

SMA Technical Memo #165? (draft)

Subject: **A METHOD FOR HANDLING SMA DATA FROM SWARM CORRELATOR -**

Solving for bandpass of high- or full- spectral resolution data

Date: September 6, 2016^{\$}

From: Jun-Hui Zhao (SAO)

^{\$}Updated from the versions since February 25, 2016

Abstract -

Along with the new wideband upgrading for SMA, we are implementing **Miriad**^[1] software to accommodate large volumes of SMA data; in particular, for the products of high spectral resolution and wideband interferometer array data produced from the **SWARM** correlator, the software upgrading is indispensable. This Technical memo intends to report the progress of **Miriad** software development for the reduction of **SWARM** data. Specially, a method of pre-processing high-spectral resolution **SWARM** data in solving for bandpass is described in details. The additional subroutines implemented in the **Miriad** bandpass solver include the micro-processes prior to the matrix solver (1) visibility weighting in both *variance* and *source structure*, (2) normalization and (3) pre-processing for enhancement of signal-to-noise ratio with *moving smoothing* and/or *orthogonal polynomials' fitting* algorithms. A testing data set was produced on February 11, 2016 with two quadrants (2Q) of the **SWARM** correlator, producing four chunks of high-resolution spectral data (0.1 MHz per channel and 16,384 channels per chunk); eventually, the **SMA SWARM** correlator will produce a total of eight such spectral chunks. The target source is Sgr B2 North, a site of forming a massive star cluster. Five antennas were used in the array operation during the test track. A demonstration is shown using the testing data.

1. Motivation -

As the **SMA SWARM** correlator comes online, reliable bandpass corrections become a pressing issue for probing fine structure of kinematics in imaging molecular lines with a high spectral-resolution as well as comprehensive distributions in continuum emission with wideband (WB) data employing the multiple-frequency synthesis (MFS) imaging technique. For some special objects, the line width appears to be very narrow; for example, the velocity width of molecular outflows from a brown dwarf candidate, such as sOri-2, is only 0.5 to 1 km/s^[2]. The full spectral resolution (0.1 MHz per channel or 0.1 km/s at 1 mm) provided by the **SWARM** correlator is needed for the study of these narrow line systems. However, a poor value for signal-to-noise ratio (S/N) provided for such high-spectral resolution data becomes an issue in solving for a reliable bandpass, given the fact that a lack of strong QSOs at higher frequencies in the range of submillimeter wavelengths and structures of non-point source for the strong dust objects from the solar system. The bandpass solvers that were developed for lower frequencies need to be upgraded with new optimizing algorithms in confront with the limited budget in S/N. On other hand, for wideband continuum, in addition to bandpass and delay corrections, the phase between individual spectral chunks must be aligned well in order to achieve a high-dynamic range in imaging process. We have been developing new algorithms to implement **Miriad** software in order

to meet the requirements for a variety of science cases in the reduction of **SWARM** data. In this memo, micro-processes prior to as well as auto-editing subsequent to solving for antenna-based bandpass solutions for an interferometer array are described and discussed in details. These algorithms are implemented in the **Miriad** program **smamfcal**, multi-frequency (MF) calibration for **SMA** wide-band (WB) data.

2. Bandpass solver **SMAMfcal** -

The program **smamfcal** is a **Miriad** task which determines calibration solutions for antenna-based corrections in aspects of antenna gains, delay terms and passband shapes from a multi-frequency observation, which is developed based on the original **Miriad** program **mfcal** coded by Bob Sault according to the **Miriad** code signature. In the past decade, **smamfcal** has been powered gradually by implementing micro-processes for handling data from observations at submillimeter wavelengths. The algorithm of matrix solver of **mfcal** is retained in **smamfcal**. The micro-processes used prior to or after solving for antenna-based bandpass and delay are listed below along with descriptions and discussions in details:

2.1 Weight

In addition to equal weighting, two mechanisms of visibility weighting have been implemented as a micro-process: 1) use variance (σ^2) of visibility; 2) use amplitude (A) of visibility, usually with the n-th power of the amplitude (A^n and $n>1$). The second method is dependent of source structure, less weighting the contribution from the visibilities with lower S/N data due to source structure in a resolved or a partially-resolved source. A subroutine **accumwt** accumulates the visibility weight in parallel to the visibility accumulation with one-to-one mapping for each data point. Three options have been provided for weighting in **smamfcal**:

1. weight = $1/\sigma^2$ or 1, the same fashion as that used in **mfcal**
2. weight = $1/(A/\sigma)^2$
3. weight = $1/(A/\sigma)^4$

Option 3 can effectively give less weighting to the lower S/N visibility near nulls for a disk-like object.

2.2. Normalization

Solving bandpass is usually carried out before phase correction made. A pseudo continuum visibility is computed by vector average of each spectral chunk given an integration. A normalization of the visibility in each spectral channel is carried out with the pseudo continuum visibility. Thus, such a self-normalization can remove any temporal phase drifts while retain bandpass variations across a spectral chunk. The normalization is accomplished with two subroutines **avgchn** and **divchz**. The former determines the vector average of spectral visibilities in the basis of per-baseline and per-chunk given an integration. Then, a normalized spectrum is computed with later.

2.3. Smooth and LSQ fit

The subroutine **smoothply** hosts two micro-processes, i.e. moving smooth^[3] and orthogonal polynomials' fitting^[3] to each of the visibility spectra prior to solving for bandpass. Options **msmooth** and **opolyfit** are given for users to select one of the two micro-processes for handling poor S/N spectral data. For high spectral resolution data, such as **SWARM** data, **msmooth** appears to be a better choice for the reasons discussed in details as follows.

For a spectral chunk with a total number of channels n , the complex quantity in each channel visibility consists of two parts, the true value of the measured quantity η and a measurement error ε ,

$$y_i = \eta_i + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (1)$$

We assume that η_i is a polynomial in frequency ν and the error ε_i to be normally distributed about zero. In the high-spectral resolution case $\Delta\nu \sim 0.1$ MHz, such as in the spectral production from the **SMA SWARM** correlator, the variation trend η_i is often embedded in the fluctuation of error ε_i for most of QSOs at submillimeter wavelengths. Thus a smoother function of ν is needed to replace every value of y_i ,

$$u_i = \sum_{j=i-k}^{i+k} y_j / (2k+1), \quad (2)$$

If η is a linear function of ν in the frequency interval $\Delta\nu = \nu_{i+k} - \nu_{i-k} = (2k+1) \Delta\nu$,

$$\eta_j = \alpha + \beta\nu_j, \quad j = -k, -k+1, \dots, k \quad (3)$$

α and β are constant. They can be determined from the data by linear regression. For a higher order polynomial η ,

$$\eta_j = a_1 + a_2\nu_j + a_3\nu_j^2 + \dots + a_{l+1}\nu_j^l, \quad (4)$$

The coefficients of the polynomial function can be determined from the data by least square fitting to the smoother function u_i , Eq(2).

If the variance of y_i is σ^2 , then the variance corresponding to the smoother function u_i is,

$$\sigma_u^2 = \sigma^2 / (2k+1), \quad (5)$$

The ratio of the variance from the original data to that of the smoother function is proportional to $(2k+1)$,

$$\sigma^2 / \sigma_u^2 = 2k+1, \quad (6)$$

For $k=5$, the variance of a spectral smoother is an order in magnitude smaller than that of the original spectrum. The default provided in **smamfcal** for moving smooth is $k=3$ and linear regression $l=1$. For the **SWARM** spectra, the linear regression approaching appears to be appropriate for even larger value of k . For the trial data used in Section 5, a smoother with $k=7$ is applied. The pre-process for **options=msmooth** in **smamfcal** should be adequate in solving for the bandpass trend of a telescope with data taken from a relatively weak calibrator. For **SMA SWARM** high-spectral resolution chunks, a smoother with $k=10$ can greatly enhance S/N ratio of the spectral data while a slow variation of bandpass response is retained. Thus with the pre-process of moving smooth, we can correct for the bandpass of **SWARM** spectral data with a high confidence.

2.4. Auto-editing BP solution

Solutions of bandpass may be contaminated due to individual channels with large errors, e.g. central spikes produced from an instrument. A function for

auto-editing of bandpass solutions is being implemented, which rejects solutions that are highly deviated from the actual bandpass trend; and replacements can be made with the values derived from orthogonal polynomials' fitting to overall solutions.

3. Help deck -

Task: [smamfcal](#)

Responsible: Jun-Hui Zhao

[SmaMfCal](#) is a [Miriad](#) task which determines calibration corrections (antenna gains, delay terms and passband shapes) from a multi-frequency observation. The delays and passband are determined from an average of all the selected data. The gains are worked out periodically depending upon the user selected interval. [SmaMfCal](#) implements algorithms for weighting, continuum vector normalization, and moving smooth prior to solving for bandpass and gains, which are necessary for handling data at submillimeter wavelength when the S/N is poor and phase dispersion is large. The basic solving algorithms are the same as in [MfCal](#).

Keyword: vis

Input visibility data file. No default. This can (indeed should) contain multiple channels and spectral windows. The frequency set-up can vary with time.

Keyword: line

Standard line parameter, with standard defaults.

Keyword: edge

The number of channels, at the edges of each spectral window, that are to be dropped. Either one or two numbers can be given, being the number of channels at the start and end of each spectral window to be dropped. If only one number is given, then this number of channels is dropped from both the start and end. The default value is 0.

Keyword: select

Standard uv selection. Default is all data.

Keyword: flux

Three numbers, giving the source flux, the reference frequency (in GHz) and the source spectral index. The flux and spectral index are at the reference frequency. If not values are given, then [SmaMfCal](#) checks whether the source is one of its known sources, and uses the appropriate flux variation with frequency. Otherwise the default flux is determined so that the rms gain amplitude is 1, and the default spectral index is 0. The default reference frequency is the mean of the frequencies in the input data. Also see the 'oldflux' option. (This function has not been implemented for [SMA](#) data).

Keyword: refant

The reference antenna. Default is 3. The reference antenna needs

to be present throughout the observation. Any solution intervals where the reference antenna is missing are discarded.

Keyword: minants

The minimum number of antennae that must be present before a solution is attempted. Default is 2.

Keyword: interval

This gives one or two numbers, both given in minutes, both being used to determine the extents of the gains calibration solution interval. The first gives the max length of a solution interval. The second gives the max gap size in a solution interval. A new solution interval is started when either the max times length is exceeded, or a gap larger than the max gap is encountered. The default max length is 5 minutes, and the max gap size is the same as the max length.

Keyword: weight

This gives different ways to determine weights (wt) prior to solving for bandpass:

- 1 -> wt = 1; the same weighting method as used in *MfCal*.
- 1 -> wt ~ amp0**2/var(i); for a normalized channel visibility, the reduced variance is proportional to amp0**2/var(i), where amp0 is the amplitude of the pseudo continuum and var(i) is the variance of visibility for the ith channel.
- 2 -> wt ~ amp0**4/var(i)**2;

Default is 2 for *SMA* and -1 for other telescopes.

- if you have stable phase, use -1;
- if the phase stability is poor, use 1 or 2;
- for a larger planet, 2 is recommended.

For antenna gains' solver:

- 1 -> wt = 1; the same weight method that is used in *MfCal*.
- >0 -> wt = 1/var, where var is the visibility variance.

Default is 1/var.

Keyword: options

Extra processing options. Several values can be given, separated by commas. Minimum match is used. Possible values are:

- delay Attempt to solve for the delay parameters. This can be a large sink of CPU time.
- nopassol Do not solve for bandpass shape. In this case if a bandpass table is present in the visibility data-set, then it will be applied to the data.
- interpolate Interpolate (and extrapolate) via a spline fit (to the real and imaginary parts) bandpass values for channels with no solution (because of flagging). If less than 50% of the channels are unflagged, the interpolation (extrapolation) is not done and those channels will not have a bandpass solution.
- oldflux This causes *SmaMfCal* to use a pre-August 1994 *ATCA* flux density scale. See the help on "oldflux" for more information. (not relevant to *SMA* data)
- msmooth Do moving average of the uv data (the real and imaginary parts) using the keyword smooth parameters

specified prior to solving for bandpass.

`opolyfit` Do least-square fit to the bandpass solutions (the real and imaginary parts) with an orthogonal polynomial of degree `n` which can be given in keyword `polyfit`.

`wrap` Don't unwrap phase while do fit or smooth the uv data.

`averrll` In the case of solving for bandpass of dual polarizations, `averrll` gives vector average of `rr` and `ll` bandpass solutions; the mean value is written into the bandpass table for each of `rr` and `ll`.

Keyword: smooth

This gives three parameters of moving smooth calculation of the bandpass/gain curves:

`smooth(1) = K` parameter `k` giving the length `2k+1` of the averaging interval; default is 3.

`smooth(2) = L` order of the averaging polynomial `l`; default is 1.

`smooth(3) = P` probability `P` for computing the confidence limits; default is 0.9.

Keyword: polyfit

`polyfit` gives a degree of orthogonal polynomial in least-square fit to the bandpass/gain curves. Default is 3.

`polyfit: 1` (linear), `2` (parabolic), `3` (cubic),

Keyword: tol

Solution convergence tolerance. Default is 0.001.

4. Usage -

The reduction of **SMA** data has been suggested to use ***msmooth*** in bandpass calibration, e.g. **[CalibrationProcedure](#)**[£] for data produced from the **ASIC** correlator. The function ***msmooth*** appears to be critical in solving high-spectral resolution bandpass of **ASIC** data. This micro-process appears to be inevitable for handling the high-spectral resolution **SWARM** data. Here, an example is given for usage of ***smamfcal*** in a C-shell script if one uses a QSO as a bandpass calibrator:

```
#!/bin/csh -f
set fname = 160211_rx0.lsb
set bcal   = 3c273
set edge   = 500
set refant = 1
set smooth = 7,1,0.9
... (see CalibrationProcedure£)
smamfcal vis=$fname.swarm.tsys select='source('$bcal$')' edge=$edge,$edge \
  refant=$refant interval=1000000 options=msmooth smooth=$smooth
```

[£]https://www.cfa.harvard.edu/sma/miriad/swarm/pscripts/swarmcali_a.csh.html

5. Application & Testing -

Using a recent test data from the observation track on February 11, 2016 with a hybrid configuration of both **ASIC** and **SWARM** correlators. This is the first real data set that was produced from the 2Q **SWARM** correlator, consisting of four high-resolution chunks while the **SMA** was in an array of five antennas. Here is a report for the lower-side band (LSB) data set, which is produced with **uvindex**:

5.1. Description of the testing track on 2016-02-11

UVINDEX: version 5-sept-2013

Summary listing for data-set 160211_rx0.lsb

```
-----
      Time           Source      Antennas Spectral Wideband Freq Record
      Name           Name           Channels Channels Config No.
16FEB11:17:20:13.8 3c273           8      71681      0      1      1
16FEB11:17:50:54.1 nrao530          8      71681      0      1     611
16FEB11:18:10:05.3 sgrb2n           8      71681      0      1    1991
16FEB11:18:25:55.1 nrao530          8      71681      0      1    2301
16FEB11:18:31:21.6 sgrb2n           8      71681      0      1    2411
16FEB11:18:46:41.8 nrao530          8      71681      0      1    2721
16FEB11:18:52:08.3 sgrb2n           8      71681      0      1    2831
16FEB11:19:07:28.5 nrao530          8      71681      0      1    3141
16FEB11:19:12:55.0 sgrb2n           8      71681      0      1    3251
16FEB11:19:28:15.2 nrao530          8      71681      0      1    3561
16FEB11:19:33:41.7 sgrb2n           8      71681      0      1    3671
16FEB11:19:49:01.8 nrao530          8      71681      0      1    3981
16FEB11:19:54:28.4 sgrb2n           8      71681      0      1    4091
16FEB11:20:09:48.5 nrao530          8      71681      0      1    4401
16FEB11:20:15:15.0 sgrb2n           8      71681      0      1    4511
16FEB11:20:30:35.2 nrao530          8      71681      0      1    4821
16FEB11:20:36:01.7 sgrb2n           8      71681      0      1    4931
16FEB11:20:51:21.9 nrao530          8      71681      0      1    5241
16FEB11:20:54:20.0 Total number of records                               5310
-----
```

Total observing time is 3.55 hours

The input data-set contains the following frequency configurations:

Frequency Configuration 1

Spectral Channels	Freq(chan=1)	Increment
1	218.94975	-0.416000 GHz
128	220.94094	-0.000812 GHz
128	220.85894	-0.000812 GHz
128	220.78375	-0.000812 GHz
128	220.70175	-0.000812 GHz
128	220.61294	-0.000812 GHz
128	220.53094	-0.000812 GHz
128	220.45575	-0.000812 GHz

128	220.37375	-0.000812 GHz
128	220.28494	-0.000812 GHz
128	220.20294	-0.000812 GHz
128	220.12775	-0.000812 GHz
128	220.04575	-0.000812 GHz
128	219.95694	-0.000812 GHz
128	219.87494	-0.000812 GHz
128	219.79975	-0.000812 GHz
128	219.71775	-0.000812 GHz
128	219.62894	-0.000812 GHz
128	219.54694	-0.000812 GHz
128	219.47175	-0.000812 GHz
128	219.38975	-0.000812 GHz
128	219.30094	-0.000812 GHz
128	219.21894	-0.000812 GHz
128	219.14375	-0.000812 GHz
128	219.06175	-0.000812 GHz
128	218.94094	-0.000812 GHz
128	218.85894	-0.000812 GHz
128	218.78375	-0.000812 GHz
128	218.70175	-0.000812 GHz
128	218.61294	-0.000812 GHz
128	218.53094	-0.000812 GHz
128	218.45575	-0.000812 GHz
128	218.37375	-0.000812 GHz
128	218.28494	-0.000812 GHz
128	218.20294	-0.000812 GHz
128	218.12775	-0.000812 GHz
128	218.04575	-0.000812 GHz
128	217.95694	-0.000812 GHz
128	217.87494	-0.000812 GHz
128	217.79975	-0.000812 GHz
128	217.71775	-0.000812 GHz
128	217.62894	-0.000812 GHz
128	217.54694	-0.000812 GHz
128	217.47175	-0.000812 GHz
128	217.38975	-0.000812 GHz
128	217.30094	-0.000812 GHz
128	217.21894	-0.000812 GHz
128	217.14375	-0.000812 GHz
128	217.06175	-0.000812 GHz
16384	212.79985	0.000102 GHz
16384	217.09965	-0.000102 GHz
16384	216.79985	0.000102 GHz
16384	221.09965	-0.000102 GHz

The input data-set contains the following polarizations:
 There were 5310 records of polarization XX

The input data-set contains the following pointings:

Source	RA	DEC	dra(arcsec)	ddec(arcsec)
3c273	12:29:06.70	+02:03:08.60	0.00	0.00
nrao530	17:33:02.70	-13:04:49.55	0.00	0.00
sgrb2n	17:47:19.88	-28:22:18.38	0.00	0.00

This is a 4hr observation track pointing on sources of between 3c273 (bandpass calibrator), nrao530 (gain calibrator) and sgrb2n (Sgr B2 North, target). The data consist of a total of 52 spectral chunks, excluding the first single channel of the pseudo continuum chunk. The first 48 uniform spectral resolution chunks (0.812 MHz per channel, 128 channels per chunk) are produced from the legacy **ASIC** correlator. Highlighted in red, the bottom four high spectral resolution chunks (0.102 MHz per channel, 16384 channels per chunk) are produced from the first 2Q **SWARM**.

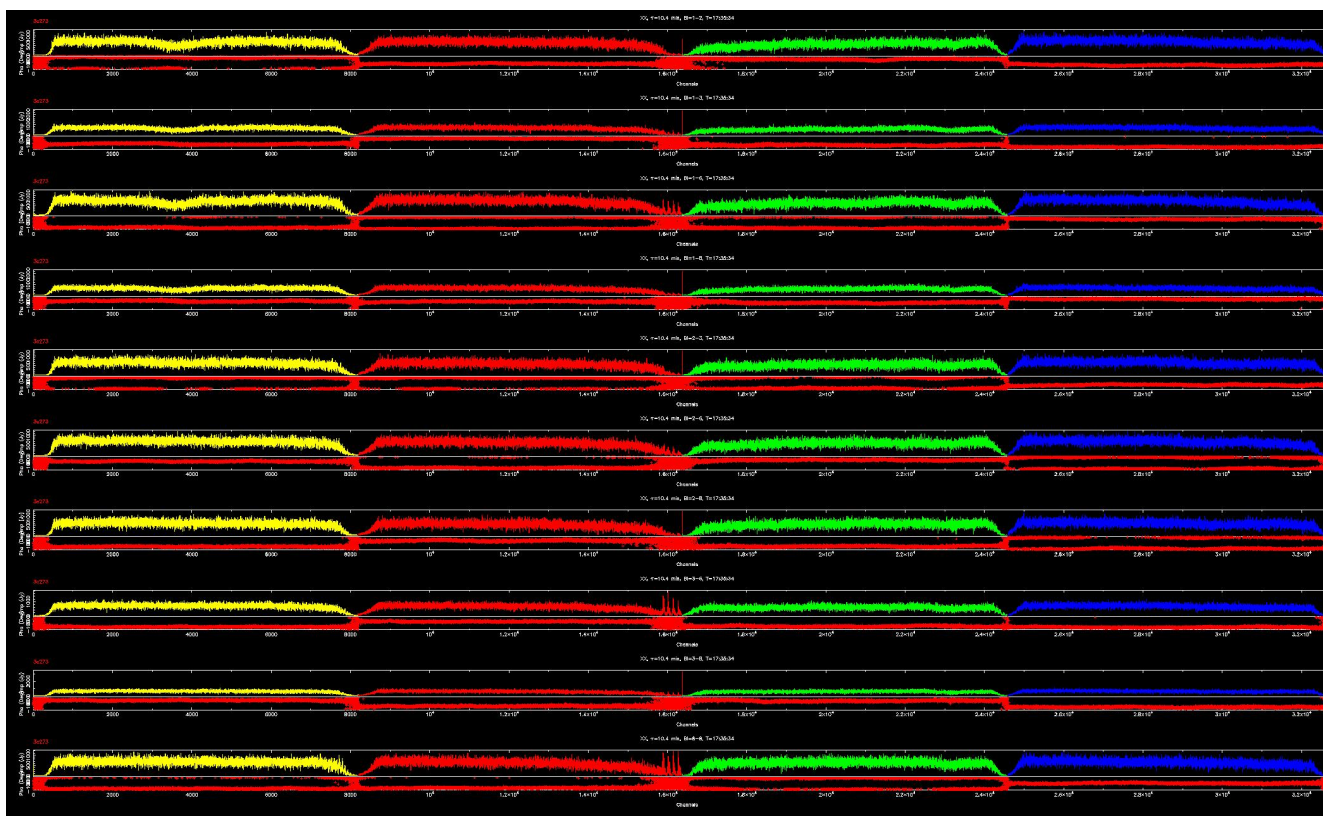


Fig. 1 - Raw spectra of 3C 273 from the ten baselines. Each row consists of four spectral chunks produced from the first 2Q **SWARM** correlator. Each of the spectra shows the amplitudes (top) and phases (bottom). Click the figure for enlargement (for webpage version only)€.

Below, using the data extracted for the four high spectral resolution **SWARM** chunks, we demonstrates the process of bandpass solving with **smamfcal**, showing the results from the **Miriad** bandpass solver **smamfcal**.

Using an **SMA** pre-process program **swarmsplt**, we binned every two channels together before solving for bandpass. We note that the current code works for 8192 spectral channels per chunk, or the chunks binned every two channels from the full resolution (16384 channels) data due to the issues of the integer type and memory allocation/relocation for a large volume of data in some Fortran 77 programs. Otherwise, **Miriad** appears to be able to have provided a full capacity in handling **SWARM** data for **SMA** science.

5.2. Raw spectra of 3C 273 - bandpass calibrator

The beginning of the observation tracks was on 3C 273, bandpass calibrator, for 30 minutes. Fig. 1 shows a set of spectra of the four **SWARM** chunks from 10 baselines. Each **SWARM** chunk is marked in different color. 3C 273 at 1.3 mm is a strong point source (~ 11 Jy). Thus, weighting with visibility variance (default) would be good enough. Using the setup given as an example in Section 4, we run **smamfcal** to solve for bandpass.

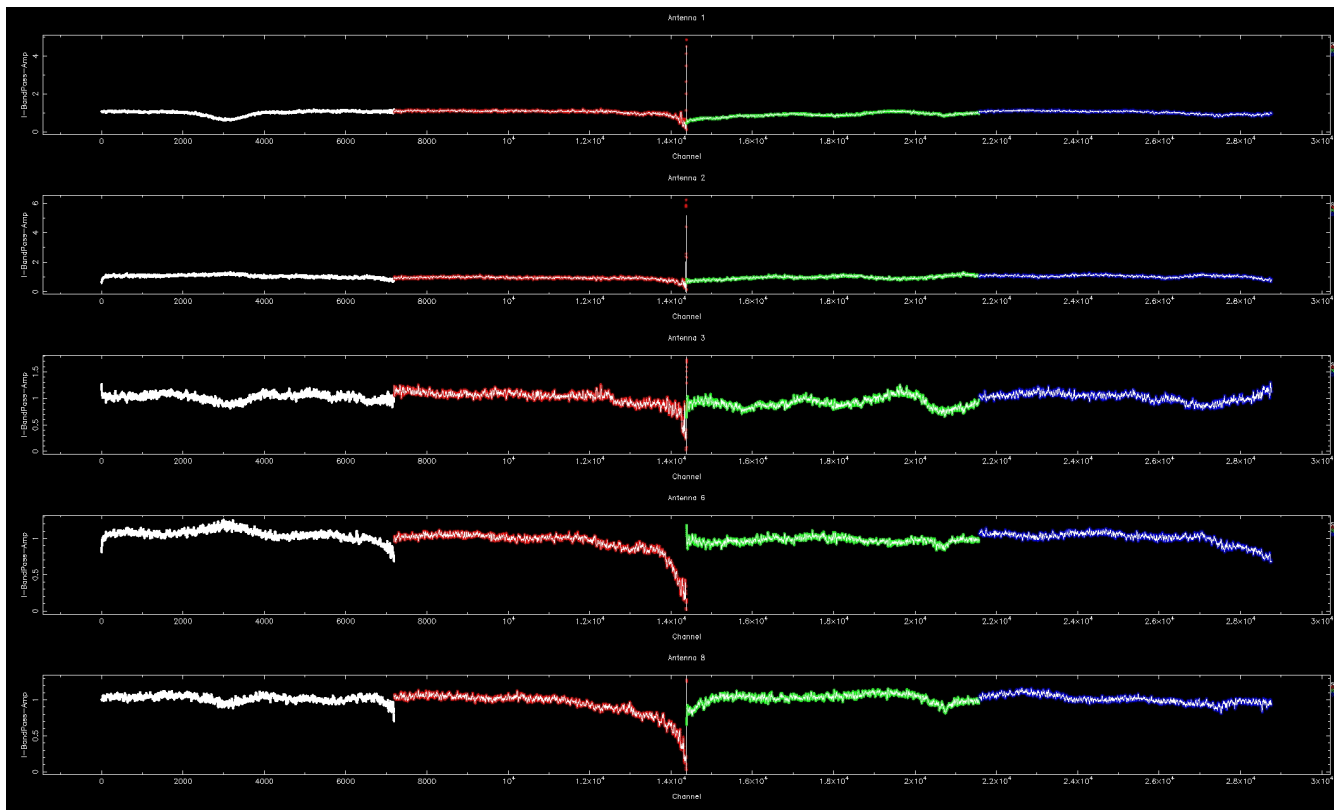


Fig. 2 - Antenna-based solutions in amplitude derived from the bandpass solver **smamfcal**. Click the figure for enlargement (for webpage version only)€.

5.3. Bandpass solutions

The bandpass solutions are obtained from the solver *smamfcal* in *Miriad*. Fig. 2 shows the antenna-based bandpass solutions in amplitude for the four **SWARM** chunks (color coded). We note that 500 channels, or 1000 channels for the data with the original channel resolution, have been cut off from both edges of each chunk. The chunk 2 at the large channel series number end may have more low signal channels that need to be thrown away.

Fig. 3 shows the antenna-based bandpass solutions in phase for the four **SWARM** chunks coded with the same color corresponding to the amplitude solutions. We note that the antenna 1 is used as reference and the rest of antennas show significant phase offsets between the spectral chunks while small phase variations appear in each spectral chunks. The alignment in phase is critical in continuum imaging. Applying the bandpass solutions to the target sources, the instrumental issues are taken care automatically.

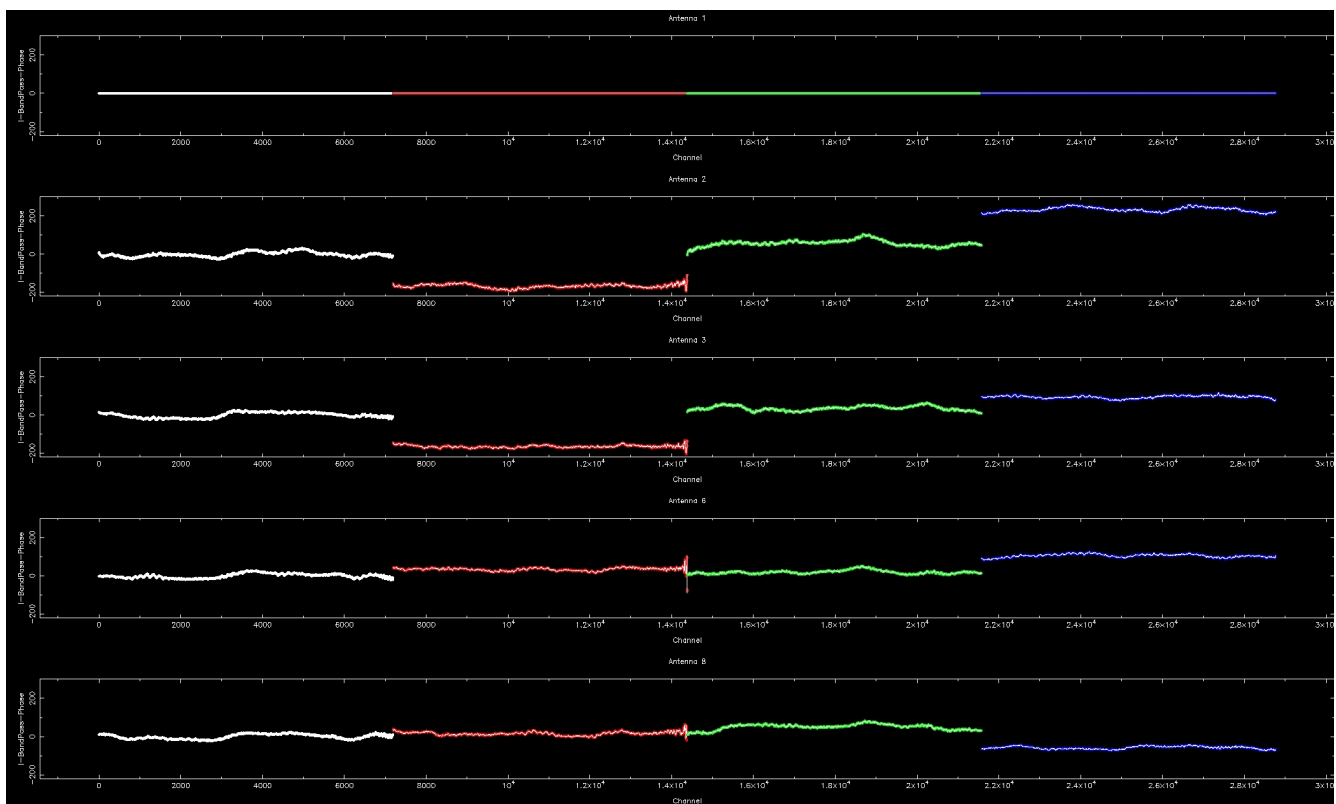
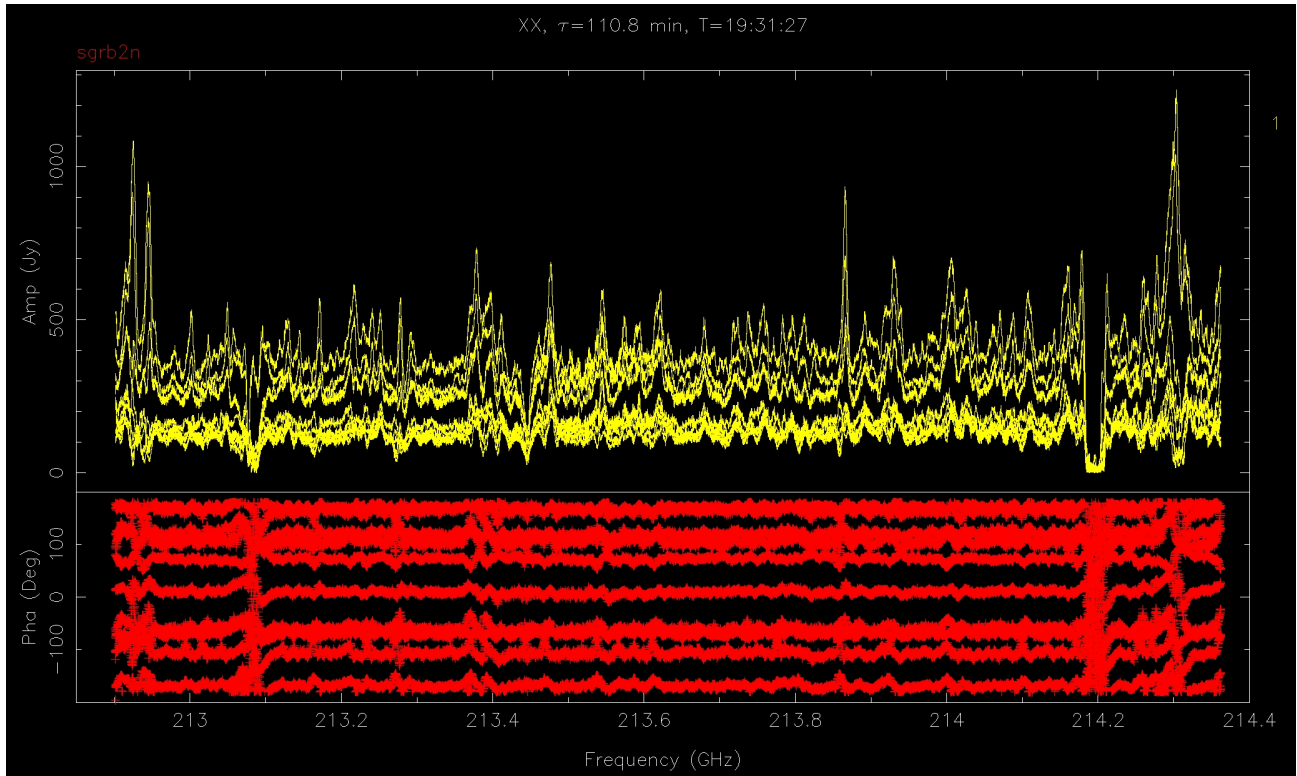
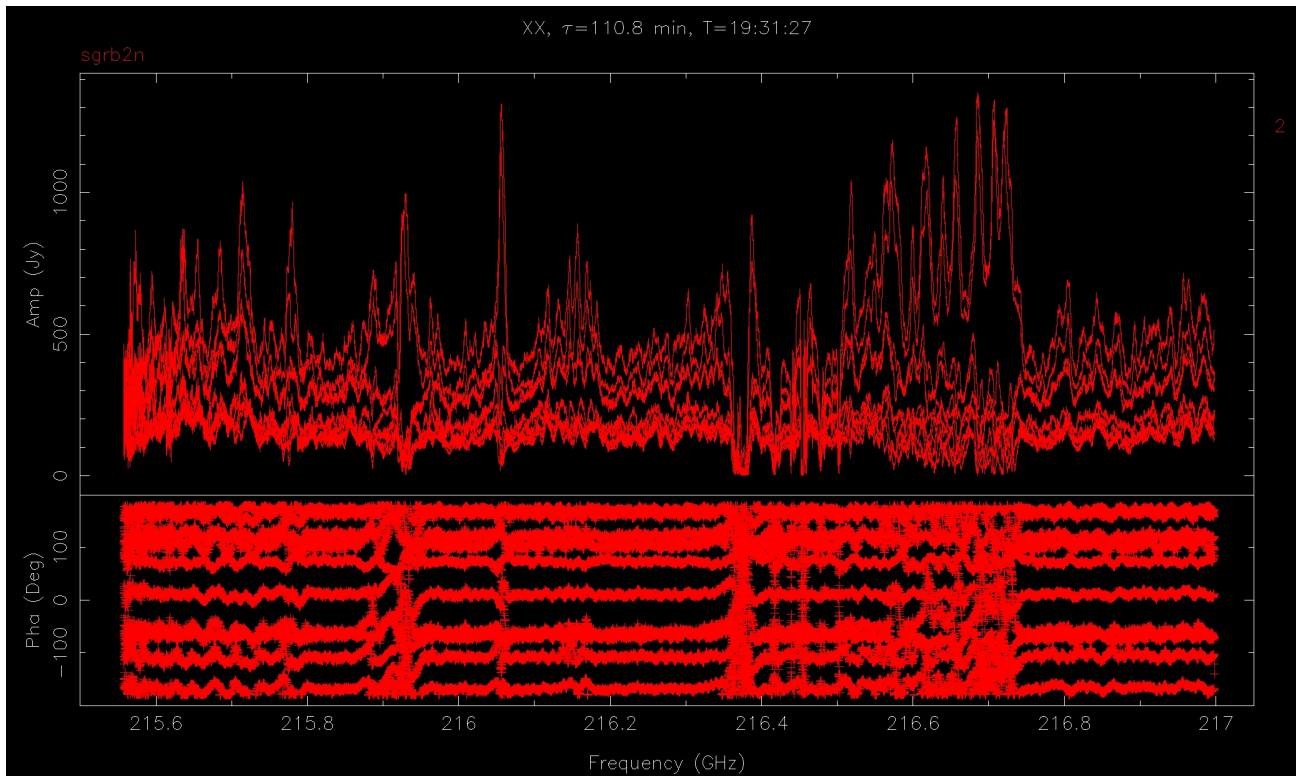


Fig. 3 - Antenna-based solutions in phase derived from the bandpass solver *smamfcal*, corresponding to the solutions in amplitude shown in Fig. 2. Antenna 1 is the reference antenna. Click the figure for enlargement (for webpage version only)€.

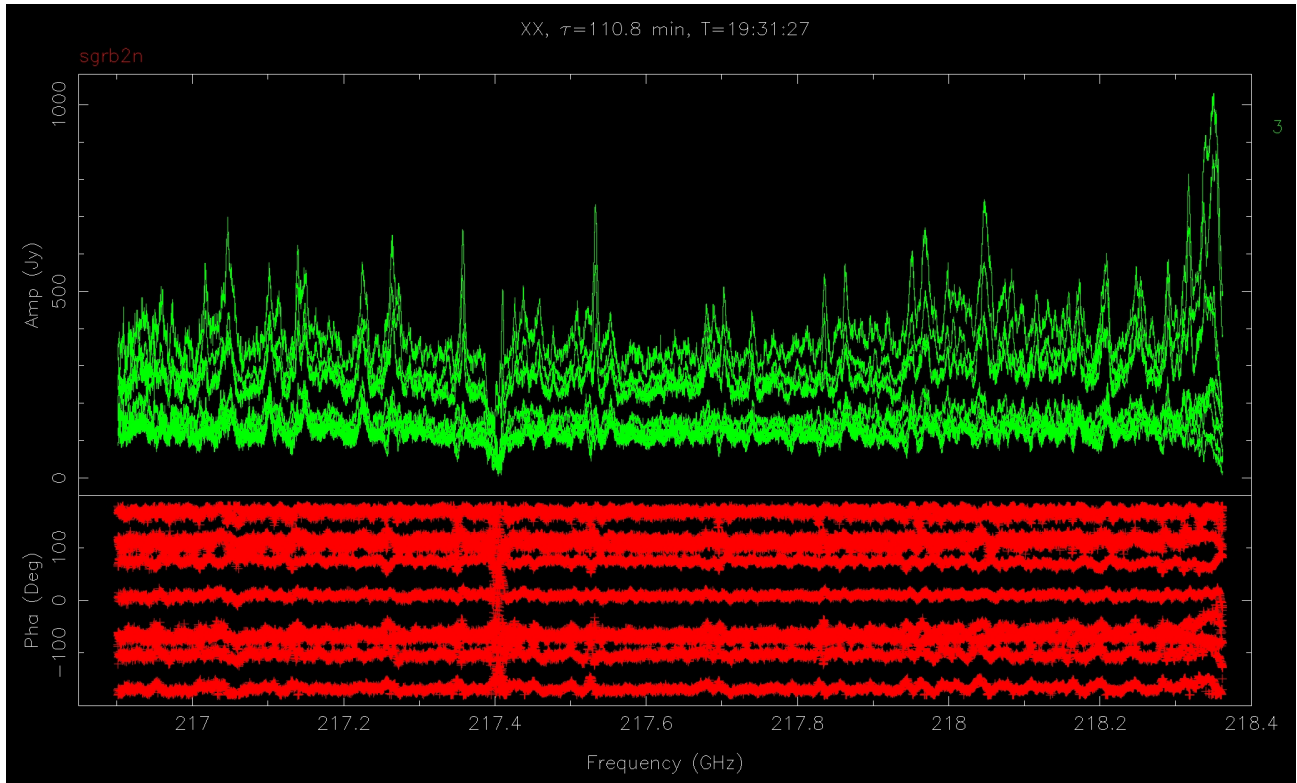
SWARM Chunk 1 -



SWARM Chunk 2 -



SWARM Chunk 3 -



SWARM Chunk 4 -

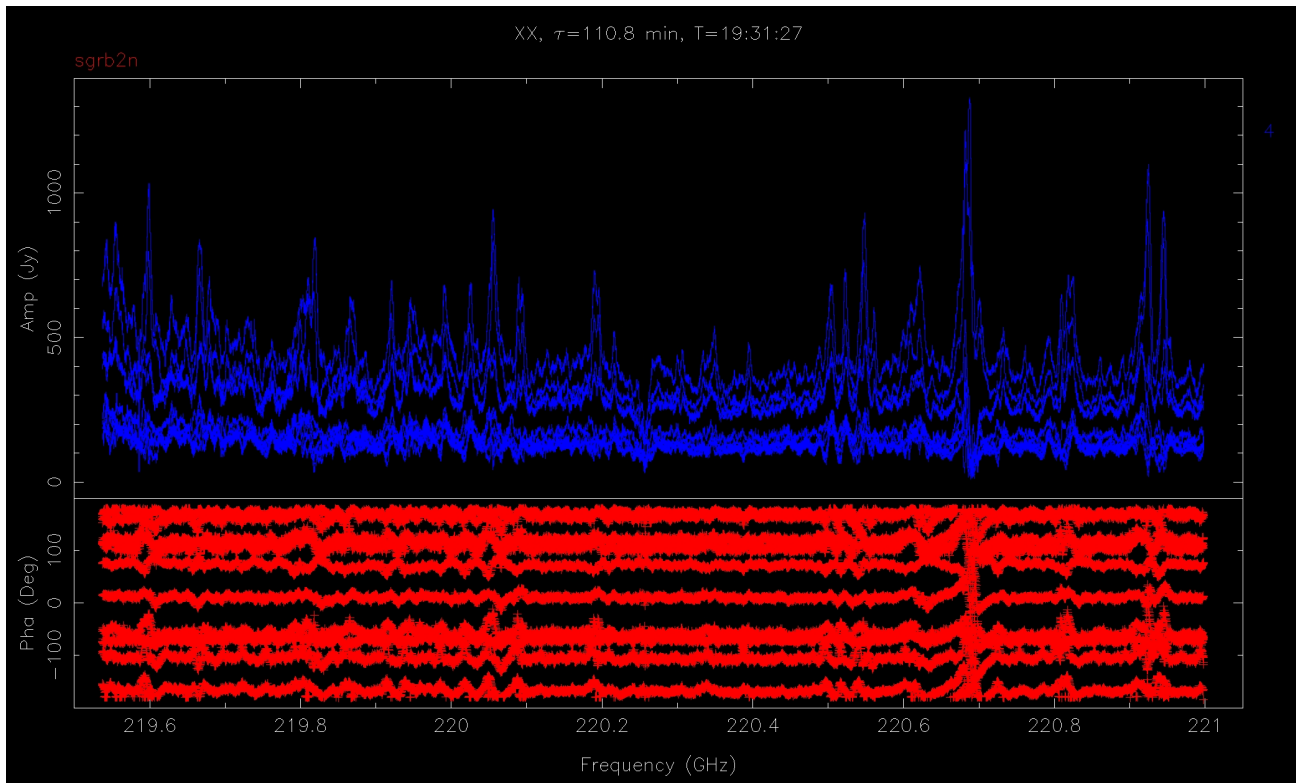


Fig. 4 - A stack of 10 spectra from each baseline in each of the four **SWARM** chunks towards Sgr B2 North after apply corrections for bandpass. From top to bottom are the spectral chunks 1, 2, 3 and 4 produced from the first 20 **SWARM** correlator. Click each of the figures for enlargement (for webpage version only)€.

5.4. Spectra of Sgr B2 North

We apply the bandpass solutions derived from 3C 273 to the target Sgr B2 North. Fig. 4 shows a stack of 10 corrected spectra from each baseline in each of the four **SWARM** chunks. A forest of molecular lines are present in the **SWARM** spectra, some show a broad absorption (for example, a feature at around 214.2 GHz in **SWARM** chunk 1) and some are very strong emission line. The hydrogen recombination line H30 α is hidden in the line forest of the upper sideband (USB) spectra (not shown in this Memo).

Finally, we note that the scale of flux density and complex gain corrections have not been applied yet to the spectra of the target source as shown in Fig. 4.

Appendix: Source code -

The Fortran code of the entire **smamfcal** program exceeds 4600 coding lines. For interested users, a copy of the source code can be found from the **SMA Miriad** distribution. The code of the Fortran 77 subroutines for the relevant micro-processes is highlighted below:

```
accumwt - weight
avgchn - vector average
divchz - normalization
smoothply - moving smooth & orthogonal polynomial fit
```

Reference -

- [1]Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, in ASP Conf. Ser. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco, CA: ASP), 433
- [2]Basmah Riaz, 2016, personel communication based on her **SMA** project: 2015B-S044
- [3]Siegmond Brandt, 1999, *Data Analysis - Statistical and Computational Methods for Scientists and Engineers*, Third Edition, Springer-Verlag New York Inc.

€<https://www.cfa.harvard.edu/sma/miriad/swarm/SMAmemSWARMBpass/SMAmemSWARMBpass.html>