Next Generation VLA Memo 67: Demonstration & Analysis of ngVLA core + Short Baseline Array Extended Structure Imaging

Brian Mason (NRAO)

October 9, 2019 (v1)

Abstract

I simulated observations of a spatially extended ($\sim 5'$) astronomical source using the ngVLA core and Short Baseline Array (SBA), and reconstructed images from these data in CASA. For the 93 GHz center frequency of the simulated observations the largest spatial scale that can be recovered by the ngVLA 18-m core array is at most $\lambda/b_{min} \sim 19''$ and $\sim 55''$ for the SBA interferometer. Correct *total flux densities* are obtained within $\pm 1\%$ using the combination of the Short Baseline Array (SBA) with feathered total power from the 18m single-dish component of the SBA. To obtain these results it was necessary to carefully examine the simulated single-dish results and apply significant (nearly a factor of two) corrections to the simulated image calibration. I evaluate the current ngVLA image fidelity requirement and propose a refined definition of image fidelity. The image fidelity of the SBA+TP image is 95%, exceeding the 90% ngVLA image fidelity goal by any of the considered definitions of fidelity. Jointly deconvolved SBA+ngVLA core images also recover accurate total flux densities within 1% when total power data is included by feathering. However their image fidelity ($\sim$ 70%) does not yet meet the ngVLA requirement. Likely causes of this shortfall are: 1) primary beam errors in the 6m+18m joint interferometric deconvolution— a path remains to be identified by which heterogeneous array data *other than ALMA* can be correctly handled in current CASA versions; & 2) clean instabilities (divergences) which were evident in the joint array deconvolution. Because the ngVLA aims to deliver most of its products as Science Ready Data Products (SRDPs)— and because in the future we would like to extend this analysis to other methods and science targets— the simulations and imaging were done in a fully scripted, repeatable fashion. In particular the multi-scale deconvolution was done in `tclean` with the `auto-multithresh` heuristic for automated clean masking, the parameters of which I derived for SBA and the ngVLA core by modestly tweaking the ALMA default parameter values. I found that the combination of TCLEAN and `auto-multithresh` worked robustly for SBA data, giving no indication of divergences during the deconvolution and resulting in high-fidelity results as previously noted. This procedure worked less well for joint 6m+18m interferometric

data given parameter values I was able to identify. Finally, I evaluated the magnitude of the Jorsater-VanMoorsel Effect, finding it is $\sim 14\%$ for the SBA and $\sim 8\%$ for the ngVLA core (tapered by a factor of two in this case). Applying the correction using the standard prescription did not significantly improve the image fidelity. I comment on directions for future work, which include correctly handling the 6m-18m primary beams in the joint deconvolution, better tuning of the automask and tclean control parameters, and alternative deconvolution algorithms.

1 Introduction

The Next Generation VLA (ngVLA) reference design calls for 18-meter diameter antennas with offset Gregorian optics (Selina et al. 2018a, 2018b). One consequence of this design is that the shortest baselines available to the ngVLA main array will be 35 meters. The anticipated science targets of the ngVLA include numerous sources with spatial structures larger than what can be reliably recovered using this range of baselines (ngVLA SAC 2017; Selina, Murphy & Erickson 2017). To provide the capability of imaging these larger science targets, a Short Baseline Array (SBA) has been designed and is included in the reference design (Mason et al. 2018). As currently envisaged the SBA comprises an array of 19 6-m diameter antennas operating as an interferometer, and four 18-m total power antennas. The 6-m array provides baselines between 60 meters and 11 meters. This memo presents a demonstration and quantitative analysis of the imaging capabilities of the SBA together with the ngVLA core.

The scripts used to produce most of the results in this memo can be found at

`https://github.com/teuben/dc2019/tree/master/scripts/ngVlaSbaSims`

and the modified "30 Dor" input FITS image is at

`https://astrocloud.nrao.edu/s/ZPbz95tkTi4sSPr`

Please contact me (at NRAO email "bmason") to report any issues.

2 Simulations

The ALMA image simulation library[1] image of 30 Doradus— an HII region in the Large Magellenic Cloud, also known as the Tarantual Nebula— was chosen as a target template for these simulations because the object shows structure over a wide range of spatial scales. For purposes of these simulations the target's Declination was changed to $+22:38$ in order to be observable from the Northern hemisphere. The image was also re-gridded onto a $10\times$ higher spatial resolution grid in order that the cell size adequately sample the ngVLA core beam at a $2''$ taper; the number of pixels in each orthogonal axis was also increased by a factor of 10 so that the overall image size was unchanged. The image is originally a SPITZER 8μm continuum image. Early ALMA observations of a small region of 30 Doradus are presented in Indebetouw et al. (2013).

[1] `https://casaguides.nrao.edu/index.php?title=Sim_Inputs`

The simulations were done with custom `simobserve` scripts configured to observe a $4'.5 \times 4'.5$ region at 93 GHz, resulting in a 187-pointing mosaic with the ngVLA core and 23 pointings with the SBA. The simulated single dish maps covered a larger region ($6'.75 \times 6'.75$) in order to provide a "buffer region" around the interferometer image; regions much smaller than this tend to introduce artifacts in the `feather` step. The ngVLA core simulation comprised 187 10-second integrations, for a total integration time of 1870sec, and the SBA simulation comprised 23 180-second pointings (with $18\times$ $10sec$ integrations for each pointing) for a total integration time of 4140sec. This integration time ratio $t_{sba}/t_{core} = 2.2$ is appropriate to match sensitivities in the overlapping range of baselines for the configurations, `ngVla-core-revC` and `ngVla-sba-revC`. None of the simulations included thermal noise or other systematic corruptions (*e.g.* tropospheric phase noise or antenna pointing errors). The single dish simulation total integration time was chosen so as to completely cover the required spatial area, but since thermal noise was not present in the simulations, the actual integration time was not consequential. In contrast the integration time (and integration durations) used in the interferometric simulations influences the *uv*-coverage of the data, which is an important variable in determining the characteristics of the resulting images. All simulations were run with zero sky opacity.

The antenna pointing patterns for the three arrays are shown in Figure 1. All simulations used the appropriate primary beams for each antenna: an Airy pattern for 6m or 18m diameter. The mosaic primary beams and the single dish weight map which results is also shown in Figure 1. For the wavelength and array configurations used in these simulations the largest spatial scale that can be recovered by the ngVLA 18-m core array is at most $\lambda/b_{min} \sim 19''$ and for the SBA interferometer $\sim 55''$.

3 Image Quality Metrics

It is crucially important for a telescope to be able to form accurate images of the sky brightness, with minimal (or known and well-understood) deviations. The quantitative measure of this capability is referred to as *image fidelity*. Because there is no general consensus on the best way to quantify image fidelity, I examined a variety of metrics that have been used. Defining the true sky brightness M (the input model to the simulation) and the synthesis image formed as I, one "classical" definition of image fidelity is:

$$F_1 = 1 - \frac{Max(|I - M|)}{Max(M)} \tag{1}$$

i.e., the image fidelity is taken to be the maximum deviation between the true sky and our (telescope-data based) model of it, divided by the peak true sky brightness. The evaluation is implicitly carried out over some predefined, potentially use-case dependent region of interest. This metric has a maximum value of 1.0, corresponding to perfect fidelity, but is unbounded from below. It has a simple interpretation— it is the maximum error divided by the true peak sky intensity— but also has the significant drawback that it is by definition driven by the values of one or two pixels in the image, so does not quantify

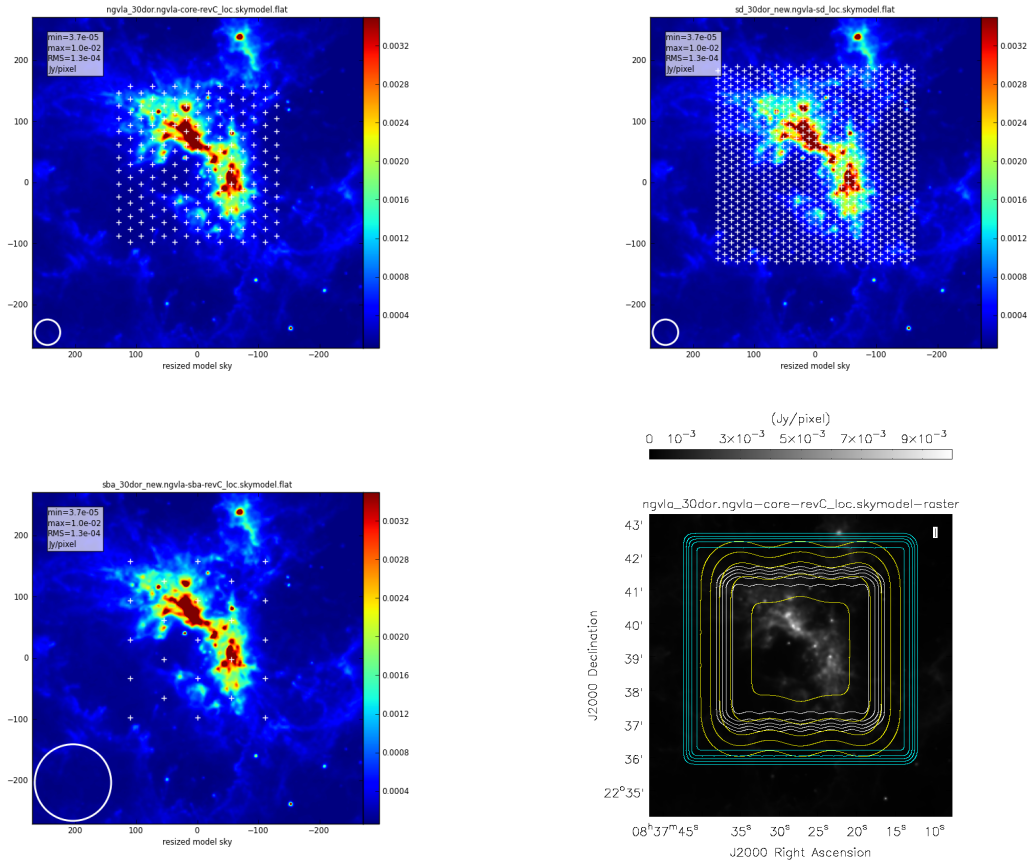

Figure 1: Simulation input with ngVLA core (top left), Total Power (top right), and Short Baseline Array (bottom left) antenna pointings. The lower right panel shows the total power weight map (cycan), the SBA mosaic primary beam (yellow), and the ngVLA core mosaic primary beam (white), along with the input model in the background. Primary beams are shown at levels of 20%, 40%, 60$, 80%, and 98% of the peak.

the typical imaging errors present in the image which in many cases are more scientifically important, and are more likely to correspond to intuitive impressions of image fidelity.

Most ALMA development studies (e.g., Tsutsumi et al. 2004) used a different metric of image fidelity[2]. Defining the fidelity of a given pixel i as

$$f_i = \frac{|M_i|}{|M_i - I_i|} \tag{2}$$

the overall image fidelity is then evaluated as the median of f_i values for pixels over some threshold of fractional peak intensity in the true sky image, for instance:

$$f_{alma,1\%} = Median(f_i|_{>1\%}) \tag{3}$$

Tsutsumi et al. (2004) used several thresholds (0.3% , 1%, 3%, 10%). This metric goes to infinity for a perfect image; unity when the M_i are $>> I_i$; and zero when the I_i are $>> M_i$. Its value can be sensitive to the chosen input image fractional intensity threshold. It also treats the true sky image and the model sky image asymmetrically in selecting the set of pixels to evaluate. This has the consequence that *image artifacts spatially removed from true sky structures do not affect the value of the image fidelity metric.* This will be discussed further below.

Versions of the ngVLA science requirements (ngVLA memo 65) prior to the writing of this document define image fidelity in two different ways. The immediately prior version (May 2019) posits

$$F_2 = 1 - \frac{\sum_i M_i |M_i - I_i|}{\sum_i M_i I_i} \tag{4}$$

which is equivalent to a weighted sum of the fractional error, with the weight being the product of the true and model sky images:

$$F_2 = 1 - \frac{\sum_i M_i I_i \frac{|M_i - I_i|}{I_i}}{\sum_i M_i I_i} \tag{5}$$

Here the fractional error is with respect to the formed image I_i. The original ngVLA definition (also adopted by Rosero et al. 2019) is

$$F_{2b} = 1 - \frac{\sum_i M_i |M_i - I_i|}{\sum_i M_i^2} \tag{6}$$

corresponding to a weighted sum of the fractional error with respect to the input (truth) image. These metrics are also unity for a perfect image and unbounded from below, and they too have the undesirable characteristic that *pixels where either the true or model sky*

[2]The formally adopted ALMA image fidelity requirement (2006-07-28-ALMA-90.00.00.00-001) is: "images shall be thermally limited at all points where the brightness is greater than 0.1% of the peak image brightness. This requirement applies to all sources visible to ALMA that transit at an elevation greater than 20 degrees at frequencies below 370 GHz."

values are zero receive zero weight. Therefore they are also insensitive to spurious structures in the reconstructed image in regions where the true sky pixels are close to zero; nor are they very sensitive to missing structures, if the reconstructed image values are close to zero. One result is that the dirty image of a point source— even when the Point Spread Function has dramatic sidelobes— can have an image fidelity quite close to unity. To quantitatively examine this behavior I calculated the fidelity metrics for an SBA point source dirty image and the corresponding cleaned image (Fig. 2 and Table 1).

In order to better quantify the effects of both missing and spurious structure I propose an alternate definition of image fidelity, F_3, defined as

$$F_3 = 1 - \frac{\sum_i \beta_i W_i |M_i - I_i|}{\sum_i \beta_i^2 W_i} \tag{7}$$

where $\beta_i = Max(|I_i|, |M_i|)$. It treats the images symmetrically, and in particular, it *provides significant weight to both missing flux and spurious features regardless of where they occur in the image*. I have also allowed a "window function" W_i which explicitly defines the region over which the fidelity is to be evaluated; within this region its value is 1, while it is 0 outside. This consideration is relevant in particular for mosaic imaging and feathering total power, since the spatial regions covered differ because of differing primary beams and the need for a total power "guard band" around the interferometric image, while the images themselves need to be co-registered for feathering. By default in this work I take $W_i = 1.0$ inside the 0.5 contour of the mosiac primary beam response. For other use cases, such as high angular resolution imaging of a debris disk, it will be much smaller. In these cases W_i could be implicitly defined simply the the size of the image chosen to reconstruct. The values of F_3 range from 1.0 for a perfect reconstruction to ~ 0 for M_i which are uncorrelated with the true sky intensity pixels P_i. It will in general only be negative if the reconstructed sky pixels are systematically anti-correlated with the true sky pixels values. The values of F_3 are also shown for the SBA point source case in Tabel 1. I define the fidelity error for an individual pixel j, following Eq. 7, as

$$\delta F_{3,j} = \frac{\beta_j |M_j - I_j|}{\sum_i W_i \beta_i^2} \tag{8}$$

Here I have assumed that this is evaluated for a pixel j within the region of interest ($W_j = 1$). This quantity is useful for determining what dominant features are responsible for deviations from perfect fidelity in the formed image. One minus the sum of $\delta F_{3,j}$ values over the region of interest in an image will equal F_3 for that image.

The Pearson correlation coefficient c was also calculated between the input model and the formed image. Like the fidelity comparisons the reference image is the model input smoothed with a Gaussian restoring beam of the size and shape given in the formed image header.

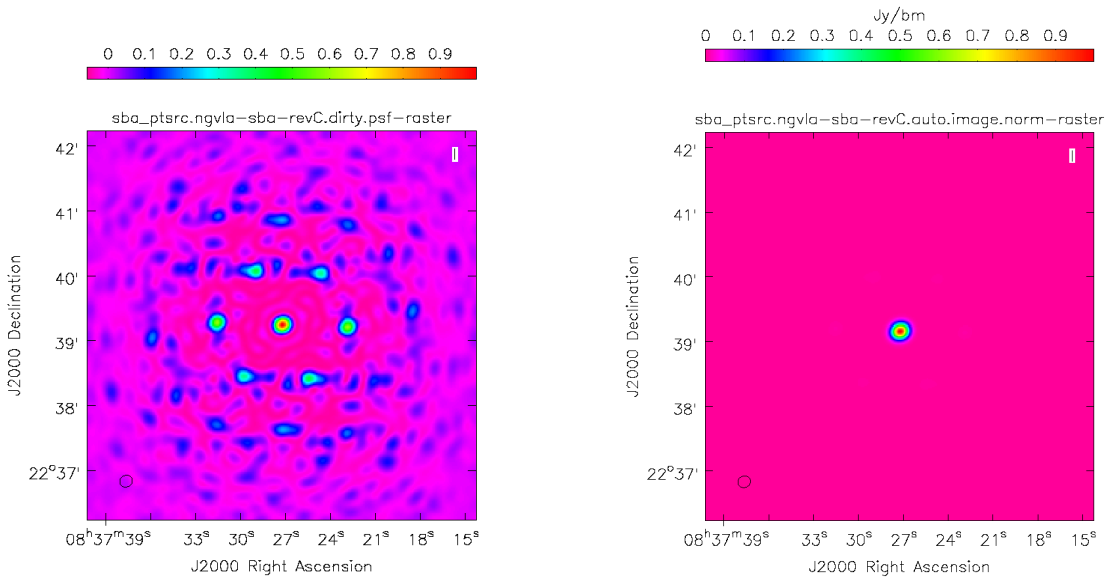

Figure 2: Dirty image (left) and auto-cleaned image of a point source observed with the SBA. By previous ngVLA image fidelity definitions (F_2, F_{2b}) both of these have fidelities within $\sim 5\%$ of unity (i.e. rather high fidelity).

Target	Array	Mask	F_1	F_2	F_{2b}	F_3	c
PtSrc	SBA	Dirty	0.615	0.946	0.949	0.507	0.71
PtSrc	SBA	Auto	0.996	0.999	0.999	0.999	0.999

Table 1: Fidelity metrics evaluated for an SBA dirty image of a point source and an auto-cleaned SBA image of a point source. c is the Pearson correlation coefficient and the image fidelities F_1, F_2, F_{2b} and F_3 are as defined in Eq. 1- 7.

4 Procedure & Results

4.1 Total Power Imaging

I found that high-fidelity simulation and imaging of total power data for this use case required care and some special procedures. My initial attempt at simulating total power observations of 30 Dor straightforwardly comprised running `simobserve` with appropriate values (e.g. $D = 18m$ and the 30 Dor model already shown) and gridding the total power "MS" into an image following the guidance in the M100 CASAguide. Specifically, a `SF` gridding function was used with a support region[3] of $N = 7$ and 9 pixels per primary beam. Results were evaluated in comparison to the model imaged smoothed by a Gaussian with a width as specified in the total power image header. **This resulted in the simulated single dish image recovering only** 56% **of total flux of the input image**. A comparable image fidelity ~ 0.6 (with no noise corrupting the simulation) is obtained. The peak flux density per beam is also only 65% of what is obtained by smoothing the input model smoothed with the 47.7″ (FWHM) Gaussian beam given in the image header.

Following this result, I simulated an observation of a known point source of 1 Jy. This resulted in a total flux low by exactly the same factor 0.56, and a peak surface brightness low by a factor of ~ 0.8. Inspecting the profile of the simulated point source, I find an actual, non-parametrically evaluated FWHM of the profile of 42.8″, while a Gaussian fit to the profile gives a FWHM 40.6″. The Gaussian fit was performed two different ways: first, with `au.fitFITSbeam`, and second, with completely independent code in IDL; results agreed to within 0″.01. The actual primary beam for an 18m dish with 0dB edge taper is 37″.91 (FWHM). The analysisUtil-ity `au.sfBeam` predicts a FWHM of the *gridded* primary beam of 42″.81, in exact agreement with the observed PB FWHM; `au.sfBeam` gives a FWHM for the Gaussian fit to that of 42″.27, somewhat larger than our observed Gaussian FWHM 40″.6. *All* of these are *significantly* narrower than the 47.7″ FWHM in the image header. For reference the relevant parameters for these calculations are: a central frequency of 93 GHz; 18m primary; 0dB edge taper; convsupport of $N = 7$ (width in pixels); and 4″.3 pixels.

Correct peak surface brightnesses are obtained if the simulated single dish data is corrected so that a simulated point source gives the expected peak surface brightness of 1 Jy/bm. For our particular use case this implies scaling up the image by a factor of $1/0.78 = 1.28$. To get correct integrated total fluxes the beam size in the header also needs to be correct. I adopted the 40″.6 FWHM Gaussian obtained by fitting the simulated point source. This procedure delivers $F_3 = 98.6\%$ fidelity for a simulated 30 Dor map and $F_3 = 97.0\%$ fidelity for a simulated point source map. The fidelity improves slightly if I use the actual simulated point source image as the PSF to compute the reference image in the fidelity calculation. In this case I find for 30 Dor $F_3 = 99.8\%$, and for the point source $F_3 = 99.6\%$. When using the exact PSF I also first rescaled the PSF to have unity peak, effectively applying the scale factor (1.28) previously mentioned; and put the Gaussian

[3]For most of this work $N = 7$ was used. The CASAguide recommends $N = 6$, which I also evaluated. The essential results were unchanged.

fitted value of 40″.6 for the beam FWHM in the image header for purposes of subsequent total flux calculations, since the brightness unit is Jy/bm. *When these physically-based corrections are applied, no arbitrary SDFACTOR is needed in the feather step to get correct total flux densities in the final, feathered image.*

Using the actual PSF and the input 30dor image, we expect a peak image surface brightness of 126 Jy/bm, in comparison to the observed 98.9 Jy/bm observed. One question is, how much of this is due to the simulation, and how much due to the imaging? To answer this I inspected the time-ordered raster data in the TP MS, finding a peak signal intensity of 106.8 Jy/bm (in scan 468, or fieldID 467). *This value is low by a factor of* 0.848, suggesting that whatever "de-gridding" procedure is used in the single dish simulation is responsible for the majority of the discrepancy[4]. However, the final SD simulation image is still low by $\sim 8\%$ from what would be expected in the time-ordered data (98.9 vs. 106.8 Jy/bm). This bias is occuring in the `sdimaging` step.

The M100 CASAguide does emphasize that it is necessary to directly calibrate the total power images using single dish maps of calibrators. One conclusion of this work is that the same sort of calibration step is needed for *simulated* total power observations as well. Preliminary inspection of an ALMA single dish pipeline cube suggests that they do *not* suffer from significantly erroneous beam sizes in the header. I did not check a manual TP reduction. Possible explanations of the observed differences include: 1) the task `sdimaging` does not correctly pick up the antenna illumination specified in the `vptable`, and instead uses defaults which are appropriate to ALMA (but not to the ngVLA) in deriving beam size from antenna diameter; 2) a different task is used for imaging (e.g., `tsdimaging`); 3) the beam size information in the header is separately corrected after imaging.

4.2 CLEAN Mask Definition

Deconvolving spatially extended structures is challenging, and the accuracy of the resulting images is often limited by the systematic errors and uncertainties inherent to the CLEAN process itself. These can to some extent be mitigated by careful CLEAN masking, which are therefore crucial in evaluating extended source imaging performance.

A general purpose, iterative auto-masking algorithm has been developed in support of the ALMA imaging pipeline (Kepley et al. submitted). This algorithm, called `auto-multithresh` in CASA, has been extensively validated against a range of real-world datasets. I chose to use `auto-multithresh` to facilitate reproducibility, and because the ngVLA operations plan calls for most data products to be automatically generated and delivered to PI's "Science Ready". I initially tried both ALMA 7m (for the SBA) and ALMA 12m-compact (for the ngVLA core) parameters. I also tuned the parameter specifically for the SBA and ngVLA core visually in an interactive clean session with somewhat better results. The main sense of the tuning was to make the box placement more conservative; to clean deeper

[4]Careful inspection of the input image at the RA and Dec of scan 468 shows that only about 2% of this can be ascribed to the fact that scan 468 is slightly offset spatially from the peak of predicted emission in the map.

Map	Total Flux [Jy]	Beam FWHM	Peak Brightness
30 Dor Input Model	**1432**	n/a	0.01 Jy/pix
Smoothed 30 Dor Model	1418	47″.7	154.3 Jy/bm
Smoothed, Regridded 30 Dor Model	1433	47″.7	154.3 Jy/bm
SD Simulation (SF, $N = 7$)	812	47″.7	98.9 Jy/bm
Point Source Model	**1**	n/a	1 Jy/pix
PtSrc, SD Simulation (SF, N=7)	0.56	47″.7	0.78 Jy/bm
PtSrc, SD Sim (SF, N=6)	0.60	46″.5	0.83 Jy/bm
PtSrc model smoothed by exact PSF	0.998	40″.6	1.0 Jy/bm
PtSrc Sim. (corrected beam and scaling)	1.000	40″.6	0.995 Jy/bm
30 Dor model smoothed by exact PSF	1436	40″.6	127 Jy/bm
SD Simulation (corrected beam and scaling)	1439	40″.6	126 Jy/bm

Table 2: Integrated flux density and peak surface brightness for a known model; smoothed and regridded versions of the known model; and images made from simulated single dish observations of these models. Two known models are considered: a point source and a 30-Dor-like science target. FWHM values are the values listed in the single dish image header after corrections discussed in the text. The final four lines show that simulated single dish observations+imaging of both a point source and the science target model give consistent results (total flux density and peak surface brightness) as what is obtained directly convolving the input model with the exact PSF (derived from the point source simulation) and applying corrections discussed in the text. It is also notable that the smoothing and regridding of the input model (top 3 lines of the table) do not introduce significant systematic errors.

Array	T_{side}	T_{noise}	$f_{minBeam}$	T_{low}	g	S	f_{cycle}
ALMA-7m	1.25	5.0	0.1	2.0	75	1.0	1.0
ngVLA-SBA	1.5	5.0	0.1	2.5	50	0.75	0.75
ALMA-12m	2.0	4.25	0.3	1.5	75	1.0	1.0
ngVLA-core	3.5	5.0	0.3	2.5	50	0.75	0.65

Table 3: Parameters used for `auto-multithresh` automated clean masking.

within each major cycle; and to make the boxes tighter. The `auto-multithresh` parameter values used are shown in Table 3.

Overall four strategies were employed for defining the CLEAN masks used:

- No mask (dirty image) – only evaluated for SBA-only case.
- TCLEAN automasking (`auto-multithresh` – Kepley et al. submitted.) using ALMA 7m (SBA) or 12m-compact (ngVLA-core) parameters.
- TCLEAN automasking (`auto-multithresh`) using parameters tuned for the SBA and ngVLA-core individually.
- Interactive, user-defined masking (ngVLA-core only).

Results from these different approaches are discussed in § 4.4. Note that the scripts on github only include those needed to reproduce the primary results of this investigation (the tuned automask method).

4.3 Restoring Beam Flux Bias (Jorsater-VanMoorsel Effect)

When imaging extended objects the flux scale difference between the cleaned and residual maps can be a sigificant source of systematic error (Jorsater & VanMoorsel 1995, hereafter JvM). The effect can be particularly pronounced when multiple interferometer array configurations are combined, due to approximating a multi-resolution PSF with a single Gaussian PSF. Walter & Brinks (1999), for example, find that integrated flux densities from their multi-configuration VLA data (B+C+D) would be overestimated by a factor of two without the correction.

I calculated the magnitude of the restoring beam area bias correction using the expressions given in JvM and find it is a modest correction for the current use cases. Numerically, I find a correction factor (as defined in JvM) η of 1.14 for the SBA; 1.094 for the ngVLA core, tapered to $2''$; and 1.078 for the joint core+SBA. Applying the corrections typically resulted in a modest (1% - 2%) change in fidelity; it did not consistently improve fidelity.

4.4 Interferometric Imaging

Several deconvolution and imaging strategies were used in order to characterize the impact of reconstruction method on the final result. The methods used to image the individual

interferometric arrays' data, along with identifying names we assigned them, are as follows:

- For the SBA:
  - Sba-Dirty: form the dirty image.
  - Sba-AutoAlma: multi-scale deconvolution using automasking with ALMA 7m default parameters.
  - Sba-AutoTuned: multi-scale deconvolution using automasking (ngVLA tuned parameters).
  - Sba-SDmod: same as Sba-AutoTuned, but use the total power map as a starting model in `tclean`.
  - Sba-Feather: same as Sba-AutoTuned, but feather total power image in at the end.
  - Sba-SDmod-Feather: same as Sba-SDmod, but feather total power image in at the end.
- For the ngVLA core:
  - Core-User: interactive clean masking at each major cycle.
  - Core-AutoAlma: multi-scale deconvolution using automasking with ALMA 12m compact configuration default parameters.
  - Core-AutoTuned: multi-scale deconvolution using automasking (ngVLA tuned parameters).

All images were cleaned with a threshold of 1% the peak brightness seen in an initial simple clean of the brightest emission; all of the SBA cleans ultimately terminated due to this threshold. In all of the SBA cases the clean deconvolution was stable in the sense that there were no indications of divergence. The ngVLA core-only cleans were less stable and typically did show some divergence; this isn't terribly surprising considering the extent of emission in the field in comparison to the spatial frequencies that these data sample well. To feather in the total power data I followed the procedures described in the M100 CASAguide.

Current best practices suggest that the optimal approach to imaging multi-array data is to perform the deconvolution jointly, *i.e.*, including all of the interferometric data in a single deconvolution. This is the current recommendation for ALMA 7m+12m data. Unfortunately it has as yet been impossible to identify a method of correctly imaging heterogeneous array data *other than ALMA data* in CASA. This is in spite of extensive effort and discussions with other experts and developers; some of this investigation is documented on `CAS-8592`, `CAS-11271`, `CAS-11464`, `CASR-301`, and `CASR-470`. In the absence of a proper technique I chose to simply image the SBA and ngVLA-core data together *but using the 18m antenna primary beam for each.* Kundert et al. (2017) did exactly this in simulations of ALMA 7m and 12m mosaics and found a limiting dynamic range ~ 70, corresponding

Image	Flux [Jy]	F_1	F_2	F_{2b}	F_3	c
Sba-dirty	n/a	0.35	−10.8	0.08	0.08	0.28
Sba-AutoAlma	224.8 (−80%)	0.61	−1.15	0.32	0.32	0.83
Sba-AutoTuned	238.8 (−79%)	0.60	−1.16	0.32	0.32	0.84
Sba-SDmod	1436.8 (+17%)	0.74	0.87	0.85	0.86	0.98
Sba-Feather	1214.2 (−1%)	0.92	0.93	0.93	0.93	0.99
Sba-SDmod-Feather	1232.3 (+0.6%)	0.88	0.95	0.95	0.95	0.99
Core-User	17.3 (−99%)	−0.02	−9.1	0.07	0.09	0.20
Core-AutoAlma	35.2 (−97%)	−0.81	−5.63	−0.02	0.10	0.18
Core-AutoTuned	71.0(−96%)	0.38	−5.34	0.14	0.14	0.62
Joint-AutoTuned	307.4 (−75%)	0.50	−0.26	0.37	0.43	0.74
Joint-SDmod	1457.0 (+18%)	0.08	0.68	0.61	0.64	0.88
Joint-Feather	1213.5 (−1%)	0.44	0.76	0.75	0.76	0.92
Joint-SDmod-Feather	1236.6 (+1%)	0.26	0.71	0.69	0.71	0.89

Table 4: Image fidelity and total flux retrieval results for the range of datasets and reconstruction techniques discussed in the text. The total flux density of the input model over the region considered— defined to encapsulate the region within the 50% contour of the ngVLA core mosaic primary beam— was 1224.8 Jy. This region is slightly different from the one considered in Table 2. The case giving best overall fidelity is shown in **bold** for each array combination.

to $F_3 \sim 0.98$ in ngVLA terms, which is more than adequate for purposes of the current study. It is also necessary to manually adjust the relative weights of the 6m and 18m data by a factor of $(6/18)^4$. Following the advice of the Data Combination 2019 working group, I also concatenated the SBA and core MS's in *time order*, which required some foresight at the simulation stage. The methods used to image the data jointly were then as follows:

- Joint-AutoTuned: multi-scale deconvolution of SBA+Core using automasking (ngVLA core parameters).
- Joint-SDmod: same as Joint-AutoTuned, but using total power model as initial clean model.
- Joint-Feather: same as Joint-AutoTuned; feather total power in at the end.
- Joint-SDmod-Feather: same as Joint-SDmod; feather total power in at the end.

All images were cleaned with a threshold of 1% the peak brightness seen in an initial simple clean of the brightest emission. The Joint deconvolutions also showed indications of instability and divergence (more comments on this are in § 5).

The fidelity metrics for all of these images are presented in Table 4, along with the total flux density measured in the same large aperture in each image. The four best SBA and SBA+TP images are shown in Figure 3. When total power is included, the SBA robustly

recovers accurate ($\pm 1\%$) total flux densities for the object and produces a high fidelity ($F_3 \sim 0.95$) image. In the absence of total power information the best fidelity achieved is 32%; in this case only about 20% of the total flux is recovered. When the TP is used as a starting model *only*, fair results are obtained: 86% fidelity and a total flux accurate to within $\sim$ 20%. Clearly superior results are obtained by including the TP data via feathering. Images of the fidelity error (Eq. 8) and the fractional flux error are presented in Figure 4.

The ngVLA core-only results, not surprisingly, are poor: a few percent of the total flux is recovered and the image fidelity is correspondingly on the order of 10%. Tuned automasking gives the best results, though this is a question of the lesser of three evils. The best core-only image is presented in Figure 5.

The joint SBA+core images are a distinct improvement on the ngVLA core-only images, and successfully reproduce higher-resolution structures that cannot be seen in the SBA images; see Figure 6. The total flux densities are similarly accurate when total power is feathered into the images ($\pm 1\%$). However they fall short of the quantitative fidelity goal, with typical fidelities of $\sim$ 70% and a best-case fidelity of 76%. Figure 7 shows the fidelity error and fractional brightness error. One cause of the lower fidelity is the primary beam error— the fact that I used the 18m PB for the 6m antennas. A systematic pattern corresponding to the SBA antenna pointing positions can be discerned in the Joint images. This is illustrated more clearly in Figure 8. This effect has a larger impact than seen in Kundert et al. for two reasons: 1) the antenna diameters differ more (6m-18m, vs 7m-12m), leading to larger primary beam errors for the 6m antennas; & 2) by coincidence the brightest location in the sky model falls in an "anti-node" of the assumed 6m primary beam pattern, *i.e.*, a location where the incurred beam error is largest.

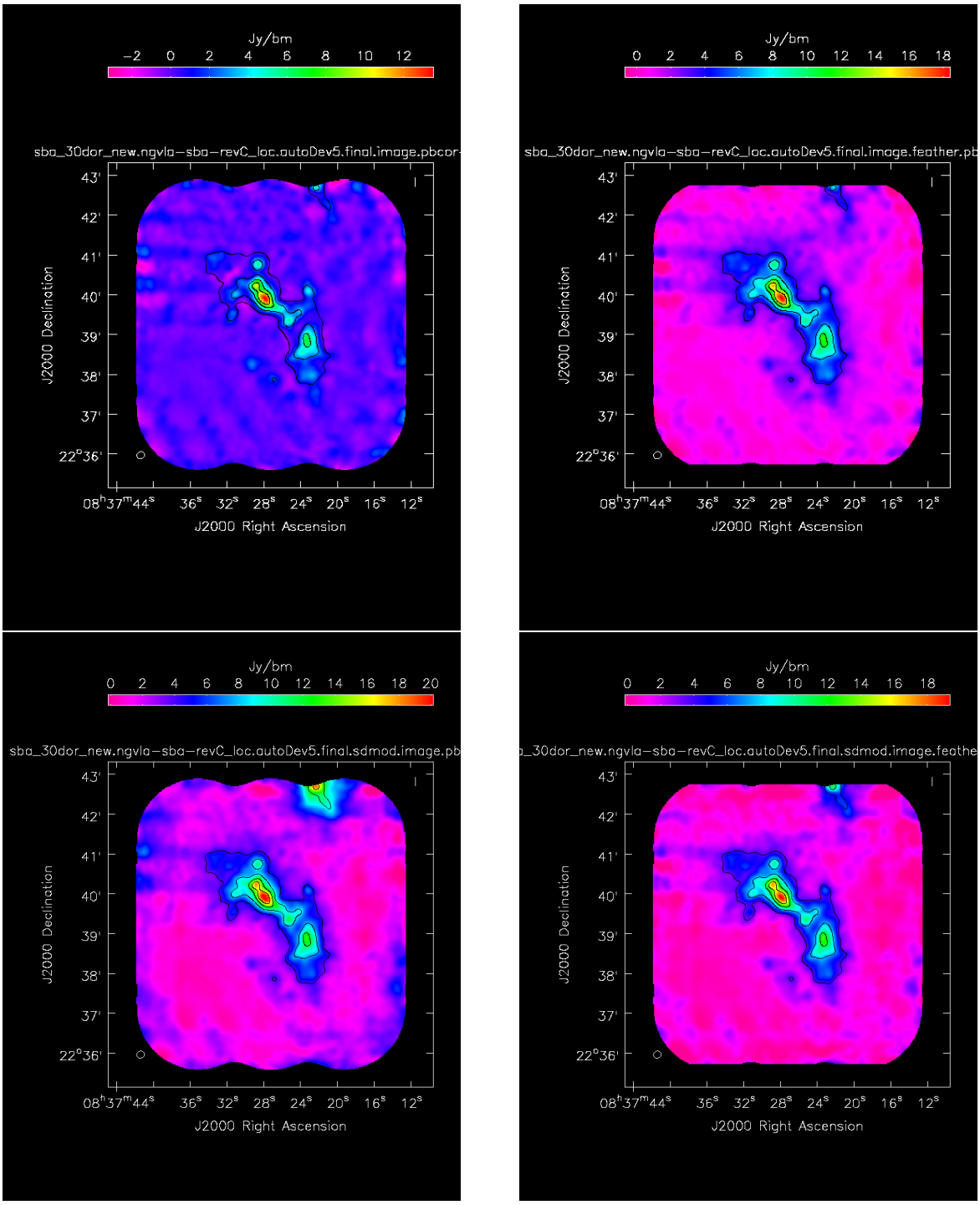

Figure 3: Recovered Short Baseline Array images. Top left: SBA alone; Top right: SBA+TP (feathered); Bottom left: SBA with TP starting model for CLEAN; Bottom right: SBA with TP start model, and TP feathered in afterwards. Black contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

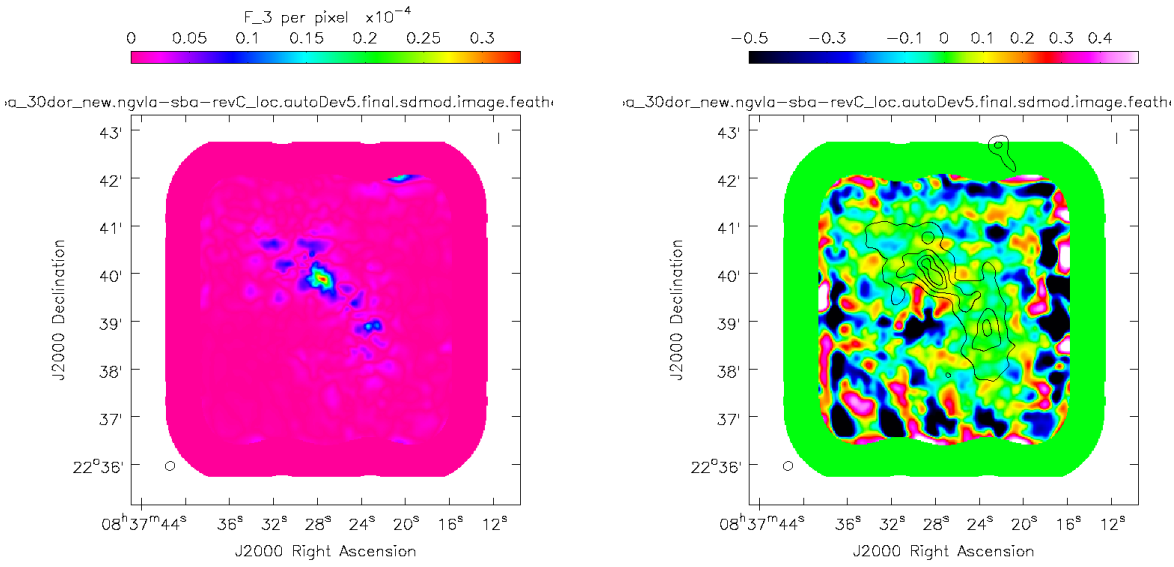

Figure 4: Fidelity error (left, following Eq. 8) and fractional error in reconstructed image (right) for the SbaSdmodFeather case. Black contours show the input model at the resolution of the restoring beam.

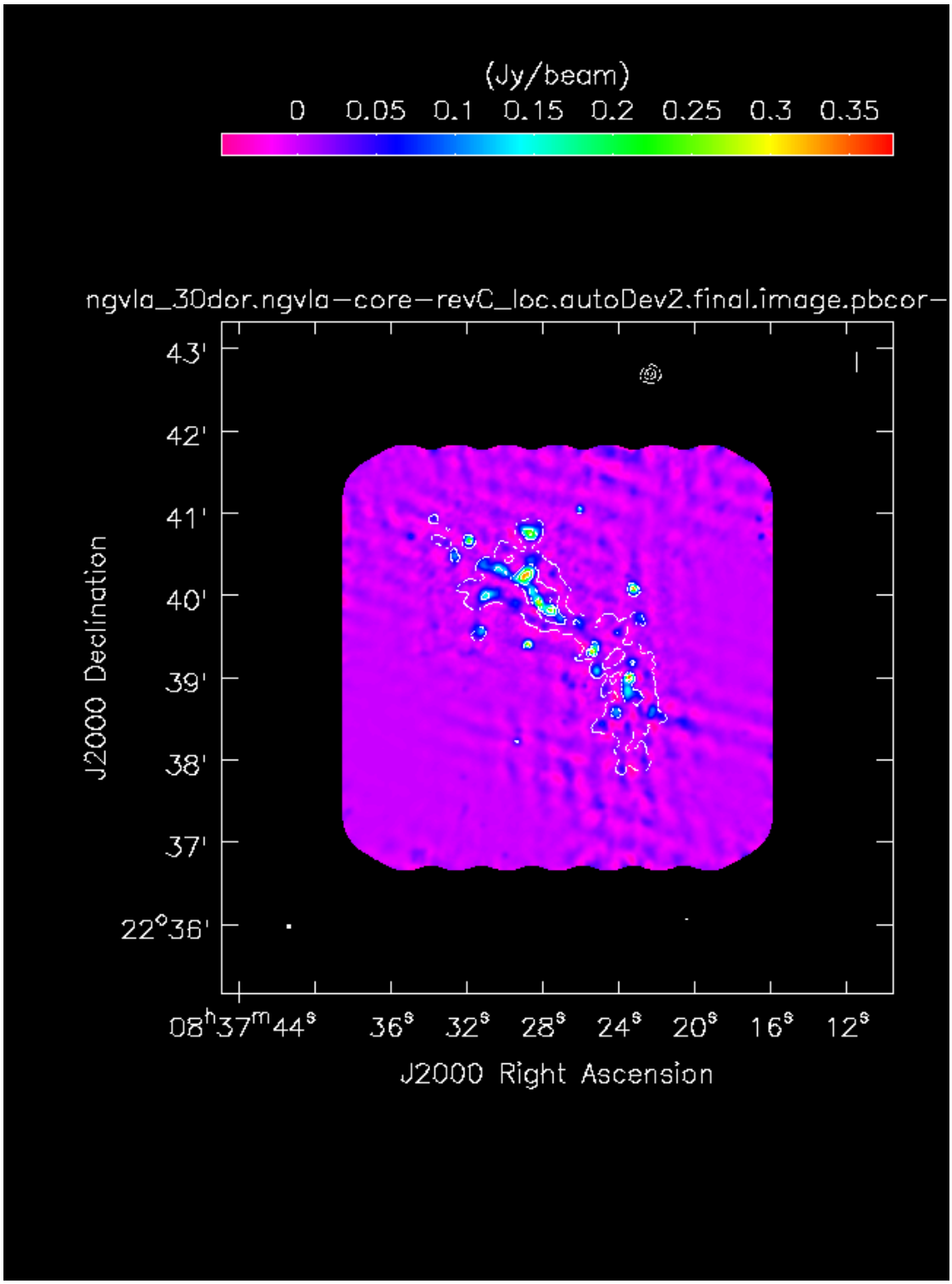

Figure 5: ngVLA core (only) recovered image. White contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

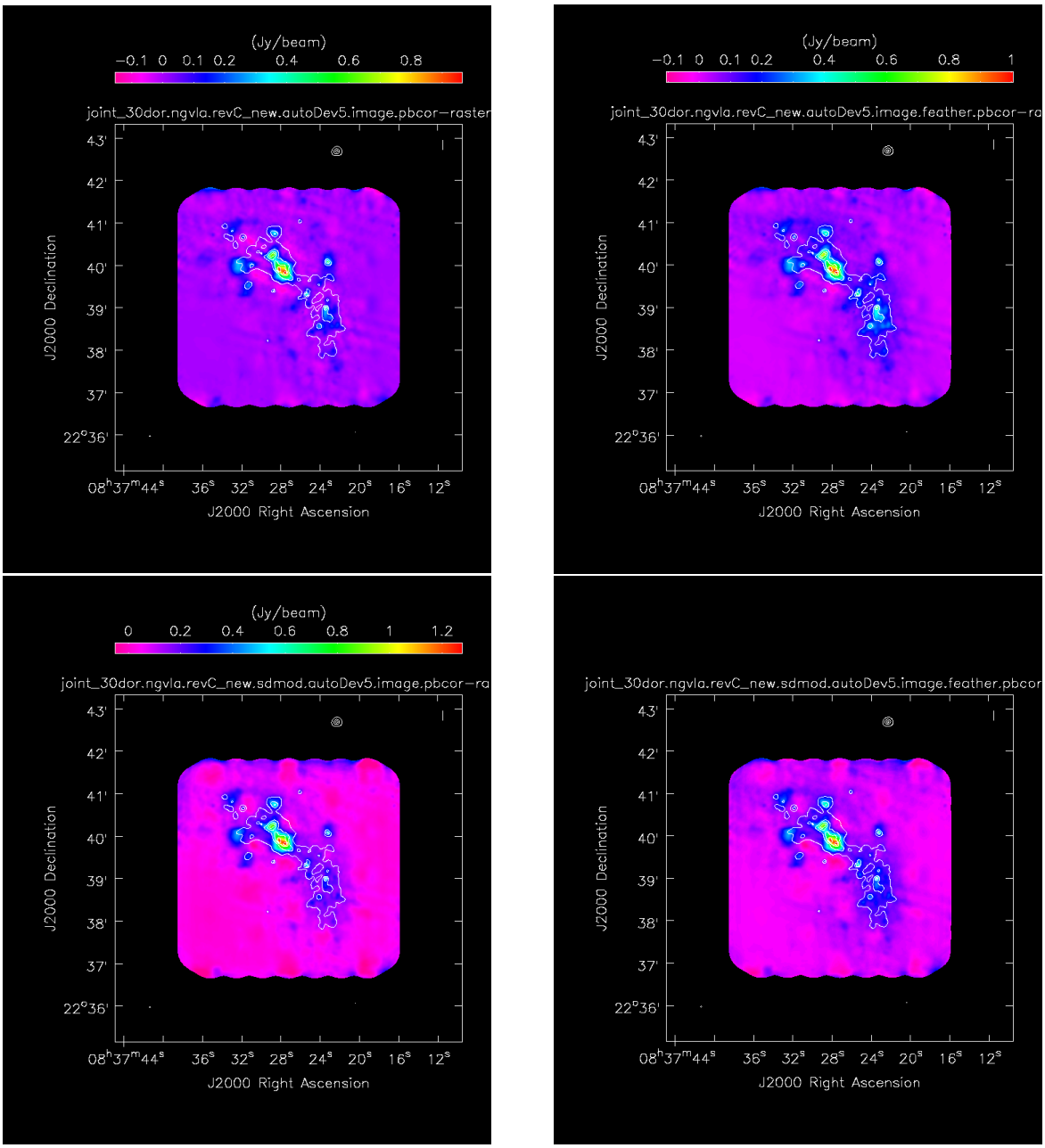

Figure 6: Recovered Joint SBA+Core images. Top left: Joint-AutoTuned; Top right: Joint-Feather; Bottom left: Joint-SDmod; Bottom right: Joint-SDmod-Feather. White contours represent 20%, 40%, 60%, and 80% of the input model peak intensity (smoothed by the restoring beam).

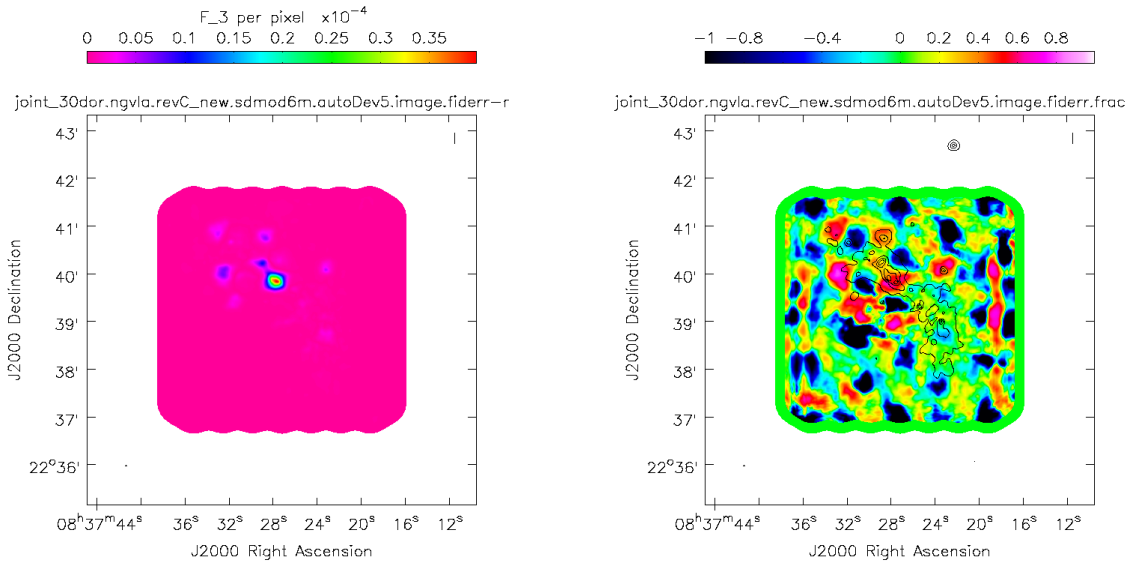

Figure 7: Fidelity error (left, following Eq. 8) and fractional error in reconstructed image (right) for the JointSdmodFeather case. Black contours show the input model at the resolution of the restoring beam.

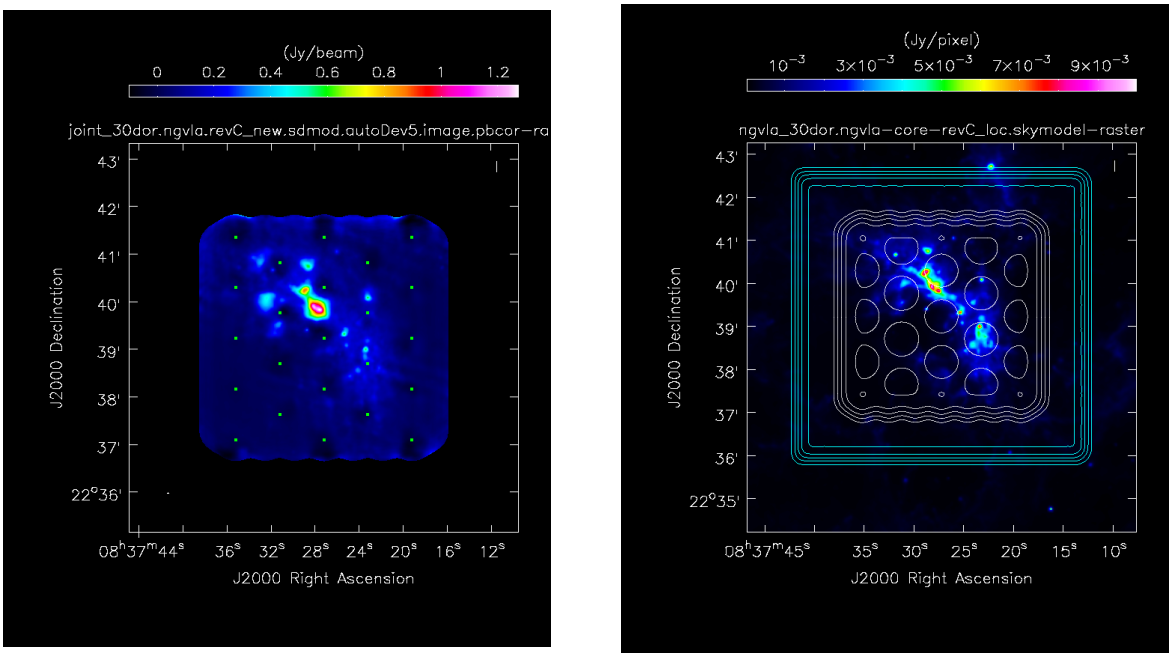

Figure 8: Jointly deconvolved SBA+Core image on a stretch which more clearly demonstrates the artifacts resulting from the erroneous primary beam (left); the 6m mosaic pointings are marked with green points. The joint primary beam mosaic is shown on the right in white, with the TP weight map in cyan. Both sets of primary beam contours are at 0.2, 0.4, 0.6, 0.8, and 0.98× the peak (the same contours as Fig. 1).

5 Conclusion

The major conclusions of this work are as follows:

- The SBA itself robustly produces high-fidelity images of structures extended over many fields of view when 18m total power data is included. In the absence of total power the image fidelity is $F_3 \sim 30\%$.
- SBA or ngVLA core images, when 18m total power are feathered in, produce accurate total flux densities to $\pm 1\%$. In the absence of total power only about 20% of the total flux is recovered.
- Feathering provides much more accurate total power information than just using the total power data to initialize the interferometric clean model, although the latter is better than not using any total power information at all. For the well-behaved (SBA) use case, the SDmod+feather method also delivers slightly higer image fidelity than feathering alone.
- Robustly producing high-fidelity ngVLA core, or core+SBA (joint) images has not yet been demonstrated. One limiting factor is the need for properly handling multiple antenna diameters for a general case (not ALMA). A request for this development has been submitted (`CASR-470`). These core and joint use cases also suffered from instability in clean which more optimal parameter choices may be able to mitigate. Inclusion of SBA data *should* stabilize the ngVLA-core imaging, but the primary beams need to be fixed first.
- Based on a quantitative evaluation, I recommend and use an improved fidelity metric F_3. For high-fidelity reconstructions, F_3 and the *original* ngVLA fidelity metric (F_{2b}) give very similar results for the main use cases I examined, though they differ for low-fidelity images. F_2 is less well-behaved.
- Care is required to produce quantitatively correct single dish simulations. This being done there is no need for arbitrary scalings of the single-dish data in `feather()` (i.e., SDFACTOR can be left at unity). The most important issue to note here is that the beam size that `sdimaging` writes in the header is incorrect for our use case, possibly because of differences in the assumed antenna illumination profile (ngVLA: airy; ALMA: more heavily tapered). This underscores the importance of having well-calibrated, well-understood inputs to any combination method.

Interferometric imaging of objects of size $>> \lambda/b_{min}$ is intrinsically challenging so it is not surprising that the core and joint deconvolutions were unstable. In the context of automasking, one of the specific challenges is that automasking tended to work better if each major cycle goes deeper, avoiding premature termination of mask updates. Deeper major cycles are also, however, a source of systematic error in imaging since errors in the approximate PSF used in the minor cycles are corrected at each major cycle. Better

results might be obtained in the future by varying the `minpercentchange` sub-parameter. Alternative deconvolution algorithms, such as SDINT (Rau et al. 2019) or TP2VIS (Koda et al. 2019), may also prove more stable and would be worth evaluating on an equal footing. One surprising result of this investigation was that the JvM corrections, while modest, did not consistently improve image fidelity. It would be worth exploring this in more detail, and in particular, if a different form of the correction is needed in the context of an image which also includes total power data, e.g., by feather. Finally, it's worth noting that while this is a continuum only simulation— important elements of the scripts probably won't work for cubes— it is in practical terms most indicative of one channel of a spectral line use case in that there is no guarantee that ngVLA TP will provide robust and sensitive continuum information, similar to the current situation with ALMA.

Acknowledgements: Thanks to Urvashi Rao and Kumar Golap for extensive and ongoing consultation in attempting to get ngVLA heterogeneous array imaging working; Amanda Kepley for helpful automasking advice; and Ryan Loomis for sharing his code to calculate the JvM correction.

References:

Indebetouw, R. et al. 2013 ApJ 774, 73

Jorsater, S. & VanMoorsel, G. 1995, AJ 110 20137

Kepley, A. et al., "Auto-multithresh: A General Purpose Automasking Algorithm", submitted to PASP

Koda, J. et al. 2019, PASP v131, 999 p054505

Kundert, K. et al. 2017, IEEE Trans. Ant. & Prop. 65, v2 "Understanding Systematic Errors Through Modeling of ALMA Primary Beams"

Mason, B. et al. 2018, ngVLA memo 43: "The ngVLA Short Baseline Array"

ngVLA Science Advisory Council 2017, ngVLA memo 19: "Key Science Goals for the Next Generation Very Large Array (ngVLA): Report from the ngVLA Science Advisory Council"

Rau, U., Nikhil, N. & Braun, T. 2019, AJ 158, 3

Rosero, V. et al. 2019, ngVLA memo 65: "Sculpting of the Synthesized Beam and Image Fidelity Study of KSG 1: Imaging of Protoplanetary Disks"

Selina, Murphy & Erickson 2017, ngVLA memo 18: "Summary of the Science Use Case Analysis"

Selina, R. et al. 2018a, Proc. SPIE 10700E, 10S

Selina, R. et al. 2018b, ASPC 517, 15S

Tsutsumi, T. et al. 2004, ALMA memo 488: "Wide-Field Imaging of ALMA With the Atacama Compact Array: Imaging Simulations"

Walter, F. & Brinks, E. 1999, AJ 118 273