Data Comb 2019 Memo: Gaussian Simulations Working group

Ed Fomalont (National Radio Astronomy Observatory, Charlottesville, VA, USA & Joint ALMA Observatory, Santiago de Chile, Chile)
Edwige Chapillon (Laboratoire d'astrophysique de Bordeaux, Pessac & Institut de Radioastronomie Millimétrique, Saint Martin d'Hères, France)
Sébastien Muller (Onsala Space Observatory, Chalmers University, Sweden)
M. Carmen Toribio (Onsala Space Observatory, Chalmers University, Sweden)

September 17, 2019

This memo was prepared as part of the workshop "Improving Image Fidelity on Astronomical Data: Radio Interferometer and Single-Dish Data Combination," held on 12-16 Aug 2019 at the Lorentz Center in Leiden, The Netherlands.

1 Introduction

The goals of this working group are to investigate the performances of different combining methods on "well-controled" simulated data. We start with an existing dataset with ACA (7m) and/or 12m and/or TP observations pointings (.ms files, that can be provided by a PI) and replace the actual data by a simulated source. We have chosen to work with a set of Gaussians, with different sizes and locations across the observed field. We have discussed on possible corruptions to alter the quality of the data (to mimick real observations). Such corruptions could be thermal noise, pointing errors, flux calibration issues, change of relative weighting, impact of TP field of view, etc. One aspect that can be effectively tested is source emission extending outside of the observed field of view, which can lead to establish the appropriate size of the TP map to be observed in order to perform a successful array combination.

1.1 Making Simulated Data

The simulation package in these initial tests starts with an existing ALMA 7-m multi-field Measurement Set (MS) and/or corresponding 12-m multi-field MS. The user then produces an appropriate sky model that is defined by a set of Gaussian components of arbitrary size and position within or near the region covered by the multi-field Ms. This model should approximate the kind of sky emission expected: for example long thin Gaussians to approximate filaments and also sky emission somewhat outside of the mosaic field of view.

The visibility amplitude and phase corresponding to the sky model is then calculated for each u-v data point in each 7-m and 12-m pointing field in the multi-field MS's. This requires knowing the position of each pointing in the mosaic and the 7-m and 12-m apriori primary beam sensitivity. The existing data in the MS's are then replaced by this model visibility. The three data sets (TP image = sky model, 7-m, 12-m multi-field data sets are then the data sets that are needed for the current and developing reduction packages.

This simulation of the data can probably be made using the present ALMA simulation package—'simobserve' to produce the appropriate multi-field MS's, and 'fakeobs' to generate the visibility data associated with an input set of Gaussian models. A short-term goal is to confirm the accuracy of the ALMA sim package and to describe its use. This will be part of the documentation for TP+ACA+12m array data combination that will be available for cycle 8 proposals.

1.2 Proposed Tests with the Simulated Data

The simulated data will then be used in the ALMA software procedures that are, or will soon be, available in order to produce images over the a large area of sky. Since the input sky model is known, the accuracy of the ALMA processing can be estimated. A initial list of anticipated comparisons are:

- What is the accuracy of the ALMA processing, assuming no noise in the data? Are the limits associated with imaging grid effects? Can edge effects and loss of integrated flux density be understood better?
- Addition of thermal antenna-based noise to the simulated data, both the sky model and interferometric data. Does the image rms increase as expected? (See Figure 4).
- Addition of typical tropospheric phase changes associated with the interferometric data. Effects on the final image.
- How easily can these tropospheric delay changes be removed by self-calibration techniques even in the presence of large-scale emission?
- What are the effects/improvements of self-calibration of the 12-m data when you have information on other scales?
- Estimate the effect of pointing errors for the interferometric data
- What are the ultimate image affects associated with inaccuracies in the shape/cutoff/sidelobe of the 7-m and 12-m antenna beams.
- How to deal with emission outside of the mosaic regions? Perhaps make the TP image larger? Small fields of view around each bright source.
- What are reasonable TP image errors that are obtained from rasters of the sky over the relevant areas. In addition to receiver noise, what are errors associated with the 'background' subtraction around the relevant field of view? How important are these in the final processed images?

2 Dataset overview

For this particular exercise we used a simulated dataset with a set of Gaussians with properties as given in Table 1. Image shown in Fig 1.

Table 1: List of Gaussians used for the simulated image.

RA Offset (arcsec)	Dec Offset (arcsec)	Peak flux	Major (arcsec)	Minor (arcsec)	P.A. (deg)
-30	-15	1.0	3.0	3.0	0.0
-19	35	1.0	3.0	3.0	0.0
19	-22	1.0	3.0	3.0	0.0
22	0	1.0	3.0	3.0	0.0
22	0	0.04	10.0	10.0	0.0
-10	0	0.15	12.0	2.0	90.0
14	18	0.02	20.0	15.0	150.0

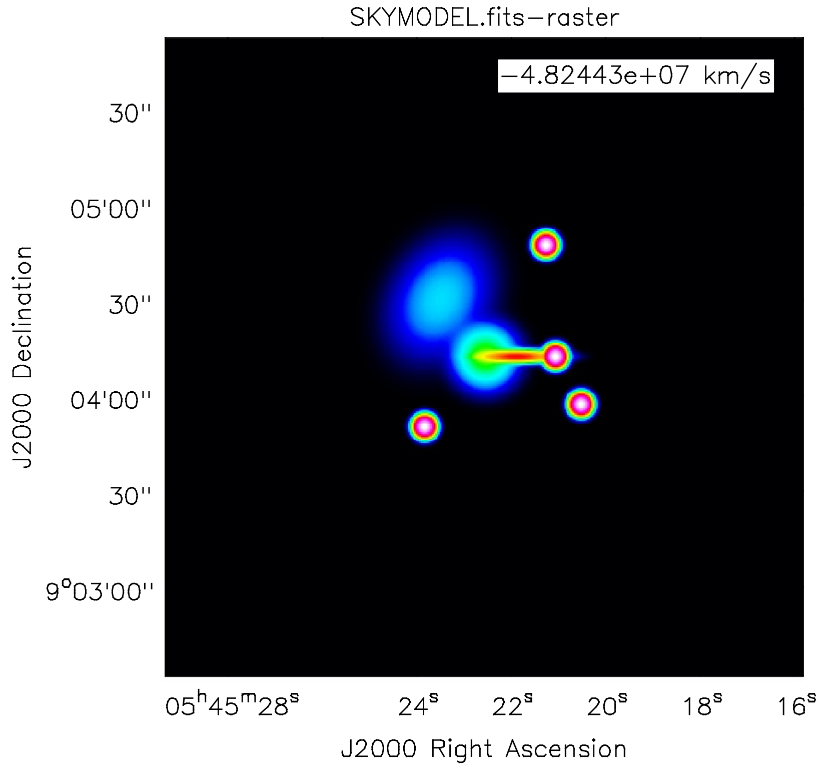


Figure 1: Image of the sky model.

The simulated observations of the skymodel consist in a mosaic of 7 pointings observed with ACA, as shown in Fig. 2. Figure 3 shows the weights of these ACA simulated observations.

2.1 Introducing errors in the observations

The first straightforward implementation was to introduce Gaussian random noise **per visibility**. This causes decorrelation (flux losses, etc) in the observations.

However, we acknowledge the approach we should adopt for being consistent with how an interferometer intrinsically works would be to introduce errors **per antenna**. For example, introduce antenna position errors, that affect subsequent calibration tables, etc.

In the following, we list several possible corruptions to apply to the simulated data and ways to implement them:

- TP flux calibration: multiply the TP map by a given factor
- TP pointing errors: shift the whole TP map in a given direction.
- Thermal noise: add a random phase/amplitude to visibilities.
- Change the relative weights of TP/ACA/Main Array.
- Test impact of incomplete mapping of an emission region, i.e. source emission located out of the observed field of view.

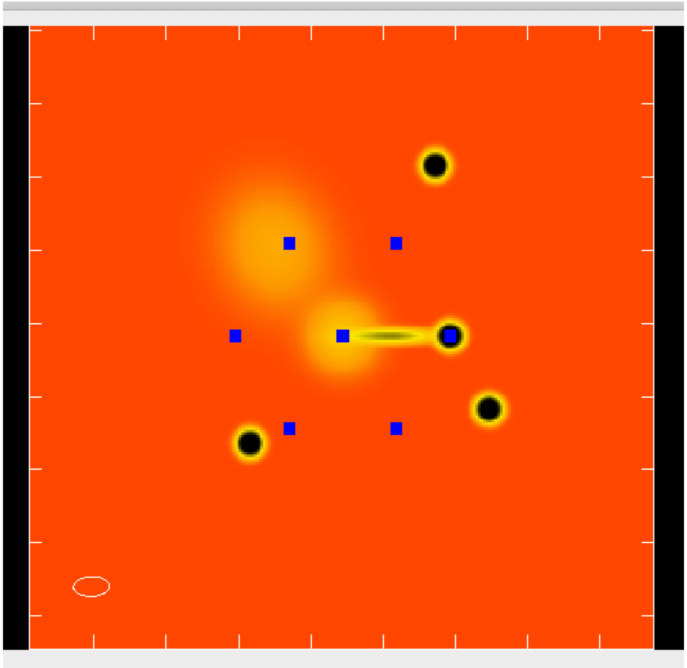


Figure 2: Input model; ACA pointings in blue

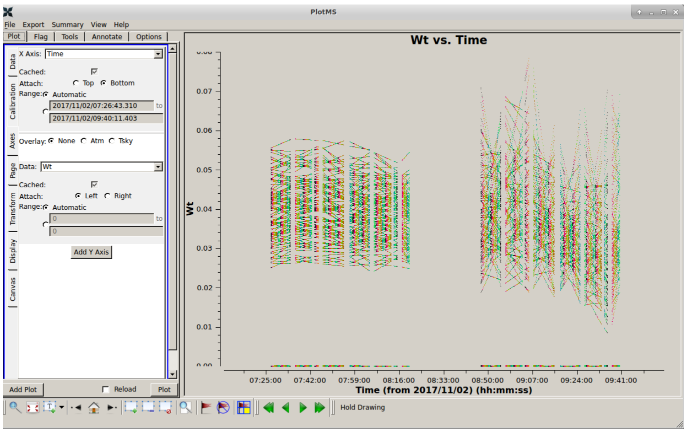


Figure 3: Input model; ACA weights

3 Combination Methods

3.1 Method 1: Feather

The team could not perform any testing on the feathering method yet. This is work in progress.

3.2 Method 2: Joint Deconvolution (tp2vis)

We have used our simulated data to test `tp2vis`.

Our SD input image is in units [Jy/pixel], therefore, we not should call tp2vis with the default parameter `deconv`, but set `deconv=True`.

`tp2vis` uses as an input a list of pointings to recreate the SD visibilities (3rd argument of the

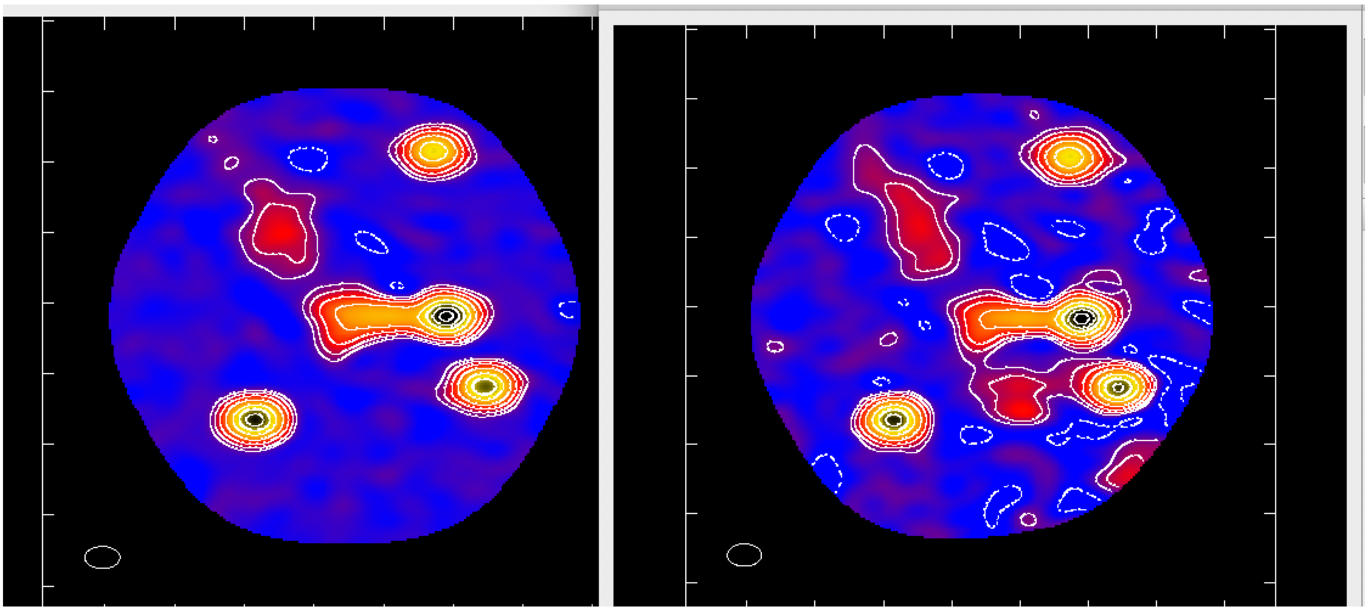


Figure 4: Left: 7-m mosaic image with noiseless data. Lowest contour is 2% of peak. Right: 7-m mosaic image with significant thermal noise added. Lowest contour is also 2% of peak. But, peak is 0.6 times that of noiseless image.

task). We are facing a limitation with respect to this: Due to CASA limitation, `tp2vis` currently uses 12m primary beam for mosaic, so the pointing spacing should be as dense as the one for 12m.

Since we are trying to combine TP+7m mosaic only, if we extract the ACA pointings from the MS, these would be too sparse for a 12m dish mosaic.

Therefore, we should use instead a 12-m pointing pattern in that region of the sky. The same original ALMA dataset that was used to generate the Gaussian Model belongs to a project that also included 12-m observations. We used the pointings for those runs as input for `tp2vis`.

Another aspect to explore is the `rms`. Our first perfect models did not include any RMs. Therefore, we called `tp2vis` with `rms=0`. Plots presented here (see Fig. 5) correspond to this case.

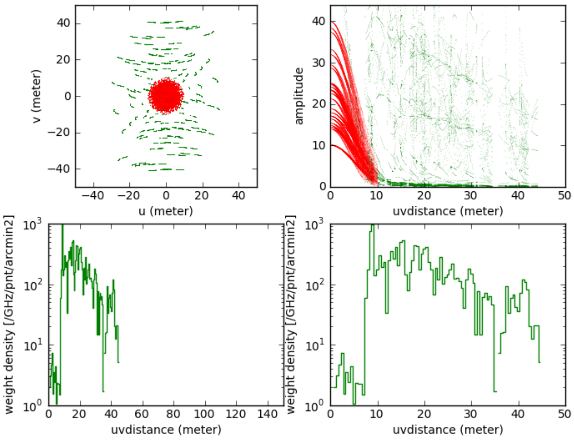
As we can see from the diagnostic plots of `tp2vispl`, there is a clear mismatch between the weights of TP and SD visibilities (The expected ratio beween weights should not be so many orders of magnitude). We are investigating if the origin of this mismatch is a result of the way the data was simulated. For that, we will compare with the use of 'simobserve'. This is work in progress.

3.3 Method 3: SDint

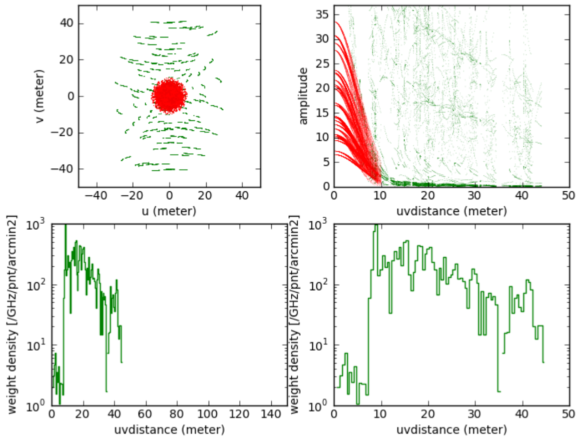
In order to run SDint, we need to transform our SD input image to units [Jy/bm], as well as generate the corresponding PSF. The team has found it challenging to follow installation and documentation for SDint to work on our simulated data. This is work in progress.

4 Next steps and future plan

The aim of this exercise is to contribute to the recommendations for the best approaches to perform data combination of different arrays.



(a) tp2vis plots when using SD image without taking into account te 7-m mosaic PB.



(b) tp2vis plots when using SD image that includes a 7-m mosaic PB.

Figure 5: Output plot of tv2vispl for the two different SD images we used

In order to do that, we should finalize the data combination following the methods provided at the workshop (feather, tp2vis and SDint), and compare and evaluate the results. The exercise can be expanded to a comparison with the tasks available in the package GILDAS.

Our simulations should also help to identify the most suitable indicators of a satisfactory data combination.

Here we will list our most immediate steps and the future plan:

- Perform data combination for the 3 different proposed methods, an expand to GILDAS package.
- Compare and evaluate the performance of the methods.
- Use as input the result of 'simobserve'/'fakeobs'
- Introduce the different type of effects that can affect real observations (see Section 2.1)
- Identify suitable indicators for array combination performance
- Provide recommendations based on the output of proposed tests (see Section 1.2)
- Generalize the scripts to allow the generation of user-defined models and make them available to the community

5 Appendix: Comments on tclean with Automasking

When setting up tclean with automasking, there is a blob of emission not captured (see Fig.6). The parameters for automasking were set up according to the recommendations of the CASA Guide (https://casaguides.nrao.edu/index.php/Automasking_Guide). Further exploration of automasking parameters is needed.

Imaging was performed under CASA 5.6.0 with the rest of default parameters for tclean.

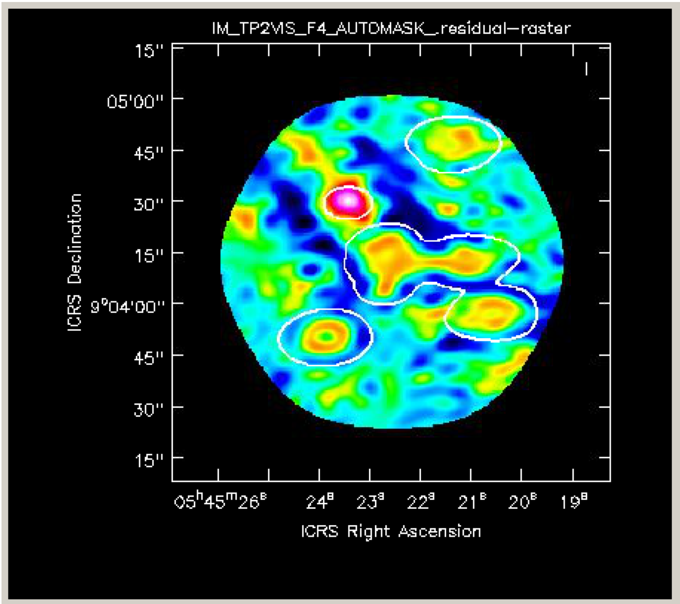


Figure 6: Cleaning boxes selected by automasking after 3(?) major cycles. The box surrounding the strongest peak of emission was manually added.