VLASS Project Memo: 17

Characterization of VLASS Single Epoch Continuum Validation Products

M. Lacy, P. Patil (NRAO),
K. Nyland (NRL)

October 25, 2022

Abstract

This Memo describes the validation of the VLASS SE continuum images with respect to the survey requirements. Images both produced by scripts prior to their implementation in the pipeline, and the versions created by the production pipeline are included in the analysis. We find that nearly all the survey requirements are likely to be met once system noise and complex gain calibration errors are taken into account, with the exception of in-band spectral indices of extended sources near strong brightness gradients. We believe this exception is due to intrinsic limitations of in-band spectral maps. We find that the mosaic gridder combined with an image plane position correction provides adequate tt0 (first Taylor term, i.e. total intensity) images over the whole sky. The tt1 (second Taylor term) images made with the mosaic gridder are, however, strongly affected by the neglect of the w-term at zenith distances $\gtrsim 45^\circ$, resulting in corrections being needed for spectral indices of compact sources unless the aw-project gridder is used.

1 Introduction

The VLASS Single Epoch (SE) continuum images are Stokes I, two Taylor-term (for spectral index information), self-calibrated images that are intended to be a significant improvement over the Quick Look images that are produced ≈ 1 month after the observations are taken (see VLASS Memo 13). The VLASS SE continuum imaging algorithm required some development work in order to produce acceptable results. This was performed under the auspices of the VLASS Imaging Project (VIP), which ran from January to April 2020. The aim of this project was the development of a script capable of producing VLASS single epoch (SE) continuum images capable of satisfying the survey requirements (VLASS Memo #15[1]). This effort concluded with the successful production of ≈ 10 VLASS SE images using a script. To show that images produced this way satisified the survey requirements, a larger ≈ 100 deg^2 imaging effort using this VIP script was begun in October 2020. This document describes the analysis of those images, and of the validation of the pipeline SE products made when the VIP script was adapted for the VLASS pipeline.

The 98 deg^2 imaged using the VIP script were divided into 12 validation fields varying in size from 1–30 deg^2 and spanning a wide range of Galactic and extragalactic fields, including several calibrators and fields with prior, deeper radio surveys in S-band (Table 1). For all tests we have results for two different griddings: AW-project (Bhatnagar et al., 2013) with 32 w-planes (denoted AW32), and the mosaic gridder (MOS, see Mooley et al., 2019). There were also some smaller tests with a single w-plane in AW-project (AW1), a single plane with with ATERMS=FALSE (AW1-ATF), and one with USEPOINTING=FALSE (AW1-PTGF). The mosaic gridder by default does not use conjugate beams (Bhatnagar et al., 2013), but they are an option, so conjugate beams with the mosaic gridder were also tested (MOS-CBT) (AW-project has CBT by default). Finally, images made with the SE pipeline (currently only using the mosaic gridder) are denoted SEPLMOS.

Cleaning differed between the AW-project and MOS runs in that for MOS, the images were initially cleaned within a mask and then cleaning was restarted, cleaning without a mask to pick up any remaining emission that was not masked. The AW1 and AW32 runs used a mask for all cleaning. All other parts of the imaging script were kept the same for the comparison. Simple component catalogs for each set of images were made using pyBDSF (Mohan & Rafferty, 2015). Comparison archival datasets included the

---

[1] current version of the script at `https://gitlab.nrao.edu/jtobin/vip/-/blob/master/vip_script_mosaic_6.1.2.py`

CNSS mosaic from Mooley et al. (2016), made by combining individual single pointings, and the COSMOS S-band survey (Smolčić et al., 2017).

Updated survey science requirements are listed at `https://open-confluence.nrao.edu/display/VLASS/Updated+VLASS+survey+science+requirements+and+parameters`. We compare the ones that can be either partly or fully tested with the 100 deg^2 of validation images in Section 2 below.

2 Comparison to survey requirements

2.1 RMS noise (SCI-BDP-010)

The RMS was measured as the median of the median pixel values in each of the 1 deg^2 RMS images produced by the VIP script for each validation field. The tt0 RMS is within the requirement (170μJy) over most of the 100 deg^2, with the notable exception of the Galactic Plane and some parts of the CNSS field (the CNSS field is close to the Clarke Belt, where RFI is very strong, leading to a loss in effective bandwidth). The RMS in tt0 is about 7% higher on AW32 and AW1 than for MOS (but the beamsize is 7% smaller than that for MOS).

The survey requirement is that 90% of the sky should be below 170μJy, but the fraction of the 100 deg^2 below the requirement is only $\approx$ 65% for the scripted VIP images. However, some of these validation fields were preferentially selected to be in problematic areas such as the "Clarke Belt" centered on -5.5° where the emission from geostationary satellites peaks and the Galactic Plane, so this is not unexpected. Before the pipelined (SEPLMOS) images were made, the VLASS calibration pipeline was changed to reduce the growth of the automated target flagging and the data recalibrated. This results in about a 10% decrease in the noise in the CNSS area, and this will probably be sufficient to allow most fields near the Clarke Belt to satisfy requirements. The RMS of the tt1 images is, as expected (see the VLASS Technical Implementation Plan [2]), about six times that of the tt0 image in all cases.

2.2 Flux density calibration (SCI-BDP-020)

2.2.1 Flux density calibrator observations

Three gain calibrator fields were imaged in both AW32 and MOS. These have flux densities and spectral indices measured from their pointed observations contemporaneous with the VLASS observations, so provide a reliable check on the flux density scale. One of these fields (W1303-1051 in the VLA calibrator list) was run through the gain compression correction workflow, but still had a recovered flux density within $\approx$ 2% of

---

[2] `https://www.authorea.com/users/4076/articles/8161/_show_article`

Table 1: Validation fields

Field Name	Tile(s) drawn from (and campaign)	Approx RA, Dec	Size (deg^2)	Approx ZD (deg)
CDFS	T03t06,T04t06 (2.1)	0328−2830	9	66,62
CNSS	T10t34,T10t35,T11t34,T11t35 (2.1)	2230+0000	31	37,39,32,34
COSMOS	T11t15,T11t16 (2.1)	1000+0200	4	41,36
Cygnus	T21t21 (2.1)	2040+4230	9	19
Galactic Plane	T08t28,T09t28 (1.2)	1830−1000	9	47,55
GOODS-N	T26t10 (2.1)	1240+6430	5	36
N-SPARCS	T18t21 (2.1)	1530+2900	8	32
S. Gal. Plane	T01t35 (1.2)	1715−3830	7	73
J0555+3948	T20t08 (1.2)	0555+3900	10	40
J1546+0026	T11t24 (1.2)	1546+0030	1	36
W1303−1051	T08t20 (1.2)	1303−1030	1	46
XFLS	T25t13 (2.1)	1718+5930	4	42

the pointed data. The results of these tests are shown in Table 2. All are within the survey requirements of 10% and goal of 5% in flux density accuracy.

Table 2: Comparison of calibrators between their pointed observations and their recovered flux densities and spectral indices using total flux densities measured in IMFIT.

Calibrator	$S_{\rm Ptd}$ (Jy)	$\alpha_{\rm Ptd}$	$S_{\rm QL1.2}$ (Jy)	$S_{\rm MOS}$ (Jy)	$\alpha_{\rm MOS}$	$S_{\rm AW32}$ (Jy)	$\alpha_{\rm AW32}$
J0555+3948	3.69	0.622	3.835	3.840	0.708	3.805	0.529
J1546+0026	1.10	-0.771	1.097	1.110	-0.551	1.102	-0.814
W1303-1051	0.320	-0.048	0.333	0.326	0.133	0.324	-0.060
Diff cf. ptd	-	-	-4.0% to -0.2%	-2.1% to + 1.2%	+0.06 to +0.22	-2.7% to +0.9%	-0.04 to +0.09
Mean diff.	-	-	-2.7%	-0.3%	+0.16	-0.4%	-0.05

2.2.2 Clean bias

Besides the flux calibration, radio surveys are also subject to other forms of flux density bias near the survey limit. Clean bias, where sidelobes are erroneously cleaned due to the presence of noise close to the clean threshold, subtracting flux from real sources, has been seen in the NVSS survey (Condon et al., 1998). Rau et al. (2016) point out though that the smaller sidelobes associated with the MT-MFS technique used in the SE continuum images means that clean bias effects should be negligible. (There is also a different form of clean bias arising from insufficiently deep cleaning of faint emission, see VLASS Memo 16 for an analysis of this in the context of VLASS.)

To search for evidence of clean bias in the VLASS catalogs we compared the CNSS Epoch 2 catalog from M+16 with the pyBDSF catalog made from the MOS images in CNSS. There is a flux calibration offset between the CNSS and VLASS images of about 16% over the whole flux range (in the sense that the VLASS flux densities are higher), probably due to the fixes applied to try to mitigate the gain compression issues in Mooley et al. (2016). Comparing the flux of the $\approx$ 12mJy point source J221108.3-000302.6 in the P145 pointing as recalibrated for extended source comparisons (Section 2.7) shows that indeed the flux scale for the recalibrated data is consistent with the VLASS flux calibration, suggesting that Mooley et al. (2016) overcompensated for the gain compression. Removing this as a constant 16% offset results in the plots shown in Figure 1. (We also show, in Figure 2 the results from comparing to the COSMOS survey, where no overall changes to the flux density scale have been applied.) These plots show that there is no trend for VLASS sources to have lower flux densities than their CNSS counterparts close to the survey limit, consistent with no measureable clean bias. In fact the VLASS peak flux densities in particular seem to be higher, probably a combination of the smaller average beam in CNSS and Eddington Bias (Eddington, 1913), where positive noise fluctuations enhance the probability of picking up a source near the survey limit. This effect will be characterized as part of the catalog validation process.

2.3 Beamsize (SCI-OBS-005)

2.3.1 Size recorded by tclean

The MOS beams are about 7% larger than AW beams (in linear size; 14% in area) (Table 6). Nevertheless, in all but the 9 deg^2 of the J0555+3948 field the beamsize is smaller than requirement (a 3-arcsec beam). This requirement needs only be met over 60% of the sky, so the beamsize requirement will be easily met.

2.3.2 w-term effect on beamsize/peak depression

The beamsize given in the header is derived from fitting the beam produced by tclean (see VLASS Memo 14). For the mosaic gridder, this neglects the w-term, and the actual beam is distorted. Simple simulations suggest this effect is a few percent in the far south of the survey (VLASS Memo 14). To test this, we convolved the CDFS field AW32 and MOS images to the same beam (3-arcsec round) and compared the peak to total flux ratios from pyBDSF catalogs (see Figure 3). Though a KS test shows that the null hypothesis that the data are from the same underlying distribution can be rejected at the 98% level, the differences in the mean peak-to-total flux ratios are very small (mean of 0.74 for AW32 and 0.75 for mosaic).

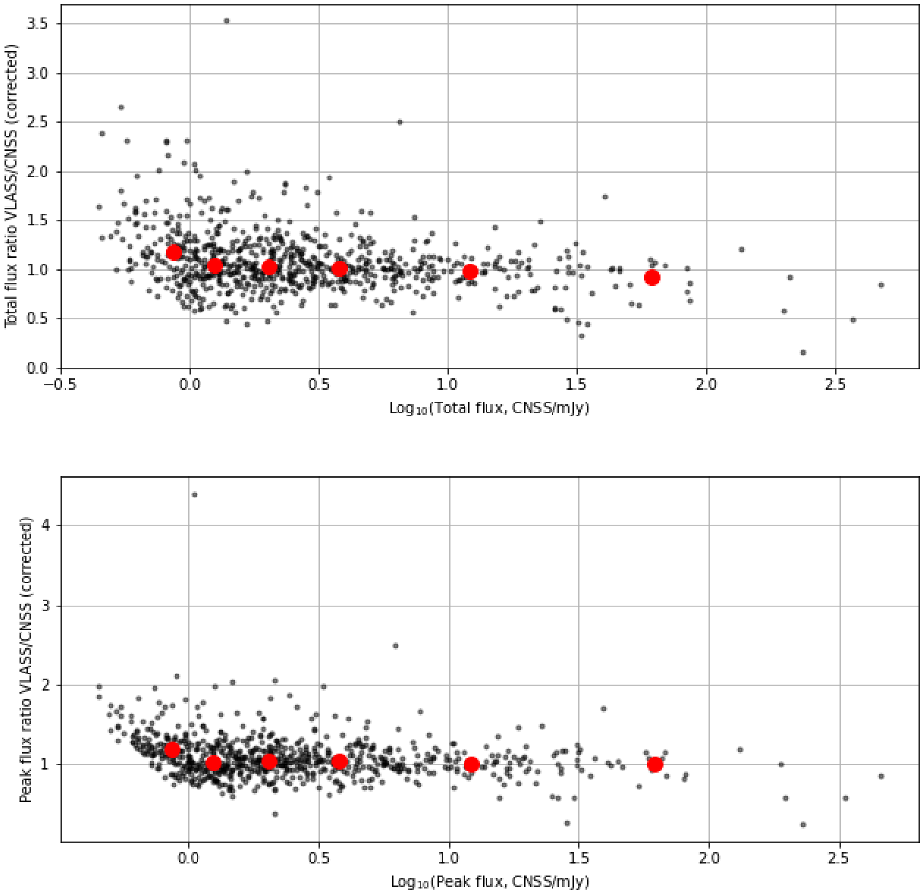

Figure 1: Comparison of the total and peak flux densities of VLASS and CNSS sources as a flux density (note that the ratios have been corrected by 16% to allow for a likely mis-calibration of the flux scale in CNSS). Note that the outliers are likely to be extended sources that are fit with different components by pyBDSF in the two sets of images. The red points show a running median.

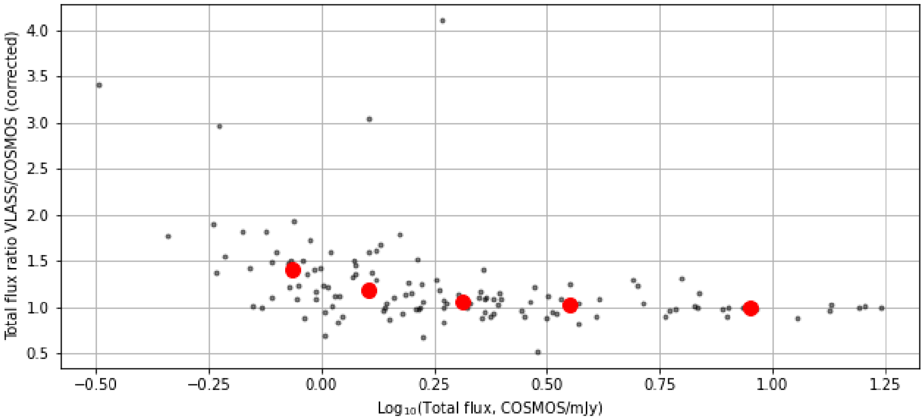

Figure 2: As Figure 1, but for total flux densities the COSMOS field using the survey of Smolčić et al. (2017), where no overall correction of the flux density scale has been applied.

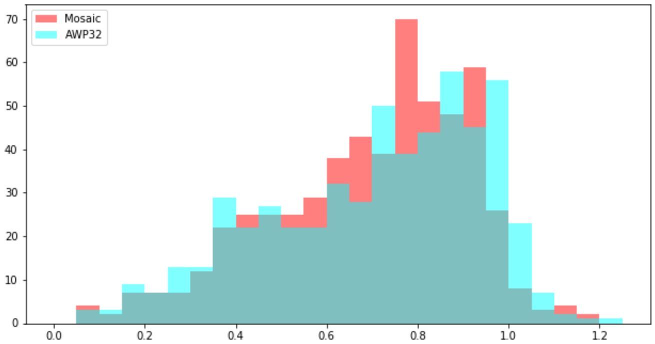

Figure 3: Peak to total flux ratio distributions in CDFS for MOS and AWP (images smoothed to a common 3 arcsec round beam).

The prediction from the simulations is $\approx 2\%$ at the typical Zenith distance this field was observed at, too small to be measured with these data.

2.4 Positional accuracy (S1)

The Gaia mission provides highly accurate (sub-milliarcsecond) astrometry for point sources in the optical. When matching Gaia to VLASS, nearly all the matches are quasars, compact in both the optical and radio. For select fields, we performed a positional match to both Gaia (EDR3), and between images from different imaging algorithms. Note that image-plane position corrections (VLASS Memo 14) were applied to the MOS and AW1 images before analysis. Table 3 shows that both the corrected MOS and the AW32 images have good astrometric accuracy (within ≈ 0.1 arcsec in median offsets in RA and Dec). [3] The scaled Median Absolute Deviation (MAD) differences in Table 3, a statistically-robust measure of the RMS, indicate the typical positional accuracy of a individual source, and are dominated by noise. The relationship between positional error and signal-to-noise, ρ, for a Gaussian source is given by:

$$\sigma(x) = \frac{1}{2}\sqrt{\frac{1}{\ln 2}}\frac{\theta_{\rm FWHM}}{\rho} \approx 0.60\frac{\theta_{\rm FWHM}}{\rho} \quad (1)$$

in each axis (Condon, 1997), so for a typical 10σ VLASS point source observed with a 2.5 arcsec beam, the position error should be about 0.15 arcsec in each axis, corresponding to an RMS radial position error of 0.21 arcsec overall. The scaled MAD values in Table 3 are in reasonable agreement with this.

To investigate the flux dependence of the position errors, we examined the combined catalogs made from the MOS and AW32 processings of the CDFS, CNSS and N-SPARCS fields. Table 4 shows that as the flux density limit/median SNR of the sources is increased, there is an initial decrease in the MAD position uncertainties, but at flux densities $\gtrsim 10$ mJy, corresponding to $\rho \approx 100$, they no-longer follow the $1/\rho$ scaling of Equation 1, and do not fall significantly below 0.1 arcsec, even for $\rho > 100$. As these errors persist for both the AW32 and corrected MOS positions, these are unlikely to be related to the w-terms, and probably arise from phase errors. This explanation is consistent with the magnitudes of the position shifts obtained using the simple model of phase calibration errors discussed in Appendix A.

2.5 In-band spectral index (S2) (compact sources)

In-band spectral indices are calculated using the ratio of the tt1 to the tt0 image. In the catalogs produced from the VLASS SE images, the spectral index value corresponding to the peak pixel of the source component

---

[3] Note that there is a small ($\approx$0.04 arcsec), but statistically-significant, difference between the AW32 and corrected MOS images in the first line of Table 3. Both the position fix to the MOS images and the AW32 algorithm itself are approximations, and this residual offset is within the expected uncertainty of both.

Table 3: Astrometric offsets in the validation fields (using Gaia EDR3). Scaled Median Absolute Deviation(MAD) is a robust estimate of RMS.

Dataset	N pairs	Median RA diff arcsec	Median Dec diff arcsec	Scaled MAD RA diff arcsec	Scaled MAD Dec diff arcsec
CNSS MOS-AW32	2083	0.02	−0.04	0.06	0.08
CNSS AW32-Gaia	237	−0.04	0.02	0.16	0.21
CNSS MOS-Gaia	244	−0.03	−0.01	0.17	0.19
T10t34 AW1-Gaia	77	−0.10	−0.03	0.14	0.26
CDFS AW32-Gaia	100	−0.08	0.05	0.12	0.13
CDFS MOS-Gaia	99	−0.02	−0.04	0.13	0.16
N-SPARCS AW32 - Gaia	87	0.04	0.02	0.15	0.13
N-SPARCS MOS - Gaia	88	0.03	0.00	0.18	0.14

Table 4: Postional MAD versus flux density limit for the combined CDFS-CNSS-NSPARCS catalog.

Flux limit	N (MOS)	RA diff MAD (MOS) arcsec	Dec diff MAD (MOS) arcsec	Median SNR (MOS)	N (AW32)	RA diff MAD (AW32) arcsec	Dec diff MAD (AW32) arcsec	Median SNR (AW32)
None	431	0.16	0.17	20	425	0.16	0.17	19
2mJy	280	0.13	0.15	31	293	0.13	0.15	30
5mJy	154	0.12	0.14	76	156	0.11	0.14	74
10mJy	92	0.10	0.13	122	95	0.10	0.12	118
20mJy	51	0.08	0.10	193	59	0.10	0.14	178
40mJy	27	0.13	0.12	379	26	0.09	0.18	334

is used unless otherwise stated. This section only discusses compact source components, see Section 2.8 for a discussion of extended sources.

Figure 4 compares the in-band spectral index distributions in the CNSS equatorial field and the southern CDFS field for source components matched within 1.5-arcsec and total flux densities > 2 mJy for the MOS and AW32 gridders. In CNSS, the distributions from the two gridders are slightly offset (with a statistically significant difference using a K-S test), but the difference between the medians is <0.2 and neither is strongly preferred on the basis of what is known about the true distribution of spectral index at this frequency and flux density level. In the CDFS field, however, it is clear that the spectral index distribution using the MOS gridder has shifted by about $+0.2$ relative to that for the CNSS field, and thus is suspect. The results for calibrators are consistent with the conclusion that the MOS gridder is producing spectral indices for point sources that are flatter than the truth (from pointed observations) by about 0.15–0.2 (see Table 2).

We also examined the effect of setting conjbeams to True in the mosaic gridder in the T10t34 field (MOS CBT), this steepened the median spectral index to -1.25 (compared to -0.78 for the mosaic gridder with conjbeams False). This is steep compared to all the other estimates, and too steep on physical grounds.

We compared the in-band spectral indices of VLASS sources in CNSS to those between FIRST and CNSS (Figures 5, 6). Given that most sources are expected to have power-law spectra, this shows that both imaging methods result in reasonable spectral indices compared to FIRST.

One adjustment to the default MOS CBF spectral indices that can be made (in catalog space at least) is to average the spectral index values of the pixels around the component peak in tt0, rather than just take the value corresponding to the peak pixel. As shown in Table 5, this tends to steepen the MOS spectral indices and bring them closer to the AW32 values. For observations at low zenith distances, the 3x3 box results in a closer agreement between the MOS and AW32 spectral index, but overall the 5x5 box works better, and does not steepen the mean in the low zenith distance fields by more than -0.1 compared to the AW32 value. This benefit of averaging the spectral index values is in line with the expected effect of conjbeams False described in VLASS Memo 14, and a 5x5 average has been recommended for the SE catalog generation.

A simple model for the effects of w-terms on the in-band spectral index is discussed in VLASS Memo 14, but the results are summarized here for completeness: uncorrected w-terms cause a chromatic abberation effect in both the tt0 and tt1 images that scales with the tangent of the zenith distance. The effect in tt0 is at most only a few percent (consistent with our observation in Section 2.3.2), but that in tt1 with CBF is very significant once the zenith distance exceeds about 45 degrees. (This effect is correctable in the cubes

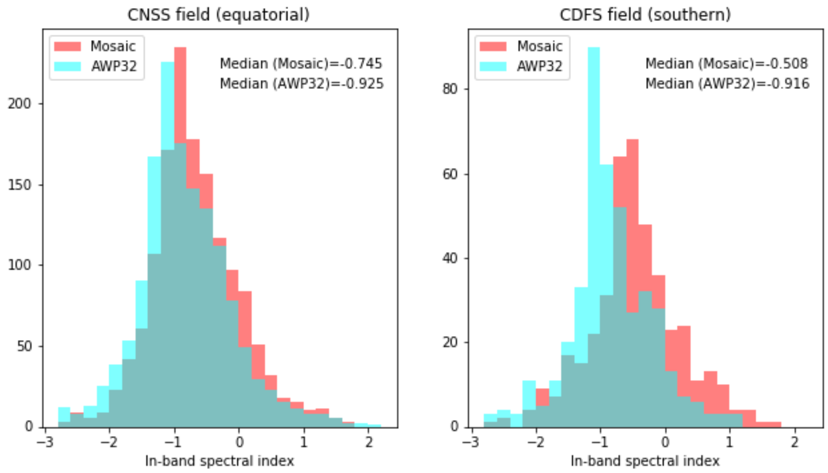

Figure 4: Spectral index distributions compared for matched VLASS sources made using AW32 and MOS (peak pixel only, not averaged) in *left* the CNSS field and *right* the CDFS field.

Table 5: Effect of changing the box size for the MOS spectral indices (for sources >2mJy).

Field	AW32 median(pk)	MOS (pk)	MOS (3x3)	MOS (5x5)
NSPARCS	-0.89	-0.80	-0.83	-0.93
T10t34	-0.92	-0.81	-0.88	-1.01
CDFS	-0.92	-0.58	-0.63	-0.90

by shifting each plane by a frequency-dependent position correction.)

2.5.1 Comparison to in-band spectral indices from a pointed survey in CDFS

By comparing the in-band spectral indices from VLASS to those from a pointed survey, any possible effects from the determination of spectral indices from the OTF versus pointed observations can be made. As part of a science program to investigate sources in the upcoming LSST deep-drilling field located in the wide Chandra Deep Field South (W-CDFS), 10 deg^2 of sky were observed on the VLA in C-configuration in S-band (2–4 GHz) as program 21A-017 between 2021-07-24 and 2021-09-09 (about 10 months after the field was observed in VLASS2.1). 511 pointings were observed for an integration time of ≈2 min per pointing. The data were imaged using the mosaic gridder (at the ≈ 10-arcsec resolution of the C-configuration data the effects of w-terms can be neglected). The RMS reached in the final mosaic was ≈ 50μJy. We matched 431 sources to the VLASS SE MOS continuum image. Figure 7 shows the comparison of in-band spectral indices from this study (both datasets had their spectral indices measured by averaging the values in the spectral index maps in a 5×5 pixel box around the peak position). To search for systematic effects, we selected only objects that were bright ($>$ 4mJy) in the C-configuration data, and isolated (source code "S" in pyBDSF) and compact (peak flux/total flux $>$ 0.7) in both VLASS and the C-configuration data. This resulted in a sample of 27 sources (black crosses in Figure 7). The median spectral index difference (VLASS – C-configuration) for these objects was 0.04±0.04, and scaled median absolute deviation of the spectral index difference was 0.22, indicating that systematic effects, even in this southern field that was observed at high zenith distance, are within the survey specification. The Figure also shows that the in-band spectral indices also compare well to spectral indices obtained by cross-matching the RACS survey (McConnell et al., 2020) with the two VLA datasets.

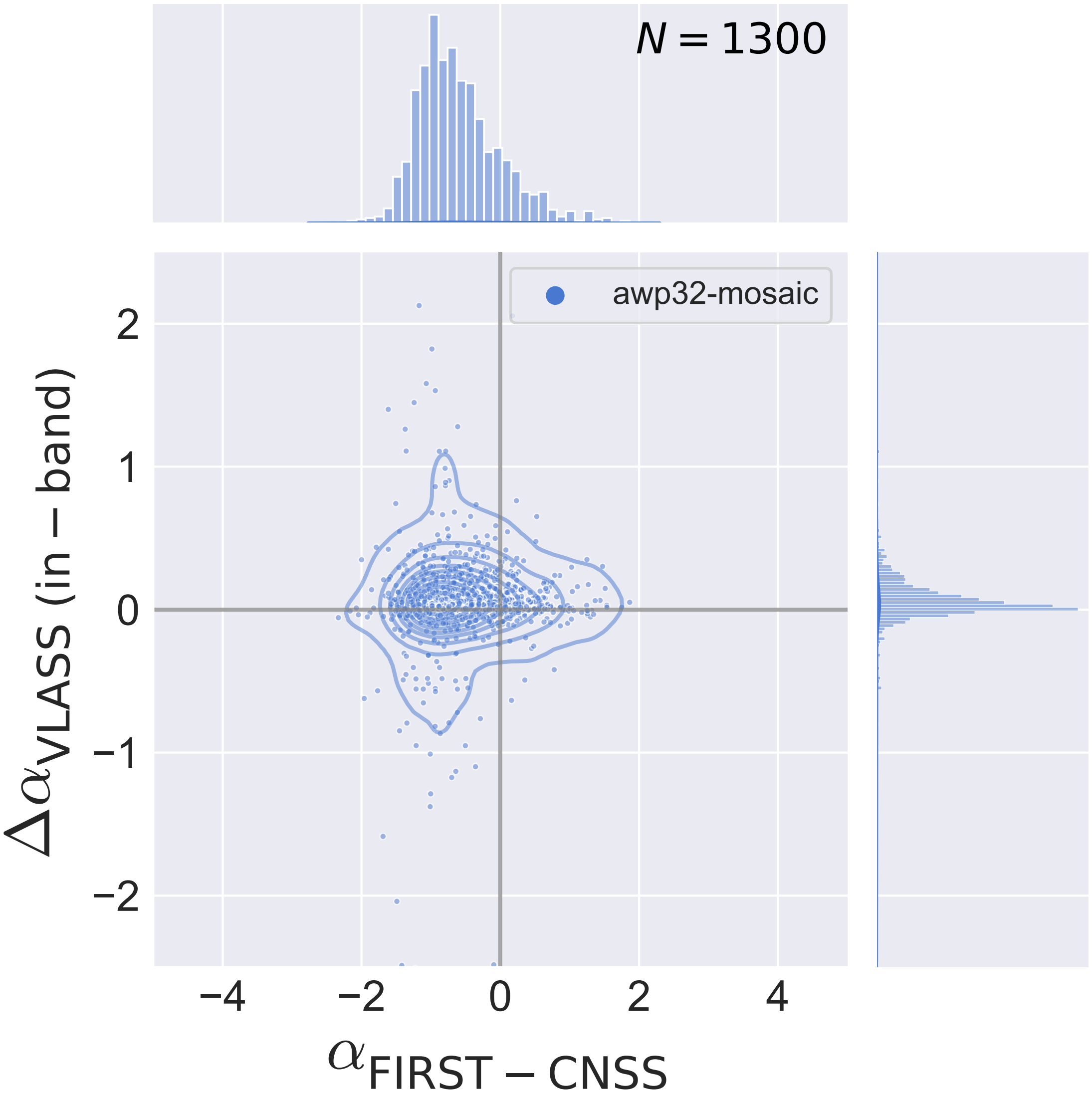

Figure 5: Spectral index of sources detected in both FIRST and the CNSS equatorial field vs./ the difference in the VLASS in-band spectral index between the AW32 and MOS gridders (i.e. in the sense AW32 minus MOS).

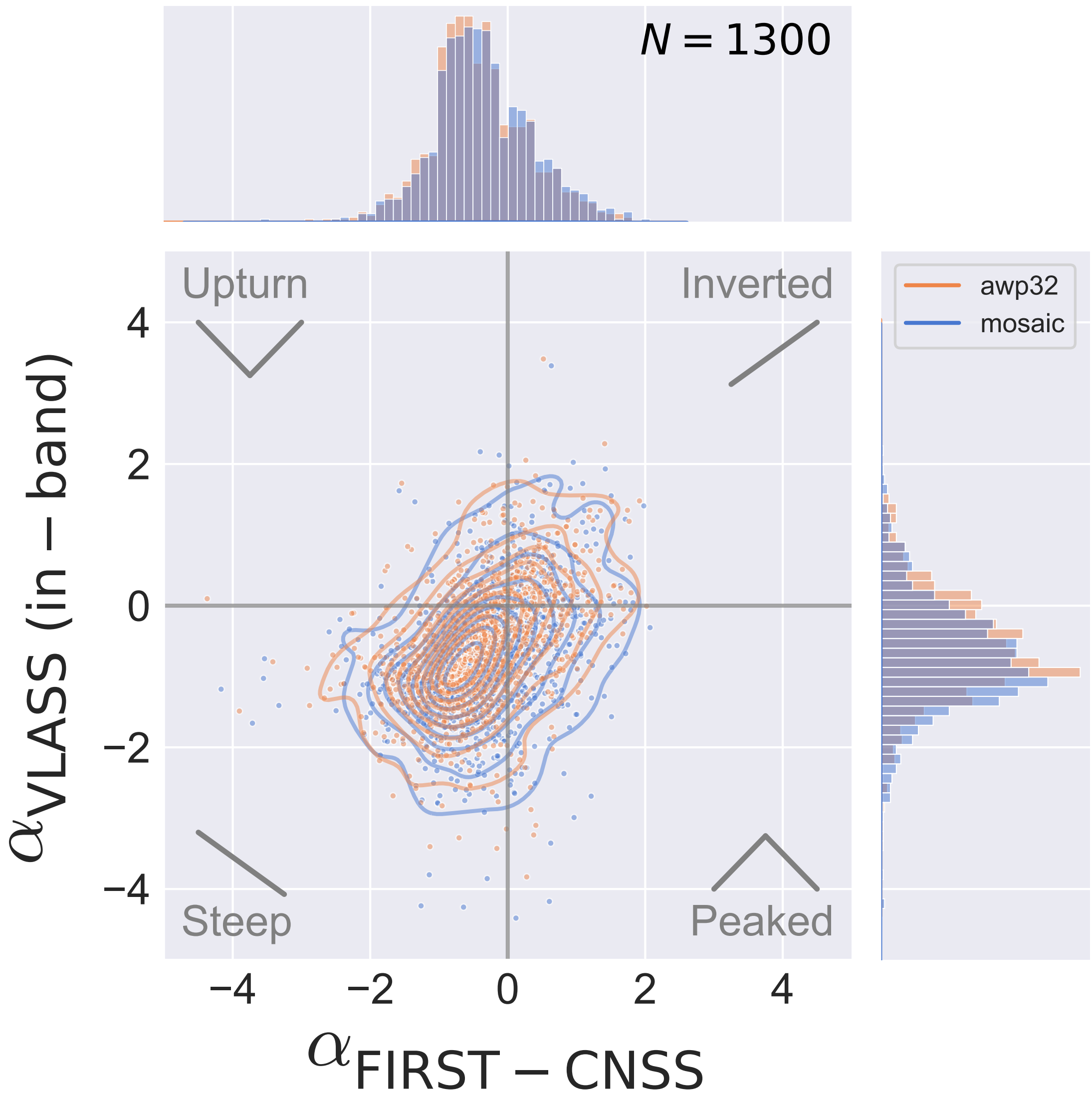

Figure 6: Radio color-color plot showing the FIRST–CNSS vs./ VLASS in-band spectral indices for the AW32 (orange) and MOS (blue) gridders.

2.5.2 Spectral index uncertainties for compact sources

The tests documented above suggest that, in addition to the uncertainty on the spectral index due to the noise in the tt1 image, an additional component to the uncertainty exists that does not decrease with increasing signal-to-noise. The likely origins of this are either in the imaging algorithm, or in the bandpass calibration and its transfer to the target fields. Further studies of the image cubes should shed light on this. For now, we note that the spectral index of calibrators and the comparison to the CDFS survey and RACs described above are consistent with the value of this noise-independent addition to the uncertainty, $\Delta\alpha_{\rm ni} \approx 0.2$, and thus the spectral index uncertainties for compact sources being given by:

$$\Delta\alpha \approx \sqrt{0.2^2 + (6/\rho)^2}, \tag{2}$$

where ρ is the signal-to-noise of the source and the factor five in the numerator of the right-hand term reflects the additional noise in the tt1 image compared to the tt0. We note that the CIRADA team recommends the use of a more complicated form for $\Delta\alpha_{\rm ni}$ (rather than simply assuming a constant value of 0.2) based on a comparison to FIRST:

$$\Delta\alpha_{\rm ni} = 0.144 + \frac{0.08}{\log_{10}(S_\nu)}. \tag{3}$$

In practice both formulae lead to similar error estimates at most values of flux density.

2.6 Dynamic range (S3a & S3b)

Two measures of dynamic range are calculated. The "near-field" dynamic range is the absolute value of the largest peak in an annulus from 9 to 15 arcsec (approximately three to five beams from the source). The far-field estimate is the ratio of the peak to the noise in an annulus from 180 to 200 arcsec. These measures were picked so that the near-field estimate represents the largest artifact that is likely to be found near a bright source, and the far-field estimate represents the increase in overall noise in the vicinity of a bright source.

For both measurements, a cut at total flux of 70 mJy is imposed on the input catalog (which limits the dynamic range at ≈440 for an object at the flux limit and a typical noise of 160 μJy). Most of the VIP images (with either algorithm) satisfy the dynamic range requirements for near-field and far-field dynamic range (50 and 400, respectively), with the exception of some fields in the Galactic Plane, where the brightness and complexity of sources makes dynamic range difficult to estimate.

2.7 Image fidelity of extended sources (S9)

We require that the difference between VLASS and a pointed observation of the same source be the same within the noise. As originally defined in the requirement, this noise was the RMS in the far-field, however, it is clear in practice that this significantly underestimates the uncertainty in the surface brightness of extended sources. This is due to the contribution of multiplicative terms that arise from errors in the complex gain calibration and other issues that scale with source flux density, e.g. primary beam uncertainties, antenna pointing errors etc. We therefore assume that uncertainties in surface brightness for extended sources can be expressed as the combination of an additive noise term from the antenna/receiver system, $\sigma_{\rm sys}$ Jy/beam (which can be obtained by measuring the RMS in the image far from bright sources), and a multiplicative term due to complex gain calibration and antenna pointing errors, $f_g\, S$, where S is the surface brightness in Jy/beam (see Condon (2009) [4] for an analogous discussion, and Perley (1999) for an analytic treatment of dynamic range). Then, adding these in quadrature, we obtain the surface brightness uncertainty,

$$\sigma_S = \sqrt{\sigma_{\rm sys}^2 + (f_g S)^2}. \tag{4}$$

The factor f_g is hard to estimate without accurate simulations as it depends on the very near-field dynamic range. For data meeting the survey requirements for near-field dynamic range (50:1), $f_g < 0.02$ three beams from a bright source, but most of the noise originates closer in than that, within 1-2 beamwidths.

---

[4] `https://www.skatelescope.org/uploaded/12336_114_Memo_Condon.pdf`

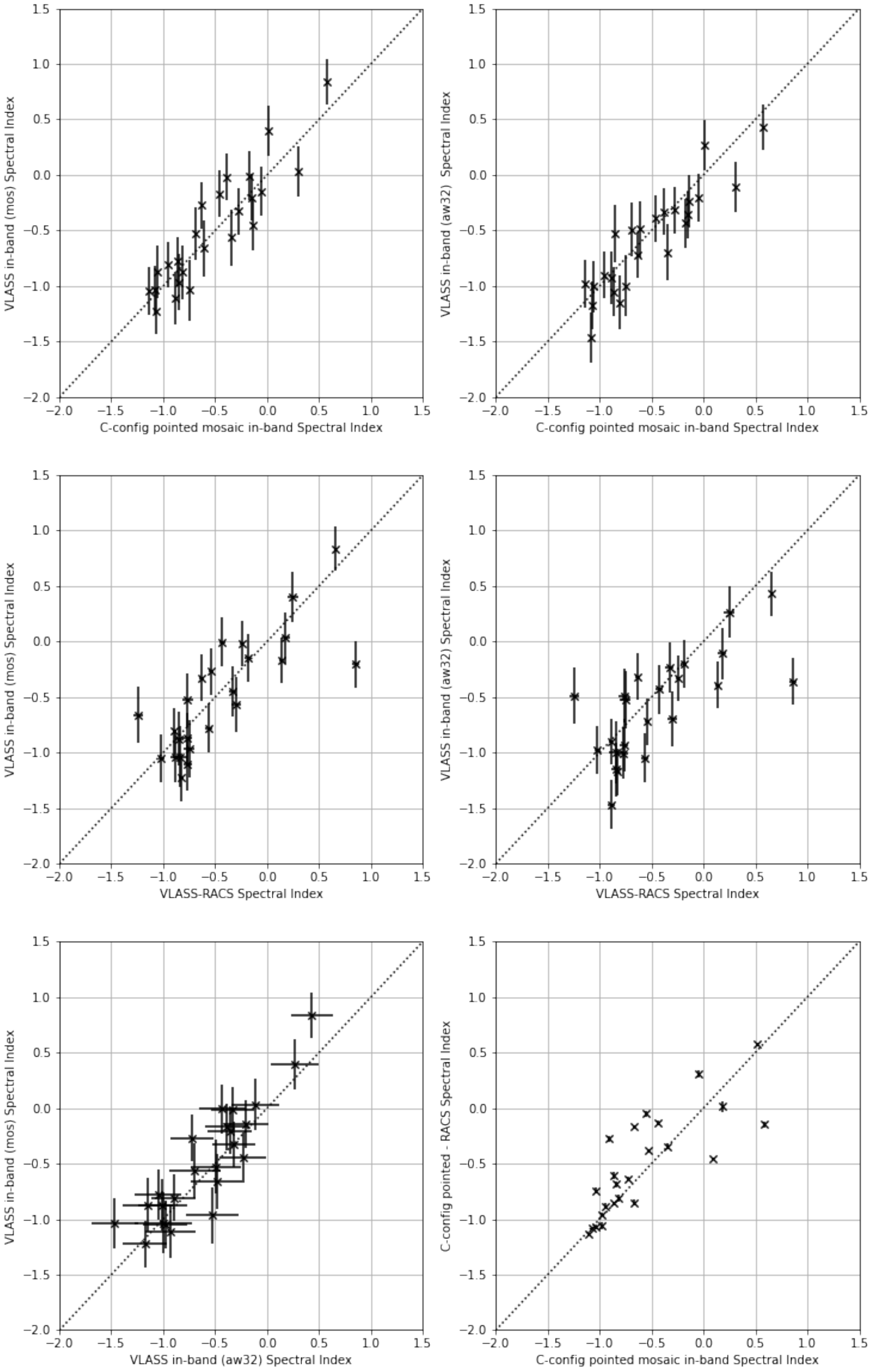

Figure 7: Spectral index measurements for bright (> 4 mJy), isolated sources in the 10 deg^2 of the W-CDFS (centered on 03hr30m -28d30m) from VLASS and the C-configuration pointed mosaic from project 21A-017. *Upper Left:* VLASS in-band spectral index (MOS) versus C-configuration in-band. *Upper right:* VLASS in-band spectral index (AW32) versus C-configuration in-band. *Middle left:* VLASS in-band spectral index (MOS) versus RACS-VLASS spectral index. *Middle Right:* VLASS in-band spectral index (AW32) versus RACS-VLASS spectral index. *Bottom left:* VLASS in-band (MOS) vs VLASS in-band (AW32).*Bottom right* RACS to C-configuration spectral index vs C-configuration in-band spectral index. Note that error bars on the C-configration in-band spectral indices are not plotted for clarity, but are expected to be similar to the VLASS in-band uncertainties.

Simulations using the simple model described in Appendix A, however, also suggest typical values of the very near-field residuals of only a few percent in the self-calibrated SE images (unless spurious structures from QL phase errors are "baked in" by the self-calibration process). To obtain f_g for VLASS, we therefore make an empirical estimate based on comparisons to pointed data.

In Figure 8 we show the difference images between the VLASS MOS, AW32 and AW1 images, and the pointed image from CNSS (shifted by the position offset between the CNSS and VLASS, Section 2.4). We also show the difference images divided by $\sigma_{\rm sys}$. All images were convolved to a common 3.5×2 arcsec beam. For the most part the images agree very well, especially the two realizations of the VLASS data. However, near strong brightness gradients, for example along the ridge in the Southern lobe, it is clear that differences exceed $\pm 5\sigma_{\rm sys}$ when the VLASS and pointed data are compared. That this difference is driven by calibration errors is suggested by the fact that the subtraction of the two VLASS images, though made with different gridding algorithms and cleaning methods, shows no such strong differences.

To make this more quantitative, and to try to make an estimate for the value of f_g, in Figure 9 we plot the histograms for the difference between the images divided by the uncertainty for image pixels above 1 mJy in the pointed CNSS image, and compare to the model of a Gaussian with unit σ and zero mean. This clearly shows that $\sigma_{\rm sys}$ is much too small to account for the observed variations, but by setting $f_g = 0.1$, we can obtain histograms very close to the expected distribution (though this is based on only one source, further tests are needed to see how universal this estimate is). In Figure 10 we show the same plots, but this time divided by σ_S with $f_g = 0.1$, showing that when the multiplicative errors are included, there is no clear excess noise remaining near strong brightness gradients.

2.8 Spectral index fidelity of extended sources (S10)

2.8.1 Difference images

Analysis of the spectral indices of extended sources has revealed issues with the in-band spectral indices derived from the continuum in both the VLASS and comparison pointed data from CNSS (Figures 11, 12). Using the spectral index map from a VLASS cube as a fiducial, we see significant ($> 3\sigma$) deviations from the expected system noise (dominated by the noise in the tt1 image, which is ≈ 6 times the noise in tt0, see Condon (2015)) compared to in-band spectral index images from both the CNSS pointed data and VLASS, especially around the edges of bright regions. Their appearance in both the pointed and VLASS mosaic data, where they have similar magnitudes, but different spatial distributions, suggest they are not an artifact of VLASS, but rather generic to the approach of estimating spectral indices within the S-band.

Examining the images in Figure 11 in more detail, it seems that the areas of brighter emission agree better in the cubes, for example, in the pointed cube versus VLASS cube comparison (top right), the core and the northern hotspot agree well. It is worth noting that the two highest frequency spectral windows in the pointed data had to be flagged due to RFI, so some differences, particularly in the lower brightness regions (which are significantly flatter in the pointed cube) are perhaps to be expected.

This analysis suggests that the in-band spectral indices for the SE continuum imaging are likely to be very sensitive to differences in RFI flagging across the band. We also note that the CASA documentation [5] says that in-band spectral indices are limited in accuracy, especially for extended sources, where the accuracy of the spectral index maps depends on the range of clean scales used to model the emission.

2.8.2 Effect of chromatic aberration on the MOS spectral indices

In Figure 13 we show the spectral index maps for a source in the southern CDFS field (J0328−2841; PKS 0326−288). The MOS CBF image shows clear issues from the chromatic aberration effects, the other images are more consistent with each other, and with the AW32 image.

2.8.3 HII regions

HII regions are dominated by thermal (free-free) emission with a spectral index of -0.1 in the optically-thin limit. They therefore provide a good test of the spectral index accuracy for bright, extended sources. We picked IRAS 17160-3707 (G350.10+0.08) as a test to compare MOS and AW32. Independent observations

---

[5] https://casadocs.readthedocs.io/en/latest/notebooks/synthesis_imaging.html#Options-in-CASA-for-wideband-imaging

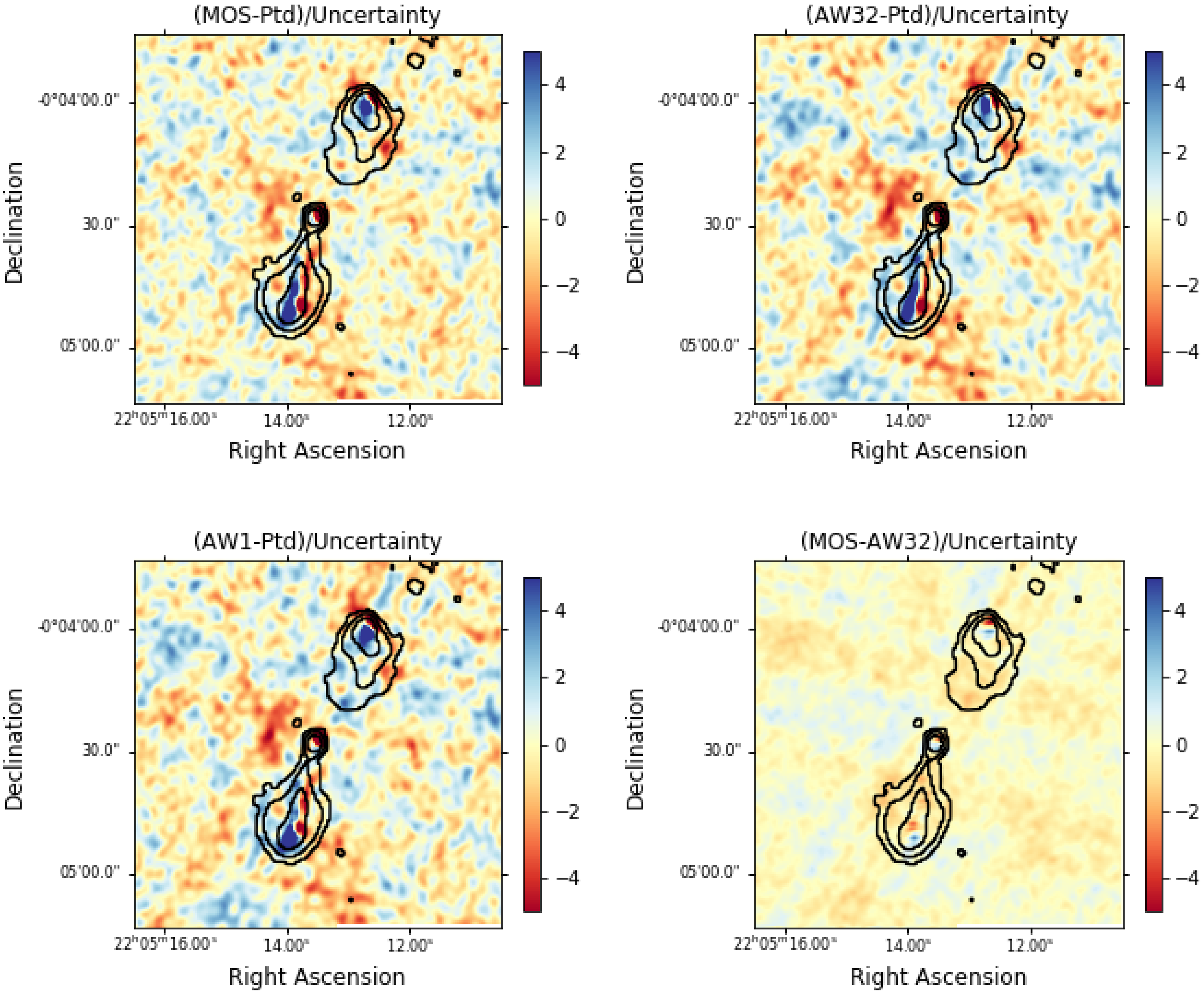

Figure 8: Comparison of residual images divided by the system noise $\sigma_{\rm sys}$. Note that the differences between VLASS and pointed images from CNSS are much larger relative to the noise than those between the two realizations of the VLASS images, seen at bottom right.

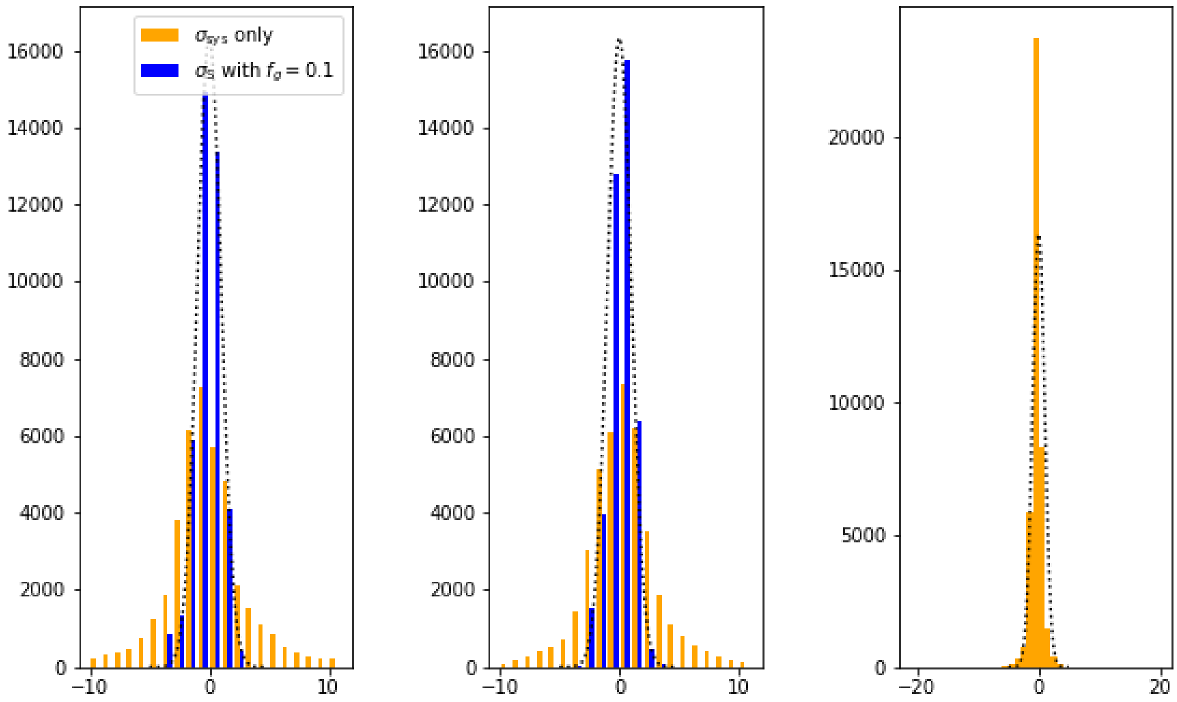

Figure 9: Histograms of the flux differences divided by uncertainties for *left* VLASS MOS minus CNSS pointed, *middle* VLASS AW32 minus CNSS pointed, VLASS MOS minus VLASS AW32. The orange bars show the system noise only, the blue bars on the left and middle plots the noise including a component that scales with flux density. The black dotted line in all three plots is a Gaussian with mean zero and $\sigma = 1$.

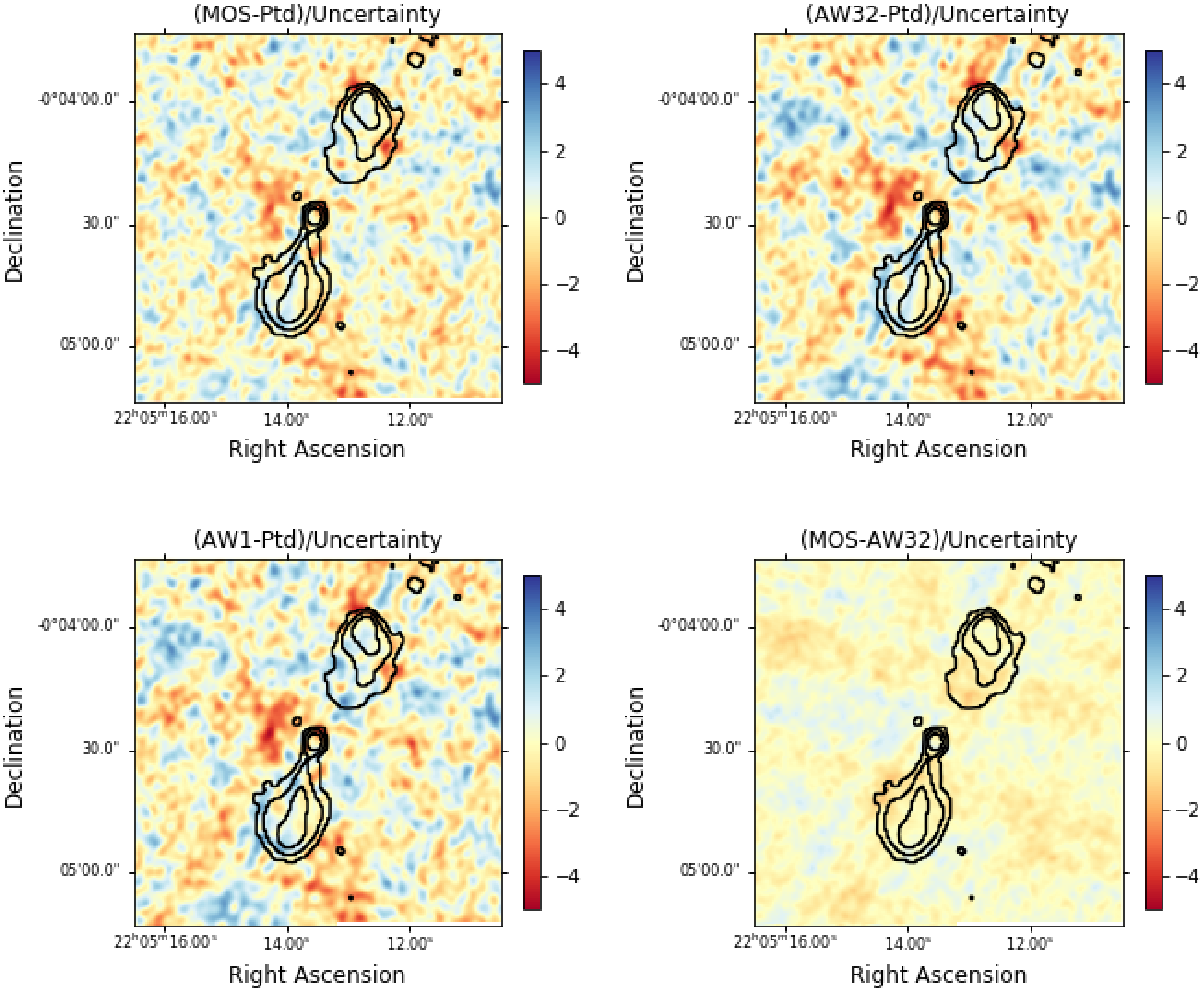

Figure 10: As for Figure 8, but this time using the full noise model (including complex gain calibration errors as an additional multiplicative term).

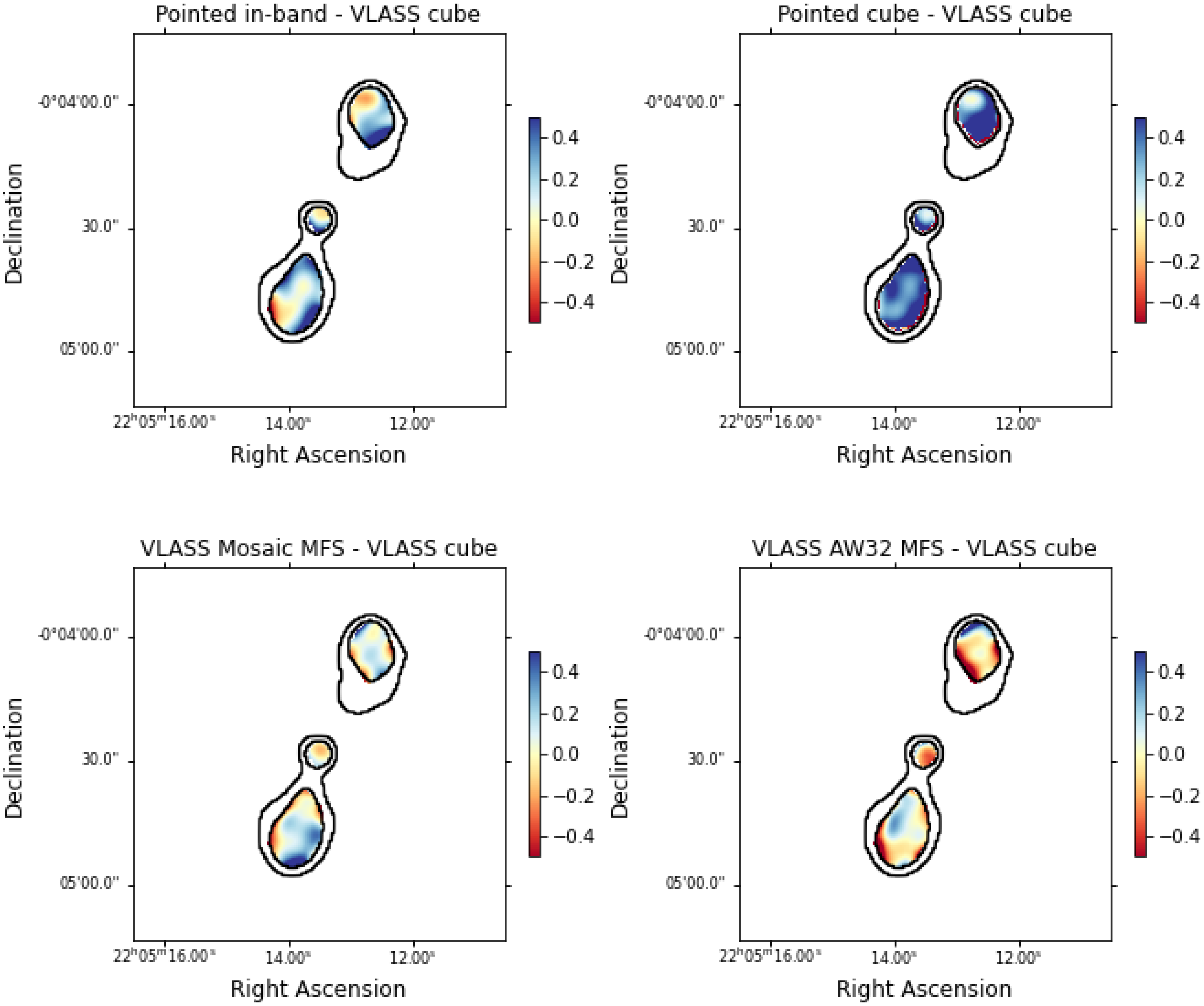

Figure 11: Pointed cube to VLASS cube and in-band spectral index comparison. All images were smoothed to the 4.5 arcsec resolution of the pointed cube before subtraction.

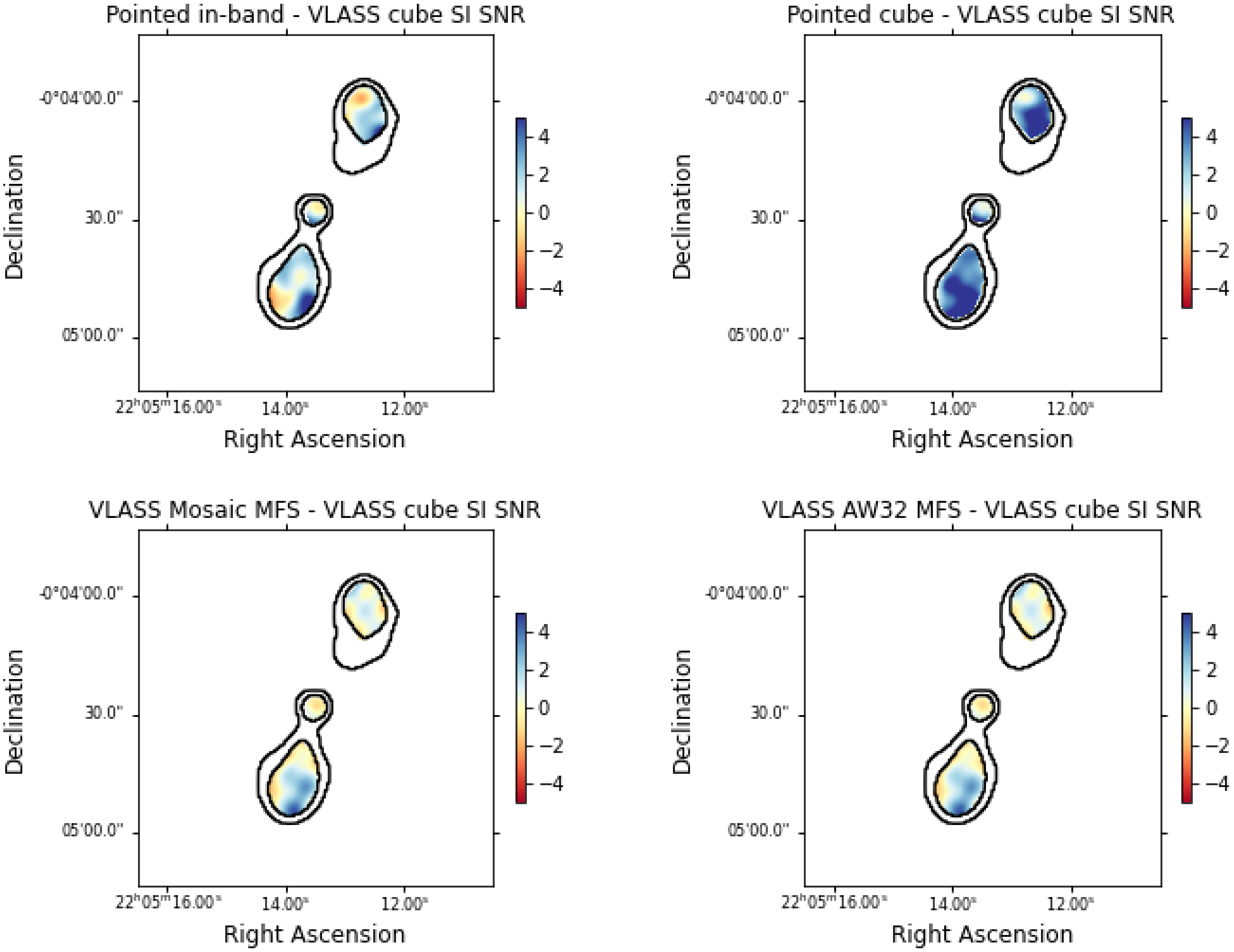

Figure 12: As Figure 11, but with the difference images divided by their uncertainties.

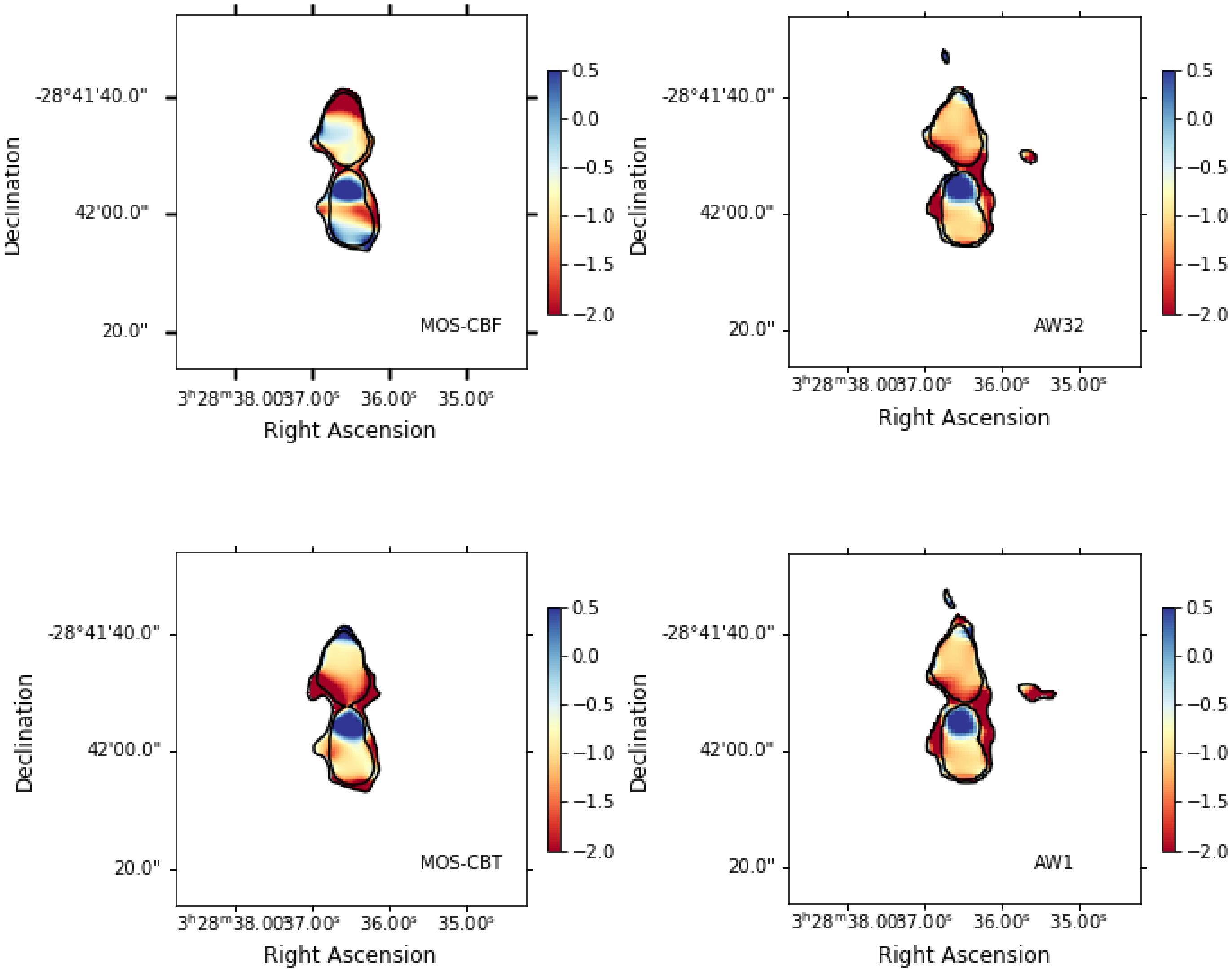

Figure 13: Spectral index comparisons for a bright source in the southern CDFS field, showing the issues arising from the neglect of w-terms in the imaging (apart from AW32).

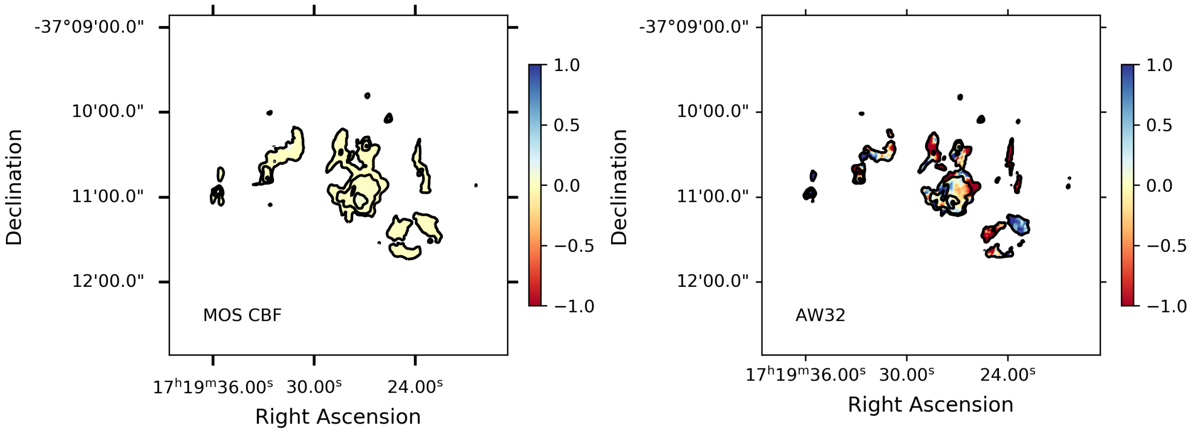

Figure 14: The HII region IRAS 17160-3707 imaged in MOS CBF and AW32.

with the GMRT (Nandakumar et al., 2016) show that indeed this object has a thermal spectrum throughout. Figure 14 shows that the AW32 spectral index is much noisier, with a standard deviation of 1.25 (though does average to a value of -0.08, very close to the correct number). In contrast, the mosaic CBF image is much more uniform, with a standard deviation of 0.01, though the mean spectral index is slightly incorrect (+0.01). This is consistent with the offsets discussed above, and also suggests that the chromatic aberration effects in MOS CBF does average out over a large, extended source, even in the far South.

2.8.4 Flux-weighted integrated spectral index for an extended source

For J0328−2841 we also constructed the flux-weighted spectral index (summing over the i pixels in which emission is detected):

$$< \alpha >= \frac{\Sigma \alpha_i S_i}{\Sigma S_i} \tag{5}$$

for the source using the MOS, AW32 and 16-plane cube spectral index maps to compare to a second-order fit in log(frequency) vs log(flux) to the literature flux values from NED at 365, 408, 1400, 2700, 5000 and 8400 GHz and the TGSS ADR (150 MHz; Intema et al., 2017) (Figure 15). This fit gives the spectral index at 3.0 GHz of -0.7673 ± 0.0005, which is reproduced well by the 16-plane cube (-0.78), compared to the AW32 SE continuum image (-0.99) and the Mosaic gridder (-0.85). Despite the issues with chromatic aberration in the MOS image, it still does slightly better than the AW32 image. Both are within the requirements, however.

2.9 Image artifacts (S11)

Image artifacts can arise from poor cleaning, residual RFI (though these seem very rare, with no clear examples in the 100 deg^2 release) or poor calibration. It is important to note that the size of VLASS images means that some apparently significant noise fluctuations are inevitable, even assuming Gaussian statistics. In a 1 deg^2 VLASS image there are ≈660,000 synthesized beams (resolution elements). This means there is about a 20% chance of picking up a random $+5\sigma$ fluctuation in a square degree. This probability increases rapidly towards lower significance, with 21 $+4\sigma$ fluctuations per square degree, and 891 $+3\sigma$ fluctuations. We therefore focus on artifacts above 5σ in the following discussion.

2.9.1 Clean artifacts around extended sources

In general, clean artifacts above 5σ are not common outside of the Galactic Plane, however, during the validation process versus pointed mosaics in the CNSS (30 deg^2), XFLS (4 deg^2) and COSMOS (2 deg^2) we noticed a few clear issues itemized below. One important difference between the MOS and AW1/AW32

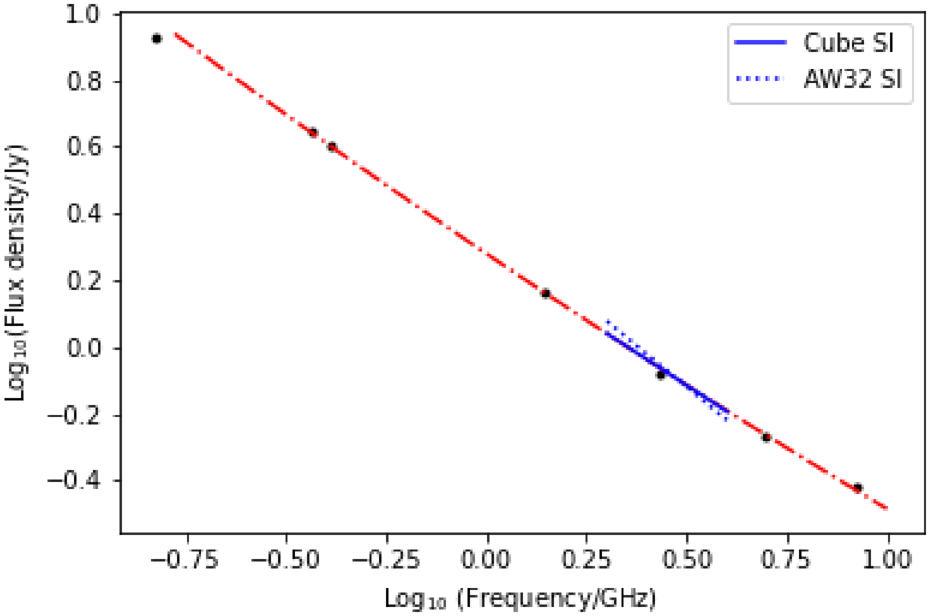

Figure 15: Spectrum of the bright FRII source PKS 0326-288 from NED. The red dot-dashed line is a 2nd-order fit in $\log(\nu)$. The blue solid line is the flux density weighted spectral index from the 16 plane cube, the blue dotted line is the same from the AW32 SE continuum image (the MOS spectral index lies between the blue solid and blue dashed lines).

images was that for MOS the final clean was performed without masking, whereas a mask was used for the aw-project images. (This masking was needed to stabilize the clean for some of the AW-project images.)

- Edge-on spiral galaxy fails to clean correctly: J 171841+603620 (NGC 6361) in XFLS AW32 (similarly bad in QL, maybe issue with being aligned with arm response) (Figure 16).
- Incomplete cleaning in extragalactic sources with faint lobes (Figure 17).

Some of the artifacts are clearly related to lack of clean depth, for example, J220516.4-004249 is (correctly) cleaned down to $4\sigma \approx 600\mu$Jy in the J220600-003000 MOS image, but the large amount of diffuse lobe emission means that clean residuals are still visible. The same source in the AW32 image was cleaned to about 500μJy in the NE lobe, but the SW lobe seems to have been missed from the mask, resulting in a peak residual $\approx 850\mu$Jy. Inside the Galactic Plane, the rate of clean artifacts is much higher in AW32 than MOS. We note too that in many cases we see an enhancement of residuals about 3 arcmin from extended sources both in MOS and AW32 (e.g. Figure 16). This seems to be related to a bump in the PSF diffraction spikes at that radius (up to about 9% in MOS, compared to a typical amplitude of the spike of about 5%).

2.9.2 Artifacts around point sources

The self-calibration employed by the VIP script was generally very successful, increasing the near-field dynamic range by a factor of ≈2 compared to Quicklook. A few issues were seen though with the VIP implementation of self-calibration, we believe these are because the initial clean went too deep and “baked in” some poor phase solutions. This issue has been addressed in the pipeline. Nevertheless, many bright ($\gtrsim$ 100mJy) sources show strong diffraction spikes that are not removed by cleaning in either the MOS or AW images. These are typically restricted to within the second null of the primary beam (≈ 30-arcmin radius), but in cases of very bright, or poorly calibrated sources can extend across most of a 1 deg^2 image. Also, the originating source can often be off the 1 deg^2 subimage. These can lead to artifacts above 5-σ

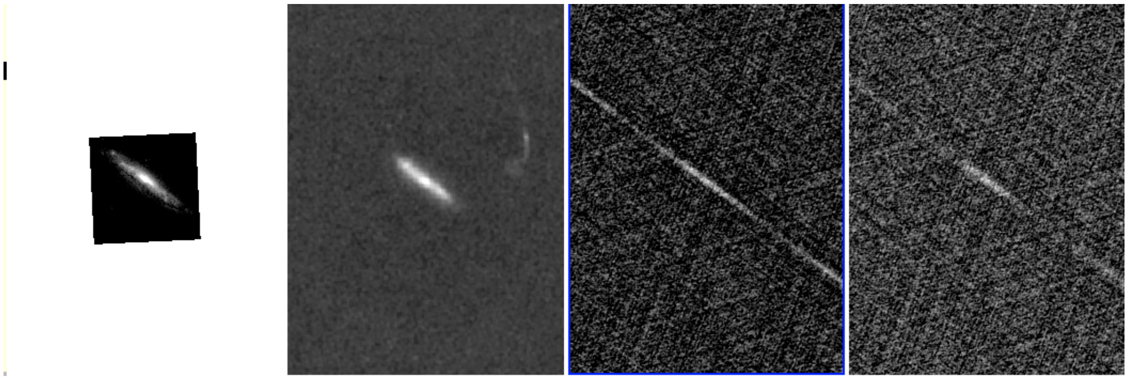

Figure 16: NGC 6361: *left* PanSTARRS g-band (2-arcmin in size), *center-left* VLA image from the survey of Condon et al. 2000, *center-right* VLASS AW32 image, *right* VLASS mosaic gridder image.

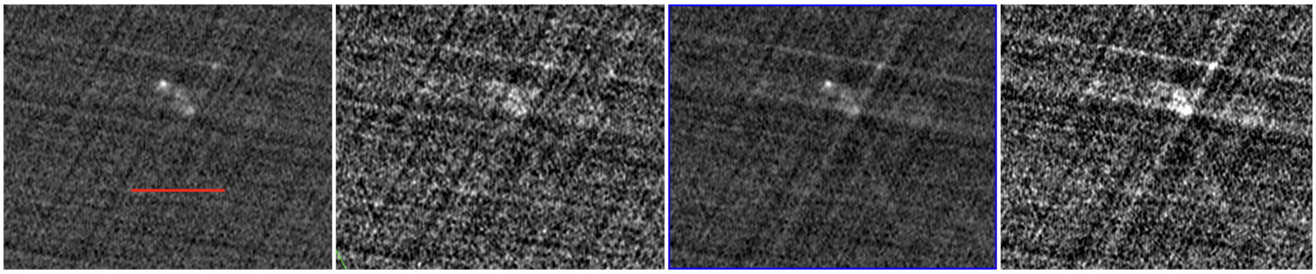

Figure 17: J220516.4-004249, an FRII with faint lobes in J220600-003000. Left to right: MOS image, MOS residual, AW32 image, AW32 residual. The length of the red scalebar is 1-arcmin.

in the mosaic. These artifacts are probably explicable via a combination of residual phase and amplitude calibration errors, (Appendix A). As their properties are fairly predictable (small range of position angles from originating source, with highly elongated morphologies) it is possible to flag likely artifacts. Indeed, there has been some success flagging these in CIRADA QL catalog using a "peak to ring" metric (Gordon et al., 2021).

2.9.3 Rate of $> 5\sigma$ artifacts

To quantify the rate of $> 5\sigma$ artifacts we examined two VLASS 1 deg^2 images, J220600-003000 in CNSS and J100200+023000 in COSMOS. A mask of SNR > 5 pixels was constructed by multiplying the RMS image by five, subtracting it from the data and masking pixels with values < 0. The masks were then compared by blinking with "truth" images from the CNSS survey and from COSMOS. In J220600-003000 three likely artifacts were found, two associated with PSF residual spikes and one in the field. In J100200+023000 one candidate artifact was found at $\approx 5\sigma$, but determined to be a likely true transient as it lies within $0.6''$ of a source detected in HST/ACS and ULTRAVISTA. We therefore estimate an artifact rate of ≈ 2 deg^{-2}, though because simple catalog flagging such as the peak-to-ring flagging criterion described in Section 2.9.2 can remove many of the bright source artifacts, the artifact rate after flagging is likely to be only ≈ 1deg^{-2} (corresponding to $\approx 98\%$ reliability for a catalog cut at 5σ). (For the cumulative images, we expect the multi-epoch beam to be significantly better than each single epoch, and thus the rate of artifacts to drop in the cumulative image data if imaged by combining visibilities from each epoch rather than simply combining the images, though that is something that needs to be tested as part of the multi-epoch development.)

3 Parameter and algorithm tests

A number of tests with different versions of the AWP imaging algorithm were made.

Table 6: RMS, beamsizes and dynamic ranges for the validation fields

Dataset	Size deg^2	RMS (tt0) (μJy)	RMS (tt1) (μJy)	Mean beamsize (arcsec)	Beam AR -	DR near -	DR far -	Median Spectral Index (> 2mJy)	No. sources
CNSS (T10t34) QL	11	155	-	2.51	1.28	78	481	-	
CNSS (T10t34) MOS	11	152	874	2.47	1.29	147	527	-	
CNSS (T10t34) MOS-CBT	11	174	1035	2.14	1.43	106	387	-1.25	862
CNSS (T10t34) AW1	11	160	1030	2.30	1.37	126	451	-0.95	919
CNSS (T10t34) AW1-ATF	9	160	1029	2.30	1.42	105	411		
CNSS (T10t34) AW1-PTGF	9	160	1030	2.30	1.38	131	451		
CNSS (T10t34) AW32	11	159	1028	2.30	1.34	119	436	-	
CNSS (T10t34) SEPLMOS	11	139	798	2.48	1.31	152	581	-0.71	992
CNSS (T10t35) MOS	5	186	1087	2.61	1.27	112	414	-	
CNSS (T10t35) AW32	5	196	1293	2.41	1.24	120	400	-	
CNSS (T10t35) SEPLMOS	10	168	974	2.61	1.20	86	447	-0.64	763
CNSS (T11t34) MOS	10	170	989	2.60	1.15	167	466	-	
CNSS (T11t34) AW32	10	181	1184	2.16	1.17	154	435	-	
CNSS (T11t34) SEPLMOS	10	154	885	2.59	1.15	159	460	-0.63	882
CNSS (T11t35) MOS	5	163	956	2.63	1.38	128	442	-	
CNSS (T11t35) AW32	5	172	1137	2.45	1.43	119	436	-	
CNSS (T11t35) SEPLMOS	5	151	887	2.64	1.41	111	474	-0.62	423
CNSS (all) MOS	31	-	-	-	-	-	-	-0.78	2617
CNSS (all) AWP32	31	-	-	-	-	-	-	-0.71	2503
T01t35 MOS (S. Gal. Plane)	7	171	-	2.59	1.32	-	-	-	
T01t35 AW32 (S. Gal. Plane)	7	180	1042	2.40	1.32	-	-	-	
XFLS (T25t13, T25t24) MOS	5	148	767	2.68	1.21	147	541	-0.78	353
XFLS (T25t13, T25t24) AW32	5	162	935	2.54	1.29	176	543	-0.77	350
CDFS (T03t06, T04t06) MOS	9	133	723	2.10	1.20	85	519	-0.58	884
CDFS (T03t06, T04t06) MOS CBT	9	206	984	1.78	1.22	67	355	-1.13	605
CDFS (T03t06, T04t06) AW32	9	143	852	1.95	1.30	148	531	-0.92	848
NSPARCS (T18t21) MOS	8	128	656	2.57	1.12	194	-	-0.80	771
NSPARCS (T18t21) AW32	8	138	792	2.38	1.14	155	-	-0.89	698
Gal. Plane (T08/T09t28) MOS	9	186	1142	2.86	1.23	46	320	-	
Gal. Plane (T08/T09t28) AW32	9	198	1319	2.69	1.24	39	312	-	
Cygnus (T21t21) MOS	9	143	767	2.55	1.03	53	367	-	
Cygnus (T21t21) AW32	9	152	900	2.38	1.05	(2.4)	366	-	
COSMOS (T11t15, T11t16) MOS	4	158	927	2.74	1.23	160	661	-	
COSMOS (T11t15, T11t16) AW32	4	169	1115	2.56	1.59	139	613	-	
COSMOS ((T11t15, T11t16) SEPLMOS	4	142	824	2.73	1.23	173	733	-0.61	344
J0555+3948 (T20t08) MOS	9	134	682	2.97	1.63	113	492	-	
GOODS-N (T26t10) MOS	5	142	733	2.70	1.53	69	404	-	
J1546+0026 MOS	1	197	1028	2.42	1.29	-	-	-	
W1303-1051 MOS	1	185	1131	2.97	1.18	-	-	-	

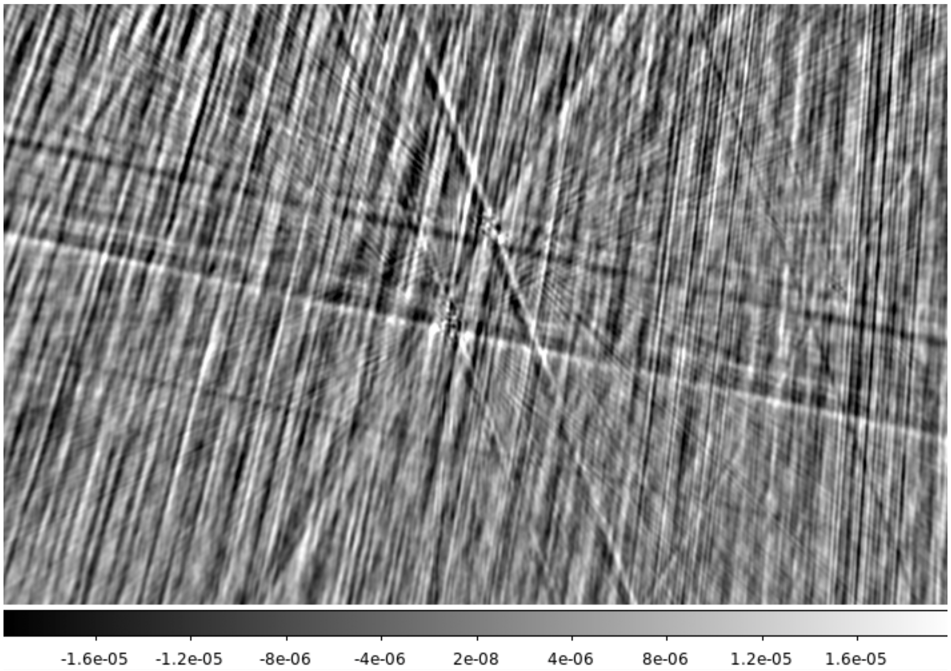

Figure 18: Difference image made with aterms-off minus aterms-on of the part of the J220600-003000 field containing the source J220513-000425 (e.g. Figure 8). The scale is linear from -20 to +20 μJy/beam.

3.1 Clipped vs non-clipped cache

Since the implementation of AW used for imaging clipped the size of the cache, we compared image J220600-003000 made with and without the clipping. This could affect the PSF and thus how well the images clean. The noise in both the unclipped and clipped tt0 images was 163 μJy, and pixel values for bright sources differ by $\lesssim 1\%$, so the clipping of the cache seems to make a negligible difference to the image quality.

3.2 A-terms off

The A-terms refer to rotated, azimuthally asymmetric, beams used during the gridding process in AW-project. tclean has an option to turn off these terms (aterms=False), which results in a simpler, azimuthally symmetric gridding function being used, and defaults to the older, azimuthally symmetric primary beams used in the mosaic gridder to make the primary beam correction. The less accurate (but computationally less expensive) aterms=False images are expected to differ from those made with aterms=True, particularly at the edges. In practice, we find that the images made in AW1 with aterms=False (AW1-ATF) are almost identical those made with A-terms enabled in the inner 1 deg^2 that is cut out to make the production VLASS images, suggesting that switching the A-terms off may be a way to reduce processing times for AW-project. The RMS of the difference image made by subtracting the aterms on from the aterms off for the J220600-003000 field is only 2.7μJy, compared to an RMS noise in tt0 of 160 μJy (Figure 18). Examination of the difference image shows that the diffraction spikes differ at the $\sim 10-20\mu$Jy level (perhaps accounting for the slightly lower dynamic range seen in Table 6), but not by enough to make a noticeable difference to the image. While this suggests that deeper data (at the level of the VLASS cumulative imaging or deeper) may benefit from A-terms, the differences in the SE between A-terms on and A-terms off are negligible.

3.3 Usepointing=False

The AW-project algorithm introduces a correction for the OTF imaging mode that fixes offsets between the correlated phase centers and the actual antenna position as recorded in the antenna pointing file (JIRA

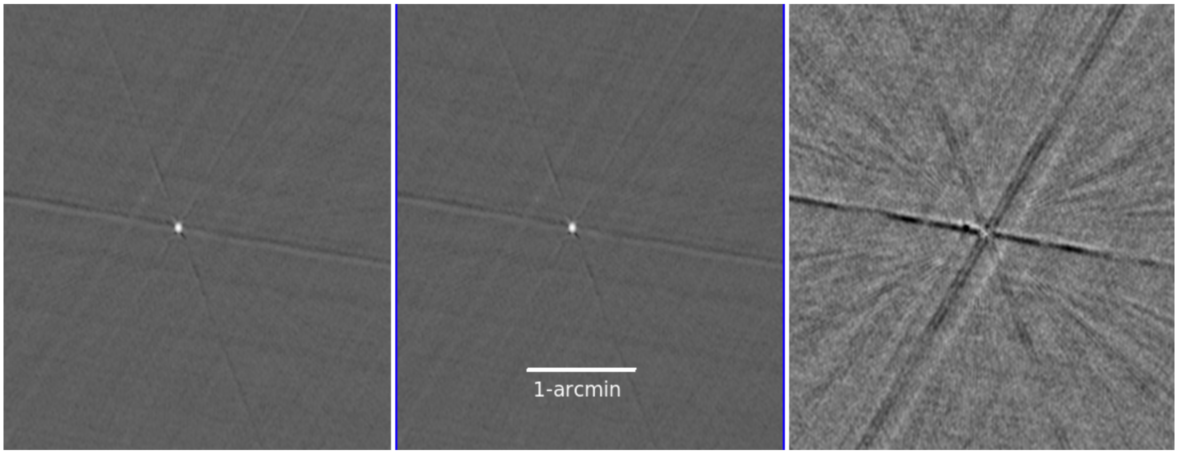

Figure 19: The region around a 100 mJy point source in J220600-003000 imaged with *left* AW1, *middle* AW1 with Usepointing=False and *right* the difference image, with a linear greyscale from -200 to 200 μJy. The biggest difference is at the position of the source, where there is a 300μJy offset.

CASA ticket CAS-12617 & corresponding CASA report; Rau 2020). This correction is not included in the MOS imaging. There are two key effects related to antenna pointing that impact VLASS data: first, there is an overall offset between the reported position of the antennas and the phase centers, second, there is a nodding in the offset because the correlator records phase center information only once per scan, and each scan has two integrations. Thus, the two integrations in each scan get assigned the mean phase center. (For VLASS1.1 data only, there is also an effect produced by incorrect positions for 2/3 of the antennas with bad ACU encoders; a correction for this can also be applied in principle.)

To some extent these systematics should average out in the MOS images. To test this, we imaged nine fields in the T10t34/CNSS overlap with the AW-project algorithm and a single w-plane with both the pointing correction on (USEPOINTING=TRUE) and off (USEPOINTING=FALSE; AWP1-PTGF). As shown in Table 6, there are no differences in the RMS or beamsize, and only very small differences in the dynamic range (4% in the near-field, no difference in the far-field) between the two sets of images. Examination of the difference image for the field J220600-003000 shows $< 1\%$ differences in surface brightnesses/flux densities and sidelobe differences of $< 200\mu$Jy (Figure 19). As in the case of the a-terms although these differences may be significant in high dynamic range situations, over most of VLASS SE the differences will be negligible, and the dynamic range limited by other factors.

4 Comparison to Quick Look catalogs

A combined component list from the SE images in the CNSS, N-SPARCS and CDFS (total 48 deg^2) was compared to Quick Look QA catalogs generated by NRAO. Results are shown in Table 7 for sources brighter than 5 mJy. As expected, the total flux densities in QL and SE MOS are very similar (as the same imaging algorithm is being used). We do see a small offset of 2–3% between the QL and AW32 total flux densities, in the sense that the AW32 flux densities are higher, which may indicate that the improvements in the AW32 imaging algorithm relative to MOS are resulting in slight improvement in the flux density measurements. This difference is well within the 10% flux density calibration requirement for VLASS though. In all cases, peak fluxes are 3–5% higher for SE than QL, probably due mostly to the self-calibration used in the SE imaging. We note that these offsets for VLASS1.1 are somewhat smaller than noted in VLASS Memo 13, where the flux densities were compared to those of bright calibrators in QL VLASS1.1 and QL VLASS1.2. Issues with the brighter calibrators may therefore have had a different underlying cause (e.g. an improvement in tclean) than the antenna pointing issue that was suspected for the poor VLASS1.1 results. Note that the large ($\approx$ 15%) flux density offset between VLASS and other surveys (in the sense that VLASS is too low) suggested by Gordon et al. (2021) on the basis of extrapolating from lower frequencies is not supported by this analysis. Even though Gordon et al. (2021) tried to restrict their comparison to unresolved, or barely

	Number	Median Total Flux ratio	Median Peak flux ratio
VLASS1.1 QL/SE MOS	716	1.00	0.96
VLASS1.1 QL/SE AW32	774	0.98	0.95
CIRADA (1.1QL)/SE MOS	735	0.99	0.96
CIRADA (1.1QL)/SE AW32	797	0.98	0.96
VLASS2.1 QL/SE MOS	966	0.99	0.97
VLASS2.1 QL/SE AW32	992	0.97	0.97

Table 7: Comparison of total and peak flux densities for QL from VLASS1.1 and 2.1 and SE (where SE is from VLASS 2.1) for sources > 5 mJy in SE total flux density over the $48\,\mathrm{deg}^2$ of the combined CNSS, N-SPARCS and CDFS fields. Uncertainties are <1% in all cases.

resolved sources, it is possible that this discrepancy is due to the fact that the resolution of VLASS is very much higher than that of the other surveys compared (which ranged from 5 arcsec for FIRST to 54 arcsec for WENSS). Thus the comparison may have been affected by source blending/confusion in the lower resolution surveys and/or resolved flux in VLASS that was not included in the pyBDSF source fits. [6]

5 Pipeline product validation

The VIP algorithm was implemented in the VLASS pipeline, and, at the time of writing, pipeline products are starting to be produced for the 100 deg^2 described above that fall in VLASS2.1 (plus other images in the same scheduling blocks). The results for the first of these images are denoted SEPLMOS in Table 6. The RMS is typically slightly lower in the pipeline images than in the scripted ones (by about 10% in CNSS), as growth of flags was turned off in the calibration pipeline for SE.

6 Summary

The comparison of the results of our validation to the relevant requirements is shown in Table 8. The mosaic gridder used with the VIP script makes tt0 images over the whole sky that are very likely to satisfy most of the survey requirements after image-plane position corrections are applied and the amount of target flagging in the pipeline reduced. Models suggest some chromatic aberration in the tt0 images at the few percent level may be occurring in the far south. These far southern tiles will most likely be processed with AW-project using the GPU gridder currently under development. Requirements that were not satisfied are noted below:

- Image fidelity of extended sources was not found to be limited by the thermal noise, as was assumed when the requirements were written. Errors in bright objects are dominated by a contribution that scales with the local source surface brightness averaged over $\approx 1-2$ beamwidths due to residual amplitude, phase and pointing calibration errors, probably exacerbated by cleaning.

- For compact source spectral indices, tt1 images made with the mosaic gridder, CONJBEAMS=FALSE (CBF), suffer from systematic errors over the whole sky due to the lack of w-term corrections, but at zenith distances $\lesssim 45^\circ$ in-band spectral indices using these are within the survey requirements of ± 0.2 for compact sources.

- None of the imaging implementations tested is able to obtain spectral indices within the survey requirements (± 0.2 in high SNR regions) over the entire emission from extended sources that have strong brightness gradients.

- The rate of $> 5\,\sigma$ artifacts is $\approx 2\ \mathrm{deg}^{-2}$, mostly from PSF spikes around bright objects arising from residual calibration errors. It is likely that many of these can be flagged using a peak-to-ring criterion or similar, reducing the rate to the requirement of 1 deg^{-2},

---

[6] Fortunately this correction was not applied to their cataloged fluxes, which agree well with the NRAO internal catalogs.

Table 8: Tabulation of validation results versus requirements for SE continuum MOS images.

Requirement	Goal/Requirement	Achieved	How tested
SCI-OBS-005: Beamsize	$< 3''$, AR < 2 60% of sky	$< 3''$, AR < 2 (100 deg^2)	image headers[‡]
SCI-BDP-015: RMS (SE)	$< 170\mu$Jy, 90% of sky	$< 170\mu$Jy, >65% of sky[†]	direct measurement
SCI-BDP-020: flux density calibration	5%/10%	3%	Calibrator observations
S1: positional accuracy	$0.1''/0.25''$	$0.2''$	Comparison to Gaia
S2: Pt. Src. spectral index	±0.1/±0.2	±0.2	Calibrator observations
S3a: NF dynamic range	50/100	≈120	Measured on images
S3b: FF dynamic range	400/2500	≈450	Measured on images
S9: ext. src. image fidelity	thermal lt/themal lt+40%	±10% of surf. brightness+themal[*]	Comparison to ptd
S10: ext. src. Sp. Index fidelity	within thermal noise	±0.2-0.5[**]	Comparison to ptd
S11: 5σartifacts	<0.03 deg^{-2}/< 1 deg^{-2}	$\sim$2 deg^{-2}[§]	comparison to COSMOS/CNSS

[†] flagging will be changed in the production pipeline and likely bring the products within the requirement.

[‡] also verified for QL over whole survey.

[*] probably limited by phase and amplitude calibration errors.

[**] Large variations are seen around regions of high brightness gradient, probably due to the intrinsic limitations of in-band spectral indices and the changing shape of the effective beam with frequency. Also, in the south of the survey, chromatic aberration effects from the neglect of *w*-terms leads to large systematics. However, for well-resolved, smooth and bright regions the spectral indices are in general reliable to within ±0.2.

[§] flagging of PSF artifacts around bright objects will bring this down to closer to the requirement of ≈ 1 deg^{-2}.

Images made with AW-project in general do not clean as well as images made with the mosaic gridder. They also have slightly higher noise (7%) and smaller beams (7%). Especially for complex sources in the galactic plane, the cleaning is borderline for acceptance, with lower dynamic range than MOS. This might be mitigated by removing the mask at the final clean stage (as was done for the mosaic images), but possibly at the price of lower speed to convergence. Spectral indices, although noisy compared to MOS, show no *w*-term related systematics, though still have issues when dealing with complex extended emission.

6.1 Recommendations

On the basis of this analysis, it was recommended that the MOS images should be used for most of the SE continuum (at least the ≈ 50% of the sky observed at Zenith Distance < 45 deg.) for speed and practicality, with the caveat that the spectral indices at high Zenith Distances will be inaccurate, and the VLASS image cubes will be needed to form accurate spectral indices for compact sources, especially in the South. Spectral indices in catalogs should be from the average of a 5x5 box centered on the peak pixel for images made with the mosaic gridder.

6.2 Quality assurance process

The quality assurance (QA) process for SE will include a tile-level examination of the data, besides the image-level already implemented for Quick Look (which currently examines the maximum residual, primary beam "black holes" (where the primary beam response falls below the cutoff due to flagging) and other major artifacts, and will be extended to include beam shape and size). Specifically:

- A script will be run to check the near-field dynamic range with the aim of identifying tiles with poor self-calibration.
- A component list will be made per tile by combining the image catalogs produced by pyBDSF in the pipeline. Data analysts (DAs) will ensure component counts are in the typical range (50-100 deg^{-2}) and measure the median spectral index and make sure it is -0.8 ± 0.2 in regions where that degree of accuracy is expected (north of Decl. $\approx -10°$).
- DAs will measure the median RMS for the tile and ensure it is below the survey requirement of 170 μJy.

Tiles that fail these checks may fail for legitimate reasons (e.g. fields close to the Galactic Plane), but if not, we will look into how they could be improved by reprocessing (for example, better cleaning).

References

Bhatnagar, S., Rau, U., & Golap, K. 2013, ApJ, 770, 91

Condon, J. 2015, arXiv e-prints, arXiv:1502.05616

Condon, J. J. 1997, PASP, 109, 166

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693

Eddington, A. S. 1913, MNRAS, 73, 359

Gordon, Y. A., Boyce, M. M., O'Dea, C. P., et al. 2021, ApJS, 255, 30

Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, A&A, 598, A78

McConnell, D., Hale, C. L., Lenc, E., et al. 2020, PASA, 37, e048

Mohan, N., & Rafferty, D. 2015, PyBDSF: Python Blob Detection and Source Finder, , , ascl:1502.007

Mooley, K. P., Myers, S. T., Frail, D. A., et al. 2019, ApJ, 870, 25

Mooley, K. P., Hallinan, G., Bourke, S., et al. 2016, ApJ, 818, 105

Nandakumar, G., Veena, V. S., Vig, S., et al. 2016, AJ, 152, 146

Perley, R. A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 275

Rau, U., Bhatnagar, S., & Owen, F. N. 2016, AJ, 152, 124

Smolčić, V., Novak, M., Bondi, M., et al. 2017, A&A, 602, A1

A The effects of complex gain calibration errors on positions and brightnesses.

In this Appendix we apply the dynamic range (DR) analysis of Perley (1999) to the case of brightness and position uncertainties in VLASS. Following Perley (1999), consider a point source with visibilities $V(u)$:

$$V(u) = \delta(u - u_k) \tag{6}$$

for all k baselines u except for u_0, which contains a phase error, ϕ:

$$V(u_0) = \delta(u - u_0)\,\mathrm{e}^{-i\phi}. \tag{7}$$

The image is the Fourier Transform of the visibilities:

$$I(l) = \int V(u)\,\mathrm{e}^{2\pi i u l}\mathrm{d}u \tag{8}$$

Thus each of the $1 - k$ good baselines contributes $2\cos(2\pi u_k l)$ (from u_k and its complex conjugate u_k^*, and the bad baseline contributes $2\cos(2\pi u_0 l - \phi) \approx 2\left[\cos(2\pi u_k l) + \phi \sin(2\pi u_0 l)\right]$. So

$$I(l) = 2\,\phi\sin(2\pi u_0 l) + 2\sum_{k=1}^{N(N-1)/2}\cos(2\pi u_k l) \tag{9}$$

and the PSF:

$$B(l) = 2\sum_{k=1}^{N(N-1)/2}\cos(2\pi u_k l). \tag{10}$$

Subtracting the PSF via clean then leaves the residual R:

$$R(l) = 2\,\phi\sin(2\pi u_0 l) \tag{11}$$

This is an odd function about the source center, so phase errors show up as asymmetric distortions in point sources in the image. A similar analysis for amplitude errors, where $V(u_0) = (1+\epsilon)\delta(u-u_0)$ leads to $R(l) = 2\,\epsilon\cos(2\pi u_0 l)$, an even function about the source center.

We adapt this simple analysis to recover the result that phase errors can also result in position shifts. By approximating $R(l)$ as a straight line for small values of l ($2\pi u_0 l << 1$), we obtain $R(l) \approx 4\,\phi\,\pi u_0 l$, and can calculate the shift of a Gaussian source $I(l) = \mathrm{e}^{-l^2/2\sigma^2}$:

$$< l >= \frac{\int_{-L}^{L} l\,\mathrm{e}^{-l^2/2\sigma^2} + 4\,\phi\,\pi u_0 l^2 \mathrm{d}l}{\int_{-L}^{L} \mathrm{e}^{-l^2/2\sigma^2} + 4\,\phi\,\pi u_0 l\,\mathrm{d}l} \tag{12}$$

where the integral is carried out over a range close to the PSF center $(-L, L)$. The first term in the numerator and second term in the denominator average to zero, and, approximating the Gaussian as a top hat for small L and neglecting constants of order unity,

$$< l >\sim \phi u_0 L^2 \tag{13}$$

The fractional shift is then proportional to ϕ and depends on the scale of the baseline affected, with errors on longer baselines having larger effects.

When more than one baseline phase error is combined, the sine functions will add together as a random walk, but differently depending on the physical origin of the phase errors. To show the effects of this we construct a toy 1D model of the VLA with 27 approximately logarithmically-distributed antennas between zero and 5 km. We simulate an observation with five frequencies from 2–4 GHz, and apply phase errors three different ways:

1. Baseline-based phase errors. The phase errors on each baseline, $\delta\phi_k$, are independent and randomly drawn from a Gaussian with width $\sigma_{\rm ph}$.

2. Antenna-based phase errors. Each antenna has an associated phase error which is independent and randomly drawn from a Gaussian with width $\sigma_{\rm ph}$. The phase error on the baseline between the ith and jth antenna is $\delta\phi_k = \delta\phi_i - \delta\phi_j$.

3. A phase wedge. The phase error modeled as a phase screen whose phase change is proportional to distance along the 1D array, $\delta\phi = \epsilon D_i$, where ϵ is the phase change per meter on the ground and D_i is the distance of the ith antenna from the origin. Then $\delta\phi_k = \epsilon(D_i - D_j)$.

Case (1), purely baseline-based errors, lead to results that are not in general consistent with the observations. Observations show that the artifacts from bright sources decay as a function of distance from the bright source, whereas, in the baseline-based case, the lack of correlation between the phase errors results in no significant increase in DR away from the source. We can thus probably eliminate this model from consideration. (It is worth noting though that baseline-dependent errors may account for the spike artifacts radiating from bright sources, which only decay slowly with distance in the VLASS images and are not completely mitigated by antenna-based self calibration.) Cases (2) and (3) (which both involve a form of antenna-based phase error) seem to result in models that better reproduce the data overall, with low DR near the source that increases further out, so we therefore focus on those in what follows.

We compare the near-field DR (described in Section 2.6) from these models to the observed values to infer the likely range of values for $\sigma_{\rm ph}$ and ϵ. For the QL data for which no self-calibration is applied, we observe DR in the range 30-100. In Case (2) (random antenna-based errors), this is reproduced quite well with $\sigma_{\rm ph} \approx 15\,$deg. Corresponding position shifts are $\approx 0.1\,$arcsec, in fairly good agreement with the observations (Figure 20). In Case (3) (phase wedge), these values of DR correspond to a phase variation across the 5 km array of $15 - 40\,$deg, and position shifts 0.12-0.17 arcsec (again, in fairly good agreement with the data). In the SE case, the self-calibration complicates the analysis as we expect the position errors to be "frozen in" by the self-calibration, but we can still observe the improvement in the near-field dynamic range, which is observed to be about a factor of two. This corresponds to a decrease in $\sigma_{\rm ph}$ to $\approx 5\,$deg (Case 2), or to a residual phase variation over 5 km to $\approx 10\,$deg (Case 3).

The implications for the surface brightness differences noted in Section 2.7 are harder to assess, though with this model we can examine the amplitude of the residual errors much closer to the centre of the beam than can be easily measured from the images. In Case (2) for SE, with $\sigma_{ph} \approx 5\,$deg, we see maximum

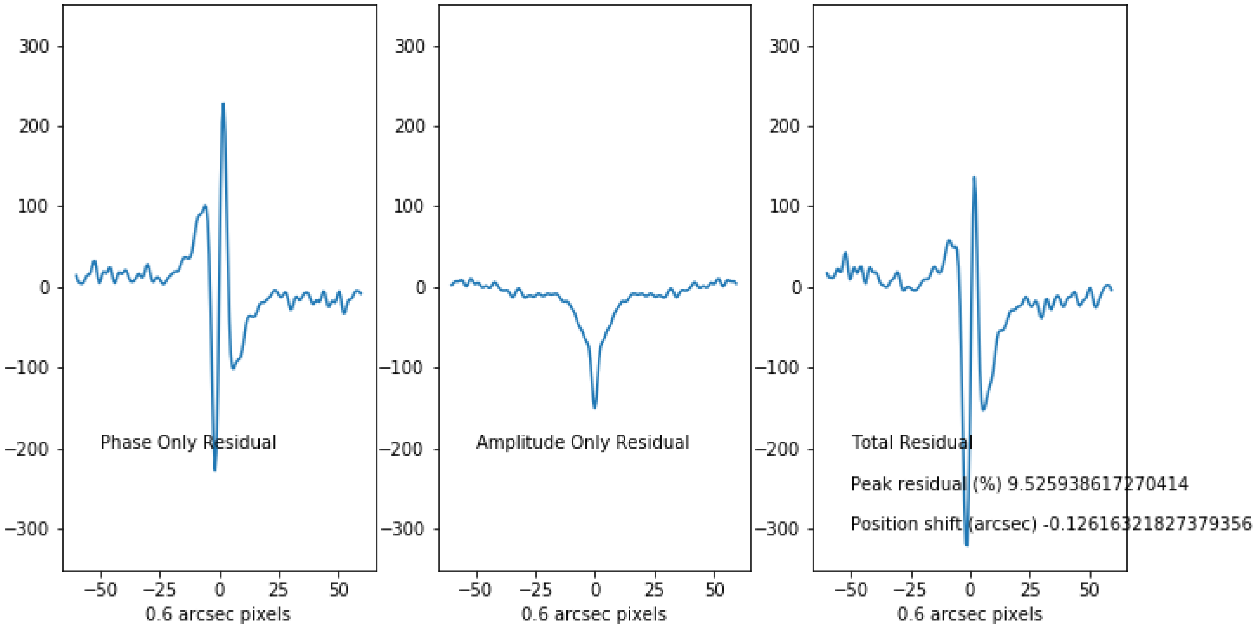

Figure 20: Example 1D simulation realization for a point source in QL. This observation has a near-field dynamic range of 60, and was made with antenna-based random phase fluctuations with $\sigma_{\rm ph} = 15\,$deg and an RMS amplitude variation per antenna of 0.1. The vertical axis is scaled to ±10% of the PSF peak.

residuals of only a few percent. Case (3) is similar. So although residual phase variations can contribute to the surface brightness differences, they may not be the dominant factor given that f_g in Equation 4 is ≈ 0.1. Other factors, such as clean instability in the presence of these calibration errors, and spurious features from QL phase errors "frozen in" by self calibration may also contribute. (It is also important to note that a full 2D simulation may give a different quantitative result.)

Residual amplitude calibration errors (possibly from pointing errors) may also play a role, especially given that the phase errors are significantly reduced by self-calibration. As these will always be even functions about the source position, they will not lead to further position shifts, and they decay as a function of distance from the source position more rapidly than the phase error residuals (Figure 20). We can constrain the residual amplitude errors as the flux densities of the bright calibrators are within $\approx 2\%$ (Table 2), suggesting a mean amplitude error $\approx 2 \times \sqrt{(}27) = 0.1$. Assuming this level of amplitude error does add significantly to the residual.