has not yielded any high signal-to-noise detections.

This memo proposes a new method for spectral line imaging that drastically reduces the computational cost and leaves obtaining the data as the major bottleneck of the processing. However, the method is *targeted*, in that there are limitations to how large a field-of-view can be used (see Section II-B). Fortunately, it is not necessary to know the frequency of the line emission (or absorption).

This short memo introduces the method in Section II, applies it to real data in Section III, and suggests future work in Section IV.

II. Method

A. Visibility subtraction

The visibilities of the MWA contain information about the entire visible sky, which is dominated by 100s of Jys of spectrally-smooth synchrotron emission from the Galactic plane (spatial scales of 1–10° and extragalactic radio sources (spatial scales of 1–5′). This emission is modulated by the primary beam, which is spectrally smooth over 10s of MHz. Lines are expected to be faint ($<$ 100 mJy beam^{-1} for a 1-Jy source) and relatively narrow ($<$ 20 kHz for Galactic, and $<$ 50 kHz for extragalactic lines). Thus, if we average the visbilities over a fairly narrow bandwidth of 160 kHz, we obtain the continuum emission of the sky, outweighing any line contribution by a factor of $\approx$ 3–16. We can do this for each frequency channel as a sliding boxcar, to produce a continuum model. This can be subtracted from the 10-kHz-resolution visiblities to produce a set of visibilities where the only remaining features are spectral lines. Figure 1 shows the effect of applying this technique to a simulated 40-kHz-wide line, and Figure 2 shows the effect of applying this technique to a calibrated set of visibilities.

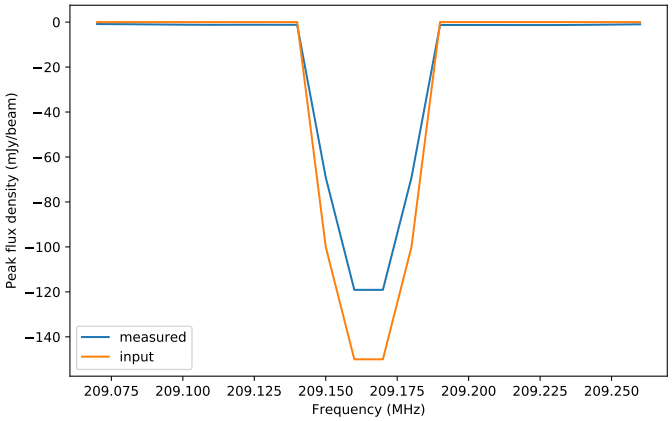

Fig. 1: Boxcar-averaging visibility subtraction applied to a simulated measurement set, in which multiple 1-Jy sources are injected into otherwise noise-free visiblities, and each source has a 40-kHz-wide line at a different frequency. The source shown has a continuum flux density of 1 Jy and a 40-kHz-wide spectral line of input brightness −150 mJy/beam, centred on 209.165 MHz (1 Jy subtracted for clarity). The source lies at the phase centre. The line is recovered with a 20 % brightness reduction.

B. Bandwidth smearing and technique applicability

We can calculate the expected bandwidth smearing of the continuum model obtained in this way. The response to a point source at position *(l,m)* is radially elongated by a factor equal to [2]:

$$\sqrt{l^2 + m^2}\Delta\nu/\nu_0 \quad (1)$$

For instance, at ν_0 =200 MHz, for $\Delta\nu$ =160 kHz, at 5° away from the phase centre, the smearing is 14.4”. The final response is the convolution of the smearing with the synthesised beam, which is 60” at this frequency, yielding 62”. Thus we can expect $\approx$ 3 % residuals around sources 5° from the phase centre. Lines of brightness $<$ 30 % of any

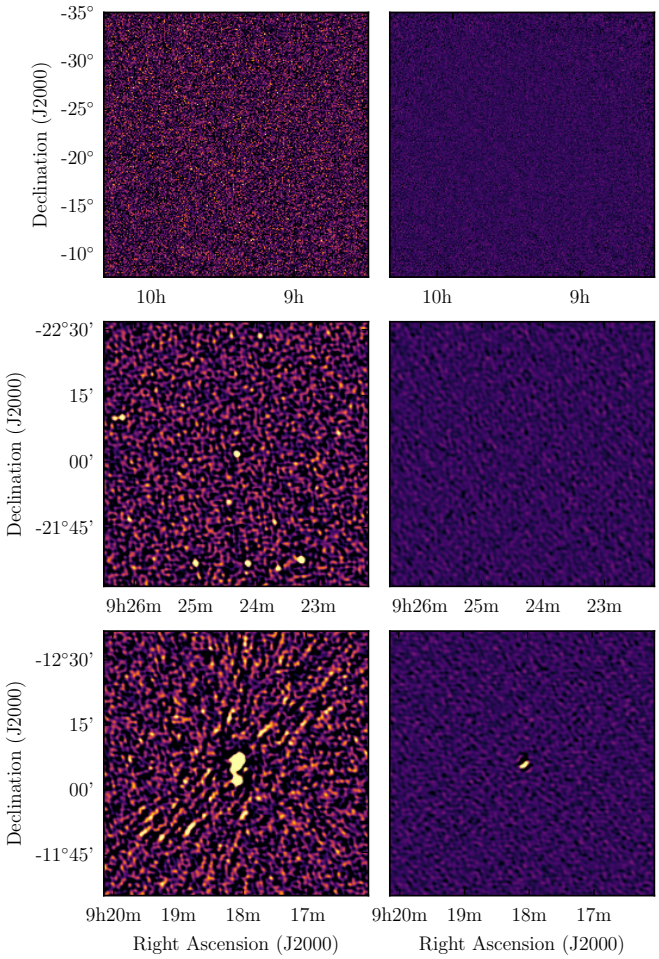

Fig. 2: The effect of applying the visibility continuum subtraction technique to the observation 1206025848, part of project G0037. The left panels show the image formed by using WSCLEAN to perform inversion and deconvolution over an 8000×8000 pixel area over the full 30.72-MHz bandwidth. The right panel shows the same imaging after visibility-based continuum subtraction has been performed. The top row shows the full image, the middle row shows the phase centre (on TN 0924-2201), and the bottom row shows Hydra A, a bright radio galaxy 10° away from the phase centre. The colour stretch in all images is linear between -50 and $+100$ mJy beam^{-1}. The RMS noise is 14 mJy beam^{-1} before subtraction and 3 mJy beam^{-1} after subtraction.

local source would thus be rendered undetectable (e.g. we would not believe a 30 mJy beam^{-1} line found on a 100 mJy source 5° from the phase centre). Additionally, the size of the bandwidth averaging chosen to produce the continuum model will suppress lines which are similar in size. Figure 3 shows the detectability of lines as a function of distance from the phase centre, for different choices of bandwidth averaging, for different line widths. This diagram helps guide a potential observer as to whether the technique would be useful for widefield searches for lines, particularly if searching in absorption.

For instance, the "High-Z" project (code G0036; PI Allison, J.[1]), postulates searching for HI absorption toward candidate $z > 5$ radio galaxies, with brighter candidates expected to require less integration time. Using their assumptions, a 1 Jy source might be expected to produce a 100 mJy beam^{-1} absorption line with a line width of ≈ 40 kHz. Figure 3 shows that visibility averaging over 160 kHz could safely be used to search for such lines up to 3.5° from the phase centre. However, since the G0036 project is targeted at specific high-z radio galaxies, a wider averaging could be used. Widefield searches are only possible for very narrow lines, where narrower bandwidth averaging can be used.

Two further effects which may guide the choice of bandwidth averaging bin are channel cross-contamination and sidelobe noise. Wider bins mean bringing in more channels, which may overlap the edges of the coarse channels, where the noise is higher, or include unflagged RFI. This will spread the effect of these channels across previously uncontaminated channels. Since wider bins also increase the effects of bandwidth smearing, there are more artefacts in the image, which produce sidelobes, which in turn raises the noise level of the image. Figure 4 shows the effect of increasing the bin size on the RMS noise of a single 10-kHz channel of observation 1206026328, part of project G0037. It would take extensive testing to determine whether a noise floor is reached with less time integration for larger bin widths, but this would be expected behaviour.

C. *Pipeline*

There are several steps needed to automatically process the spectral line data, of which this technique is just one. I have written a pipeline that performs the following stages (with time estimates per observation):

- Download the data from ASVO (highly variable: 10 min – 6 hours);
- Automatically calibrate based on the GLEAM-X-pipeline sky model (20 min);
- Apply the calibration solutions (10 min);
- Flag the (u, v) data for residual radio frequency interference (Duchesne, S., priv comm.) (5 min);
- Optionally phase-rotate the visibilities to a common phase center (necessary for projects taken in drift scan mode) (5 min);
- Perform the visibility-based continuum subtraction (15 min);
- Use WSCLEAN[3] to invert all 3,072 fine channels in blocks of 1,024 (30 min for 256-pixel targeted images, longer for wider fields);
- Delete unnecessary point spread function and duplicate images, and rename remaining images (< 1 min);
- Scale each image by the primary beam (1 min);
- Generate RMS images for optimal inverse variance weighting (1 min);

Finally, I use SWARP to combine all images and produce a final cube, which takes $\approx$ 1–2 min per fine channel, depending on number of observations used; this would scale up in

[1] http://www.mwatelescope.org/data/observing

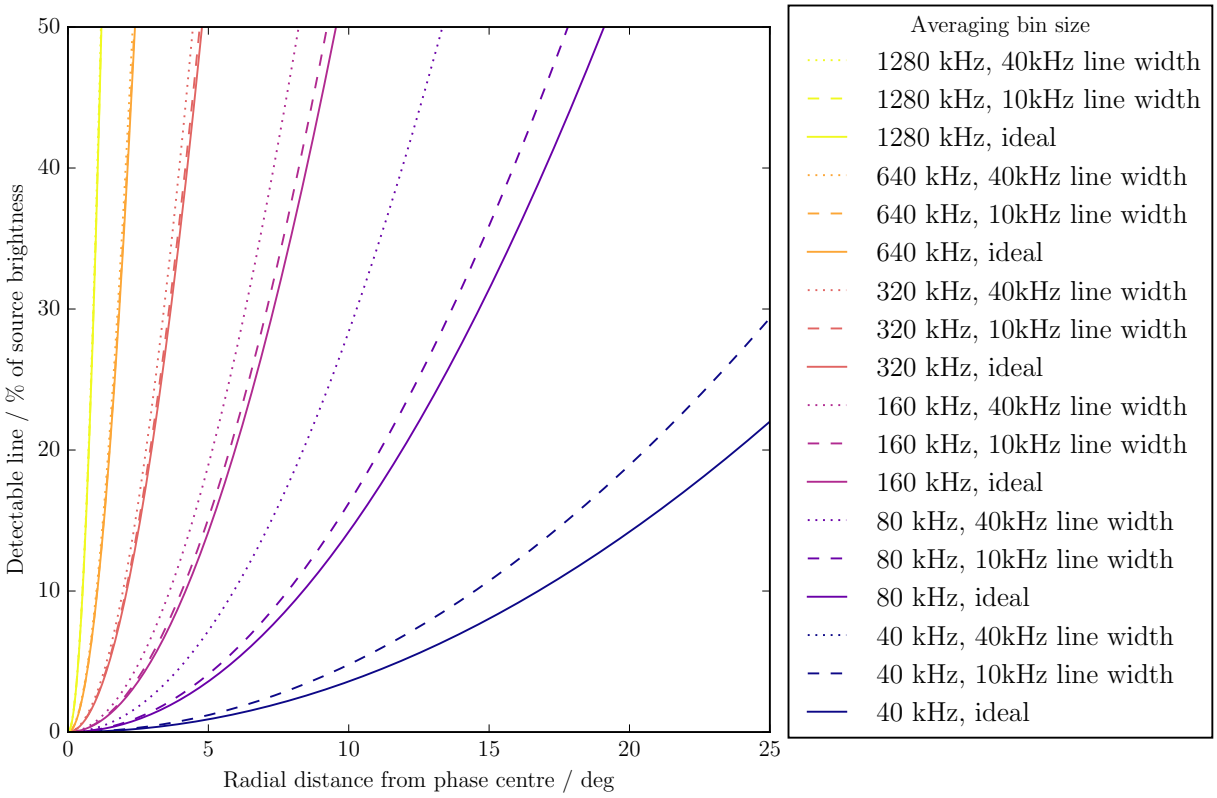

Fig. 3: The effect of bandwidth smearing and boxcar averaging width on the accuracy of the continuum subtraction. The smearing leaves a residual which is related to the distance from the phase centre (Equation 1). We assume this residual must be $10\times$ smaller than the line brightness for a line to be detectable. Bandwidth averaging also suppresses lines, moreso for lines closer to the width of the averaging bin. The plot shows the combination of these different effects as a function of distance from the phase centre. Hypothetical spectral lines above each curve are detectable for the parameters indicated in the legend. "Ideal" lines are infinitely thin and not suppressed by this method of continuum subtraction.

proportion to the area of larger images for a widefield survey. At $\approx$ 1.5 hr per observation, this pipeline can in general process data faster than it can be obtained from ASVO.

III. Results

I performed testing of this pipeline on the largest single-object spectral line project available, the search for HI absorption toward the $z = 5.2$ radio galaxy TN 0924-2201 (code G0037; PI Sadler[2]). I used the data from the nights of 2018-03-24, 25, 27, 30, and 31, comprising 409 observations over 13.5 hours, covering 208–238 MHz. Calibration showed that the upper channels were noisier than the lower channels, perhaps due to the decreased sensitivity of the MWA at these higher frequencies. Flagging also showed an increase in the amount of RFI in these channels. Since this is a targeted search, I phase-rotated all data to the location of TN 0924-2201, RA$= 09^{\rm h}24^{\rm m}20^{\rm s}$, Dec$= -22^\circ01'42"$, and imaged using 256×256 pixels.

The MFS images offer a convenient way to check the quality of the method, since any artefacts will propagate through to them, and there are far fewer to examine. Measuring their RMS is a fast and effective way to detect artefacts, and visual inspection can be used to examine their characteristics if needed. Fig 5 shows the result of checking the quality of the G0037 observations, suggesting the removal of observations 1206443352, 1206445152, and 1206533112. The same figure shows that the RMS of the highest band is consistently higher than the lower frequencies, and also changes more rapidly with time. This could be due to RFI, sources in the far sidelobes of the primary beam, or simply the field drifting more quickly through the smaller primary beam.

When combining the images, I did so cumulatively, so that I could examine whether the noise was decreasing as expected. Figure 6 plots the average RMS for the 10-kHz slices in the the lowest, middle, and highest (1.28-MHz) coarse channels as a function of integration time t. The lower end of the band decreases as $t^{\frac{1}{2}}$ as expected, but the upper end shows a shallower decrease, most likely due to the higher and more variable noise levels at the upper end of the band (Figure 5, and possibly also increased flagging to accommodate a higher level of RFI. Figure 7 shows the full spectra for each accumulated data cube. There are no clear line detections, but the noise levels reached are the lowest ever achieved for 10-kHz channels with the MWA.

[2] http://www.mwatelescope.org/data/observing

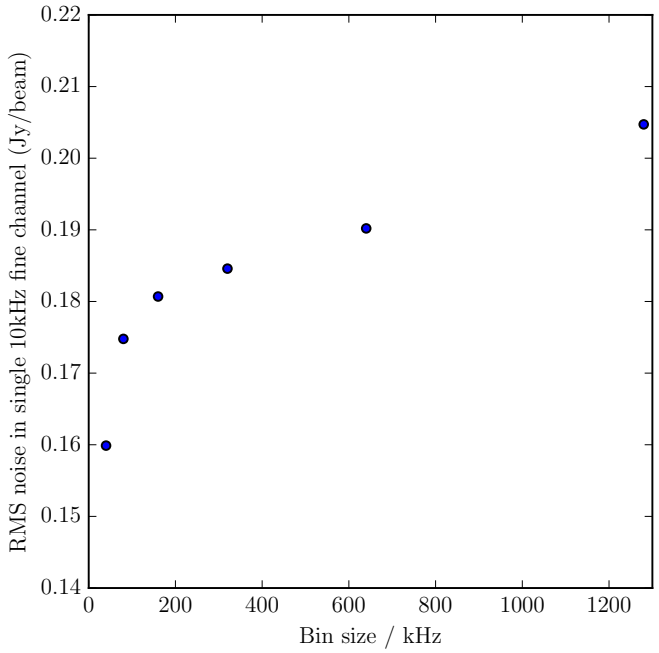

Fig. 4: The effect of boxcar averaging bin width on the noise level of channel 490 of observation 1206026328, part of project G0037. Bandwidth smearing leaves artefacts around sources outside the imaged field-of-view, and since they are not cleaned, their sidelobes increase the local noise level. There may also be some addition to the noise level from including unflagged RFI in the continuum estimation.

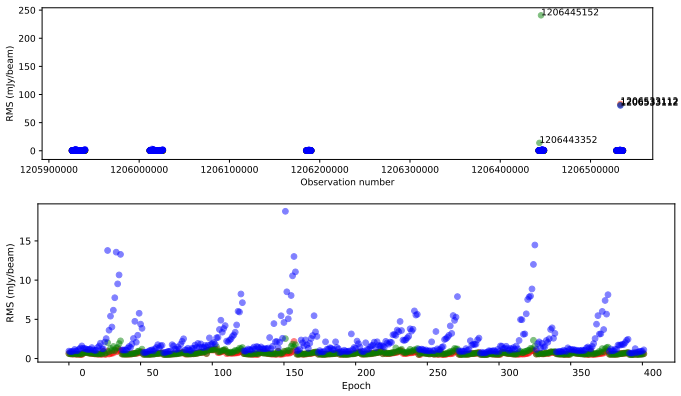

Fig. 5: RMS of the MFS images produced by the pipeline. Red indicates 208–217 MHz, green 218–227 MHz, and blue 228–237 MHz. The first panel shows the RMS as a function of observation ID before flagging. Observations 1206443352, 1206445152, and 1206533112 are clearly identified for removal from onward processing. The second panel shows the RMS after flagging, as a function of epoch, for clarity. The RMS of the 228–237-MHz band shows a repeated rising trend, perhaps due to RFI, sources in the far sidelobes of the primary beam, or simply the field drifting more quickly through the smaller primary beam.

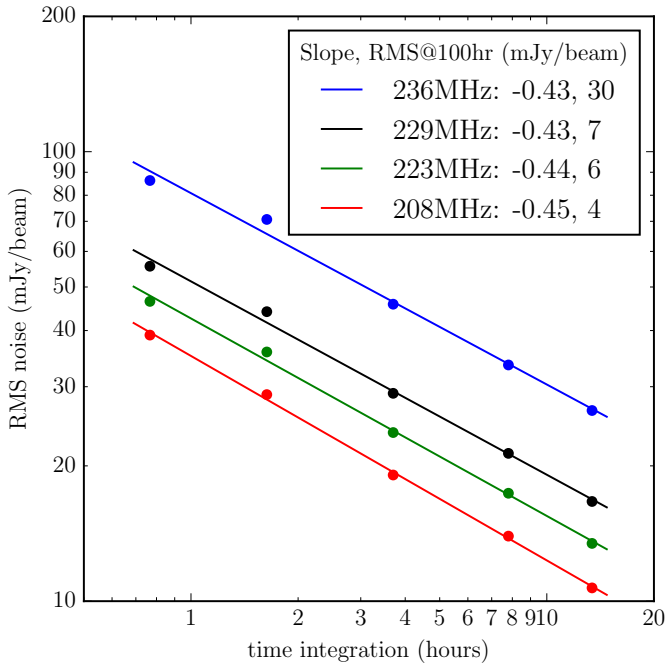

Fig. 6: The reduction in RMS noise with respect to integration time for four fine channels at the marked frequencies, from the accumulated G0037 data processed through the pipeline described in this memo. The black line at 229 MHz shows the RMS noise for the channel in which associated HI from TN 0924-2201 might be detected. The points show the data and the lines show a least-squares fit, with the slope given in the inset legend. The predicted RMS at 100 hours of integration is given in the legend. The full spectra at the location of TN 0924-2201 for each data point are shown in Figure 7.

IV. Optimisation and further work

I have not made an attempt to conduct a widefield spectral line survey using this technique, but it should be possible. Cosmology fields such as the Galaxy And Mass Assembly (GAMA) fields are only $\approx 10^\circ \times 5^\circ$, so deep integrations at their centres could use this technique to search them for bright lines ($> 15\%$ source brightness). The main challenge of searching for Galactic spectral lines is the difficulty of deconvolving Galactic structure, but this technique completely eliminates this problem, and the only challenge will be the increased system noise from the higher sky temperature in these fields. Currently, no correction for ionospheric position shift is included in the pipeline, but this would be trivial to implement, especially for small fields-of-view in extragalactic sky, and may increase sensitivity by decreasing blurring.

V. Conclusion

I have created a fast spectral line pipeline and used it to process 13.5 hours of data in one week, dominated by the time it took to obtain visibilities from ASVO. This resulted in the deepest spectral line images created so far with the MWA. This technique can be utilised to quickly perform spectral line searches of any deep observations in the archive, and may make new spectral line projects possible.

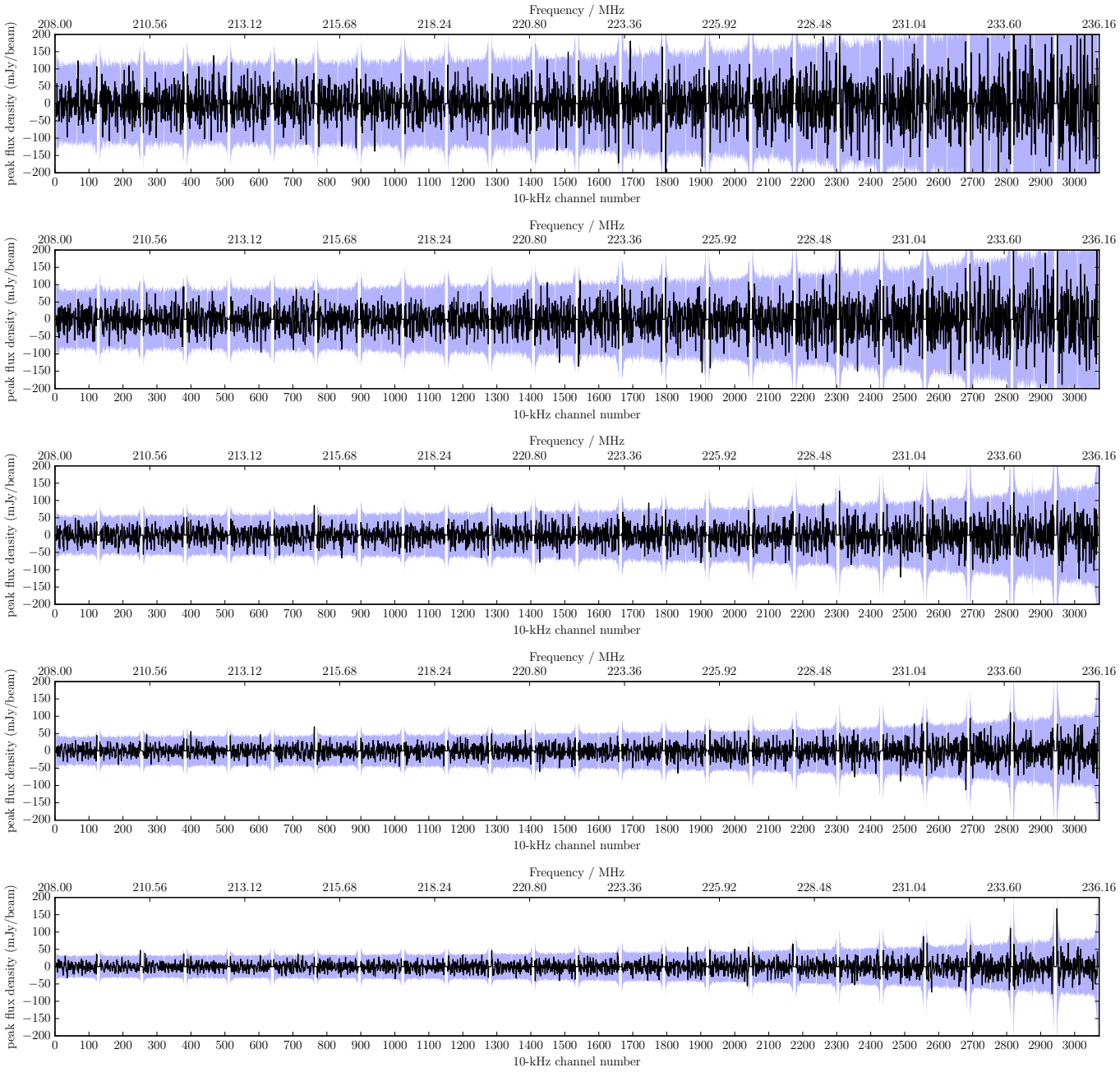

Fig. 7: Spectra for the G0037 data processed through the pipeline described in Section II-C. The light blue shading indicates 3× the local RMS of the channel, while the black lines indicate the peak flux density value at the location of TN 0924-2201. From top to bottom, the spectra show the accumulation of data over the first half of 2018-03-27 (23 observations); all of 2018-03-27 (49 observations); 2018-03-27 and 31 (111 observations); 2018-03-24, 27, and 31 (234 observations); and all data on 2018-03-24, 25, 27, 30, and 31 (405 observations). The RMS values measured in Figure 6 are taken from the same cubes.

VI. Acknowledgements

Many thanks to James Allison and Liz Mahoney for useful feedback which improved this work. I am supported by an Australian Research Council Future Fellowship (project number FT190100231) funded by the Australian Government. This scientific work makes use of the Murchison Radio-astronomy Observatory, operated by CSIRO. I acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. I acknowledge the Pawsey Supercomputing Centre, which is supported by the Western Australian and Australian Governments. I made use of the IPython package [4]; Matplotlib, a Python library for publication quality graphics [5]; Astropy, a community-developed core Python package for astronomy [6], [7]; and NumPy [8]. I also made use of the visualisation and analysis package DS9[3]. This memo was compiled in the very useful free online LaTeX editor Overleaf.

References

[1] C. D. Tremblay, N. Hurley-Walker *et al.*, "A first look for molecules between 103 and 133 MHz using the Murchison Widefield Array," *MNRAS*, vol. 471, no. 4, pp. 4144–4154, Nov. 2017.

[2] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy, 2nd Edition*, 2001.

[3] A. R. Offringa, B. McKinley *et al.*, "WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy," *MNRAS*, vol. 444, no. 1, pp. 606–619, Oct. 2014.

[4] F. Pérez and B. E. Granger, "IPython: a system for interactive scientific computing," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21–29, May 2007. [Online]. Available: https://ipython.org

[5] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[6] Astropy Collaboration, T. P. Robitaille *et al.*, "Astropy: A community Python package for astronomy," *A&A*, vol. 558, p. A33, Oct. 2013.

[7] A. M. Price-Whelan, B. M. Sipőcz *et al.*, "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package," *AJ*, vol. 156, p. 123, Sep. 2018.

[8] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: A structure for efficient numerical computation," *Computing in Science Engineering*, vol. 13, no. 2, pp. 22–30, March 2011.

[3] http://ds9.si.edu/site/Home.html