MIGHTEE: total intensity radio continuum imaging and the COSMOS / XMM-LSS Early Science fields

I. Heywood[1,2,3]⋆, M. J. Jarvis[1,4], C. L. Hale[5], I. H. Whittam[1,4], H. L. Bester[3,2], B. Hugo[3,2],
J. S. Kenyon[2,3], M. Prescott[4,6], O. M. Smirnov[2,3], C. Tasse[7,2], J. M. Afonso[8,9], P. N. Best[10],
J. D. Collier[11,12,13], R. P. Deane[14,15,2], B. S. Frank[3,11,16], M. J. Hardcastle[17], K. Knowles[2,3],
N. Maddox[18], E. J. Murphy[19], I. Prandoni[20], S. M. Randriamampandry[21,22,23], M. G. Santos[4,3],
S. Sekhar[11,24,4], F. Tabatabaei[25,26], A. R. Taylor[11,16,4], K. Thorat[15]

Author affiliations are listed at the end of the paper.

ABSTRACT

MIGHTEE is a galaxy evolution survey using simultaneous radio continuum, spectro-polarimetry, and spectral line observations from the South African MeerKAT telescope. When complete, the survey will image ∼20 deg^2 over the COSMOS, E-CDFS, ELAIS-S1, and XMM-LSS extragalactic deep fields with a central frequency of 1284 MHz. These were selected based on the extensive multiwavelength datasets from numerous existing and forthcoming observational campaigns. Here we describe and validate the data processing strategy for the total intensity continuum aspect of MIGHTEE, using a single deep pointing in COSMOS (1.6 deg^2) and a three-pointing mosaic in XMM-LSS (3.5 deg^2). The processing includes the correction of direction-dependent effects, and results in thermal noise levels below 2 μJy beam^{-1} in both fields, limited in the central regions by classical confusion at ∼8″ angular resolution, and meeting the survey specifications. We also produce images at ∼5″ resolution that are ∼3 times shallower. The resulting image products form the basis of the Early Science continuum data release for MIGHTEE. From these images we extract catalogues containing 9,896 and 20,274 radio components in COSMOS and XMM-LSS respectively. We also process a close-packed mosaic of 14 additional pointings in COSMOS and use these in conjunction with the Early Science pointing to investigate methods for primary beam correction of broadband radio images, an analysis that is of relevance to all full-band MeerKAT continuum observations, and wide field interferometric imaging in general. A public release of the MIGHTEE Early Science continuum data products accompanies this article.

Key words: radio continuum: galaxies – techniques: interferometric – surveys

1 INTRODUCTION

Radio continuum observations are a uniquely powerful tool in the pursuit of understanding how galaxies form and evolve over cosmic time, one of the key goals of modern astrophysics. Radio emission arises from the cores of active galactic nuclei (AGN; White et al. 2015; Whittam et al. 2016), and their powerful jet-driven radio lobes (Laing et al. 2011; Fanaroff et al. 2021); star-formation in 'regular' galaxies can be detected via generally-fainter synchtrotron emission (Condon 1992; Jarvis et al. 2010; Delvecchio et al. 2021) as well as from thermal emission at higher radio frequencies (Murphy et al. 2017); and, on Mpc scales, radio observations reveal diffuse radio haloes that trace the hot gas within galaxy clusters, as well as the shock-driven relic structures that can be found on their peripheries (van Weeren et al. 2019). Advances in instrumentation and data processing methods over the last decade have led to a new generation of large-scale radio surveys that are targeting the emission processes described above to efficiently gather galaxy detections in statistically significant numbers, and in a range of environments out to the highest redshifts.

Telescope time is a finite commodity, and the most efficient way to conduct extragalactic surveys is to adopt a tiered approach whereby areal coverage is traded for observational depth. Examples of on-going surveys that are essentially covering the entire sky visible to their respective observatories are the Australian Square Kilometre Array's (ASKAP; Hotan et al. 2021) Rapid ASKAP Continuum Survey (RACS; McConnell et al. 2020; Hale et al. 2021) and forthcoming Evolutionary Map of the Universe (EMU; Norris 2015) survey, the Very Large Array Sky Sur-

vey (VLASS; Lacy et al. 2020), and at lower radio frequencies the Galactic and Extragalactic All-sky Murchison Widefield Array survey (GLEAM; Hurley-Walker et al. 2017) and the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al. 2017) being conducted on the Murchison Widefield Array (MWA; Lonsdale et al. 2009) and Low Frequency Array (LOFAR; van Haarlem et al. 2013) respectively. Complementary to these are intermediate tier surveys that have higher sensitivities over areas of ∼100s–1000s of square degrees, such as the VLA Stripe 82 surveys (Heywood et al. 2016; Mooley et al. 2016), the 325 MHz observations of the Galaxy and Mass Assembly (GAMA) fields with the Giant Metrewave Radio Telescope (Mauch et al. 2013), and the imaging surveys with the Westerbork Synthesis Radio Telescope's Aperture Tile In Focus (APERTIF) upgrade (van Cappellen et al. 2021). Finally, over smaller areas, typically 1–10 deg^2, we find the deepest radio surveys reaching μJy sensitivities, well into the regime where regular starforming galaxies are dominating the radio source counts (e.g. Prandoni et al. 2018; Matthews et al. 2021), for example the deep VLA observations of the COSMOS field (Smolčić et al. 2017; van der Vlugt et al. 2021), the VLA Hubble Frontier Fields survey (Heywood et al. 2021), the e-MERLIN e-MERGE project (Muxlow et al. 2020), the LoTSS deep fields (Tasse et al. 2020), and the 'DEEP2' observations (Mauch et al. 2020) with the South African MeerKAT telescope (Jonas & MeerKAT Team 2016).

MeerKAT consists of 64 × 13.5 m dishes with offset Gregorian optics, providing an unblocked aperture. It is equipped with three receiver bands; UHF (544 – 1088 MHz), L-band (856 – 1712 MHz), and S-band (1750 – 3500 MHz). Three-quarters of the collecting area is within a dense, 1 km diameter core region, and the remaining dishes are situated around the core, providing a maximum baseline of 8 km. The large number of baselines, wide field of view (1 deg at L-band), and low (∼20 K) system temperature all conspire to make MeerKAT an exceptionally fast and capable synthesis imaging telescope. The correlator can also deliver up to 32,768 frequency channels, delivering excellent spectroscopic imaging capabilities.

Also coming in at the deep end is MIGHTEE (MeerKAT International Gigahertz Tiered Extragalactic Explorations; Jarvis et al. 2016). MIGHTEE is one of MeerKAT's flagship Large Survey Projects, a galaxy evolution survey that is using simultaneous continuum, polarimetry (Sekhar et al., *in prep.*) and spectral line (Maddox et al. 2021) measurements to investigate the formation and evolution of galaxies over cosmic time. It will use ∼1000 h of observations with MeerKAT's L-band receivers, with the goal of imaging 20 square degrees over four extragalactic deep fields, namely COSMOS, the Extended Chandra Deep Field South (E-CDFS), the southermost field of the European Large Area ISO Survey (ELAIS-S1), and the XMM-*Newton* Large Scale Structure field (XMM-LSS). The deep multiwavelength data in these fields is essential for obtaining redshifts for the radio sources, and disentangling the relative contributions of star formation and black-hole accretion to the total radio emission (White et al. 2017). The design of the MIGHTEE survey is such that the total intensity continuum images reach the classical confusion limit of MeerKAT with RMS fluctuations of approximately 2μJy beam^{-1}. The field selection and the large investment of telescope time from the observatory

provide MIGHTEE with an unrivalled combination of depth, area, and corresponding multiwavelength data, pushing both the survey parameters and the resulting science beyond the state of the art.

This article presents a description (Section 2) and validation (Section 4) of the continuum data processing strategy for the MIGHTEE survey. The wide field of view and extreme instantaneous sensitivity of MeerKAT means that direction dependent calibration methods are necessary to reach the requirements of the survey's design. We use this processing strategy to provide the 'Early Science' continuum data products for the MIGHTEE survey (Section 3), namely a single pointing in the COSMOS field, and a three-pointing mosaic in the XMM-LSS field, from which we extract component catalogues (Section 3.3).

In designing and validating our data processing strategy, we conducted an investigation of some different primary beam correction methods for broadband radio continuum imaging, and we include the results of this investigation here, as it is more broadly relevant for widefield, full-band imaging with MeerKAT, and modern radio interferometers in general. We provide a public data release[1] of the catalogue and image products with this article. In the interests of reproducibility, and to provide even more detail on our processing methods for the truly curious, the scripts that were written to process the MIGHTEE data are made available online[2] (Heywood 2020c). These have already proven to be suitable for general, semi-automatic processing of MeerKAT continuum observations, and continue to be developed.

2 OBSERVATIONS AND DATA PROCESSING

In addition to providing an initial release of continuum data to the community, the Early Science phase of MIGHTEE has also been used to develop and refine the data processing methods for the full-scale survey. This section provides a summary of the Early Science observations, and a detailed description of the full-band, Stokes I continuum imaging strategy. Numerous software packages are used, as cited in the sections that follow.

2.1 MeerKAT observations

The properties of the MeerKAT observations that make up the MIGHTEE continuum Early Science products are listed in Table 1. In total there are 25 h of observations in COSMOS for an on-source time of 17.45 h. For the three pointings in XMM-LSS there are 16.05, 16.12 and 16.03 h for XMM-LSS_12, XMM-LSS_13 and XMM-LSS_14, with 12.41, 12.45 and 12.43 h of on-source time respectively. The observational setup depends slightly on the vintage of the observing block. Typically, primary calibrators were visited for 5-10 minutes, at least twice per block, and secondary calibrators were visited for 2-3 minutes following every 20-30 minute target scan. As part of the verification of the Early Science data (Section 4) we also image and make use of all

---

[1] https://doi.org/10.48479/emmd-kf31
[2] v0.1; https://github.com/IanHeywood/oxkat

of the pointings available in COSMOS at the time of writing, although these image products are not part of the Early Science data release. These additional observations are summarised in Table B1. For all observations used in this paper the correlator integration time was 8 s per visibility point.

2.2 Flagging and reference calibration (1GC)

Each individual MeerKAT observation, as itemised in Tables 1 and B1, is a self-contained block containing scans of the target as well as scans of appropriate primary and secondary calibrator sources[3]. These calibrator sources are used to derive corrections for instrumental and propagation effects that are then applied to the target data, a process known as first-generation calibration, or 1GC.

For each observation we used the KAT Data Access Library[4] to convert the visibility data into Measurement Set format (Kemball & Wieringa 2000). Flags generated by the telescope's control and monitoring system were applied, and frequency averaging was also performed at this point to reduce the number of channels (natively either 4096 or 32768) to 1024. Basic flagging commands were applied to all fields using CASA (McMullin et al. 2007). The bandpass edges and the Galactic neutral hydrogen line were flagged for all baselines. Frequency ranges containing persistent radio frequency interference (RFI) were flagged on spacings shorter than 600 m. The auto-flagging algorithms TFCROP and RFLAG were used on the calibrator fields with their default settings.

The primary calibrator was used to derive delay and bandpass solutions (both per-scan), and frequency-independent complex gain corrections amplitude and phase terms (per integration time) using the GAINCAL and BANDPASS tasks in CASA. We used the Reynolds (1994) model (see also Heywood et al. 2020b) to predict the spectral behaviour of the standard calibrator PKS B1934−638. In the case of the primary calibrator being PKS 0408−65, we assume a point source model with a flux density of 17.066 Jy beam^{-1}, and a spectral index[5] of −1.179, the reference frequency being 1284 MHz. The calibration is iterative, with rounds of auto-flagging on residual (corrected − model) visibilities taking place before the next iteration.

The gain solutions derived from the primary were applied to the secondary calibrator. We derived complex gain corrections (one solution per scan) from the scans of the secondary calibrator in eight spectral bins. These gains were scaled according to the gain amplitudes derived from the primary calibrator, using the CASA FLUXSCALE task, allowing us to determine a polynomial model of the intrinsic spectral shape of the secondary calibrator. We then derived per-scan complex gain corrections from the observations of the secondary using this intrinsic model with the GAINCAL task. This process compensates for the effects of the large fractional bandwidth of MeerKAT. If the secondary calibrator deviates significantly from being spectrally flat, then the flux scale may be biased if this is not taken into account.

The final reference calibration step was to apply the gain solutions to the target data, which were then split out into a separate Measurement Set and flagged using the TRICOLOUR[6] package. Following the removal of the low-gain bandpass edges there is approximately 800 MHz of usable bandwidth. A further loss of about 50 per cent of the data within this region is typical following auto-flagging. The RFI occupancy is strongly dependent on baseline length, and most of this loss occurs on spacings shorter than 1 km in the core of MeerKAT (see also Mauch et al. 2020).

2.3 Direction-independent self-calibration (2GC)

The use of the target data themselves to further refine the antenna-based gain corrections is known as self-calibration, or second-generation calibration (2GC). Multi-frequency synthesis (MFS) images of the target data were made using WSCLEAN (Offringa et al. 2014). The full-band data were imaged without a cleaning mask, with deconvolution terminating after 100,000 clean components or when the peak residual reaches 20 μJy beam^{-1}, whichever occurs first. A clean mask was derived from the resulting image, excluding regions below a local threshold of 6σ, where σ is an estimate of the local pixel standard deviation. We estimate σ as a function of position following the method of Tasse et al. (2020). Consider a set of n pixel brightness measurements X, drawn from an image with pixels that follow a normal distribution with a mean of 0 and a standard deviation of 1. The cumulative distribution of $Y = \min\{X\}$ is

$$F = 1 - \left[\frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)\right)\right]^n \qquad (1)$$

where erf is the Gaussian error function. This allows a factor to be derived that converts a measurement of $\min\{X\}$ to σ for a given value of n, by finding the point where $F(y_\sigma) = \frac{1}{2}$. The number of measurements n is governed by the sliding box size within which $\min\{X\}$ is evaluated for each pixel in the image. The MIGHTEE fields are dominated by compact sources, and we find that a value $\sqrt{n} = 50$ gives high completeness, returning large numbers of true sources across the full field whilst excluding positive artefacts surrounding brighter sources. Minimum statistics are a good estimator of local noise since thermal noise should be symmetric about zero, and the corrupted PSF that dominates regions blighted by calibration artefacts also has correspondingly lower negatives. The filter is, however, not sensitive to true source confusion which manifests itself as a positive tail on a histogram of pixels drawn from a total intensity radio image. Following the creation of the deconvolution mask the original blind-clean image is discarded and the data are re-imaged using this mask.

All images extend into the sidelobes of the primary beam in order to deconvolve and model the many sources which are readily detected by MeerKAT in those regions. The total image extent is 10240 × 10240 pixels, each pixel spanning 1″.1 for a total image extent of 3.13 × 3.13 deg^2.

The multi-frequency clean components from the masked image were used to predict a visibility model in eight spectral bins, and the data were self-calibrated using the

---

[3] Polarisation calibrators are also included, however we do not make use of them for the total intensity continuum processing.
[4] https://github.com/ska-sa/katdal
[5] Note that throughout this paper we adopt the convention that the flux density S is related to frequency ν via the spectral index parameter α: $S \propto \nu^\alpha$.

[6] https://github.com/ska-sa/tricolour

Table 1. Properties of the MeerKAT observations of the COSMOS and XMM-LSS fields that form the MIGHTEE Early Science data.

Date (UT, J2000)	Block ID	Field	RA	Dec	Track (h)	On-source (h)	N_{chan}	N_{ant}	Primary calibrator	Secondary calibrator
2018-04-19	1524147354	COSMOS	$10^h00^m28\overset{s}{.}6$	$+02°12'21''$	8.65	6.1	4096	64	J0408-6545	3C237
2018-05-06	1525613583	COSMOS	$10^h00^m28\overset{s}{.}6$	$+02°12'21''$	8.39	5.1	4096	62	J0408-6545	3C237
2020-04-26	1587911796	COSMOS	$10^h00^m28\overset{s}{.}6$	$+02°12'21''$	7.98	6.25	32768	59	J0408-6545	3C237
2018-10-06	1538856059	XMMLSS_12	$02^h17^m51^s$	$-04°49'59''$	8.02	6.22	4096	59	J1939-6342	J0201-1132
2018-10-07	1538942495	XMMLSS_13	$02^h20^m42^s$	$-04°49'59''$	8.07	6.22	4096	59	J1939-6342	J0201-1132
2018-10-08	1539028868	XMMLSS_14	$02^h23^m22^s$	$-04°49'59''$	8.03	6.19	4096	60	J1939-6342	J0201-1132
2018-10-11	1539286252	XMMLSS_12	$02^h17^m51^s$	$-04°49'59''$	8.05	6.23	4096	63	J1939-6342	J0201-1132
2018-10-12	1539372679	XMMLSS_13	$02^h20^m42^s$	$-04°49'59''$	8.03	5.92	4096	62	J1939-6342	J0201-1132
2018-10-13	1539460932	XMMLSS_14	$02^h23^m22^s$	$-04°49'59''$	8.0	6.24	4096	62	J1939-6342	J0201-1132

CASA GAINCAL task.[7] Frequency-independent phase corrections were derived for every 64 seconds of data, and an amplitude and phase correction was derived for every target scan, with the solutions for the former applied while solving for the latter. The self-calibrated data were then re-imaged using WSCLEAN, and the cleaning mask was refined based on the self-calibrated image, and a lower local noise threshold of 5.5σ.

2.4 Direction-dependent self-calibration (3GC)

The dynamic range of the MIGHTEE images is limited by the presence of direction dependent effects (DDEs). At L-band these are principally caused by the time[8], frequency, and direction dependent variations in the antenna primary beam pattern, coupled with pointing errors. DDEs manifest themselves in the radio images as error patterns resembling a corrupted point spread function (PSF) around off-axis sources of moderate brightness (e.g. stronger than a few tens of mJy).

Images from all modern radio telescopes with high sensitivities and broad bandwidths tend to be DDE-limited at some level, and as such many techniques have emerged for dealing with these effects, the so-called third-generation [of] calibration, or 3GC. Visibility-domain treatments include the use of differential gains (Smirnov 2011b,c), which is a multi-directional successor to single-source peeling. Approaches that are based on image plane operations include A-Projection, which corrects for DDEs using convolution kernels applied during gridding of the visibilities (Bhatnagar, Rau, & Golap 2013), and faceting schemes that apply corrections assuming that DDEs are piecewise constant on a per-facet basis when producing a wide-field image (Tasse et al. 2018). For the MIGHTEE continuum processing we primarily adopt the latter image plane facet based approach (Section 2.4.2), with visibility domain treatments for certain pointings that require it (Section 2.4.1).

---

[7] Note that for full-survey MIGHTEE continuum data this step has been replaced by a phase and delay self-calibration operation using CUBICAL (Kenyon et al. 2018), which has been demonstrated to yield improved results (as implemented in OXKAT v0.2). The latest version is recommended always.

[8] Although the MeerKAT primary beam pattern has relatively low sidelobes, the main lobe is not circularly symmetric. The telescope mount causes this pattern to rotate on the sky as an observation progresses. The main lobe variation in the beam holography corresponds to a ~5% amplitude variation at the nominal half-power point over the course of a full track.

2.4.1 Peeling a problem source

Pointings that have a single, dominant, problematic source (typically an off-axis source with an apparent flux density upwards of ~100 mJy beam^{-1}) can be subjected to an additional processing step. This involves modelling the source and its associated DDE and subtracting it from the visibility database entirely, a process more generally known as peeling. This is achieved by imaging the data with WSCLEAN, with masked deconvolution over a high number of sub-bands (32 by default), and with lower Briggs (1995) robust weighting (-0.6 by default) in order to obtain a model with high angular and spectral resolution. Model visibilities are then predicted into two separate columns in the Measurement Set, one of which contains only the problem source, and one containing the rest of the sky model. The CUBICAL package (Kenyon et al. 2018) is then used to solve for a direction-independent gain term ($\boldsymbol{G}$, with a default time / frequency interval of 2.4 min / 256 channels) using the complete sky model, whilst simultaneously solving for an additional complex gain term ($\boldsymbol{dE}$, with a default time / frequency interval of 9.6 min / 64 channels) against a model that contains only the problem source. The model of the problem source is then subtracted from the visibilities with $\boldsymbol{dE}$ applied, and the residual visibilities containing the rest of the sky are corrected with $\boldsymbol{G}$, ready for subsequent imaging.

An illustrative Measurement Equation (Smirnov 2011a) that describes the problem of direction-dependent calibration can be written as

$$V_{pq} = G_p \left(\sum_s dE_{sp} X_{spq} dE_{sq}^* \right) G_q^* \qquad (2)$$

where V is the visibility matrix, p and q represent antenna indices, and X is the coherency matrix for source (or direction) s. When subtracting a single dominant source $s = \{0, 1\}$, and the $\boldsymbol{dE}$ terms are fixed to unity for $s = 0$. Direction 0 is represented in this case by the full clean component model covering the entire field of view down to the cleaning threshold, minus the dominant source.

The legacy approach to peeling involves phase-rotation of the visibilities to the direction of the problem source, self-calibrating on that source using a direction-indepenent solver such as the CASA GAINCAL task, subtracting the model perturbed by the best-fitting gain solutions, and then undoing the phase-rotation and gain corrections to form the residual visibilities.

The simultaneous solving of $\boldsymbol{G}$ and $\boldsymbol{dE}$ offered by CUBICAL (and first introduced by MEQTREES; Noordam & Smirnov

2010) is facilitated by the implementation of an explicit Measurement Equation such as the one above. This has two main advantages over the legacy approach, namely: (i) the inclusion of global G solution means that the residual visibilities are less prone to biases such as ghost sources (e.g. Grobler et al. 2014) and flux suppression (e.g Mouri Sardarabadi & Koopmans 2019); (ii) the process is fully general and can subtract multiple problematic sources simultaneously rather than iteratively.

For the Early Science data the peeling step was applied to only two observing blocks, which were those from the XMMLSS_13 pointing, in order to suppress the effects of the strong compact double source at J2000 $02^{h}21^{m}43^{s}.14$ $-04°13'46''.6$s. However for the full survey this process will be routine for many of the pointings in E-CDFS and ELAIS-S1, both of which are known to contain a dominant confusing source.

2.4.2 Facet-based DDE corrections

Direction-dependent corrections were made by imaging the visibilities, either those produced following the 2GC stage (Section 2.3), or those produced by the peeling step (Section 2.4.1), with DDFACET (Tasse et al. 2018). Multiple Measurement Sets containing data from a common pointing centre are brought together at this stage.

The refined cleaning mask produced following self-calibration was used. Högbom (1974) CLEAN was used for decovolution in 8 frequency sub-bands, with the termination thresholds set to 100,000 iterations or a residual peak of 3μJy beam^{-1}, ensuring a deep clean is performed within the mask, and a deep sky model is produced. The resulting model was manually partitioned into typically $\sim$10 directions, the exact number depending on the location of off-axis problem sources in the field being imaged, as well as the need to retain suitable flux in the sky model per direction. The KILLMS package (e.g Smirnov & Tasse 2015) was then used to solve for a complex gain correction for each direction with a time / frequency interval of 5 min / 128 channels.

Another run of DDFACET reimaged the data, applying these directional corrections. For all imaging steps prior to this point (except in the case where a strong source is peeled using a higher resolution model) we use a Briggs' robustness parameter value of -0.3. However for the final images we run DDFACET twice, with robust values of 0.0 to generate an image optimised for sensitivity, and -1.2 to produce a shallower image with higher angular resolution to aid with the multiwavelength cross-matching process.

An example of the directional partitioning can be seen in Figure 1, with the thirteen regions (tessellations) used in the case of the Early Science COSMOS field (the position of which is shown by the central '+' marker') delineated by the grey convex polygons, representing the patches of sky over which a unique KILLMS solution is applied. Figure 1 also shows the full area that is imaged, with the inner and outer concentric circles showing the distance from the phase centre at which the primary beam gain has nominally dropped to values of 0.5 and 0.3 respectively. The background image of Figure 1 shows a saturated version of the full COSMOS image, and many sources outside the 30% primary beam level are visible. Please refer to the caption of Figure 1 and Section 4.5 for further information.
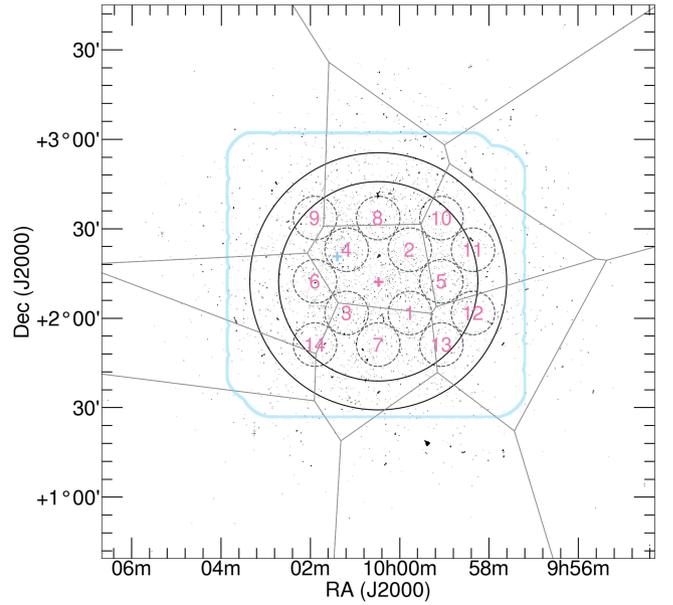
Figure 1. A schematic diagram of the COSMOS field. This shows the full extent of the images formed during the processing of the MIGHTEE continuum data, and the grey polygons show the tesselation pattern over which directional gain corrections are applied in the case of the COSMOS pointing (see Section 2). The imaged area is significantly larger than the main lobe of the MeerKAT primary beam, the 50% (0.97 deg^2) and 30% (1.62 deg^2) levels of which are shown by the concentric circles. The central '+' symbol marks the MIGHTEE Early Science COSMOS pointing centre. The numbers mark the locations of the additional pointings used in the analysis in Section 4.5, as summarised in Appendix B. The surrounding circles show the range of the nominal 97% cut of the primary beam pattern for each pointing. The additional '+' marker close to pointing 4 is the pointing centre of the CHILES survey (Fernández et al. 2013), and the blue box shows the coverage of the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017).

2.5 Restoring beams, primary beams, and mosaicking

We perform a post-processing step on the image products in order to impart uniform angular resolution to the mosaicked continuum products, within each of the four MIGHTEE Early Science fields. The clean component model produced by DDFACET is convolved with a circular Gaussian, the size of which is selected to slightly exceed the size of the largest major axis (for each of the two weighting settings) from the list of fitted restoring beams for the pointings in and given field. In the case of the Early Science data this was only applicable to XMM-LSS, however the COSMOS field was processed in the same way for consistency[9]. Following the convolution of the clean component model, we convolved the residual image with a homogenization kernel computed using the PYPHER package (Boucaud et al. 2016). The goal here is to bring the residual image to a resolution that closely matches that of the restored model, assuming that the fitted restoring beam

---

[9] This process was also used for the full COSMOS mosaic (Section 4.5 and Appendix B) to give the constituent images matched resolution.

well-approximates the shape of the main lobe of the synthesised beam (the true PSF). Following these two convolution operations, the model and residual images were summed. The differing of the synthesised beams from pointing to pointing is much less of an issue for MIGHTEE than for larger-area snapshot surveys, as the long tracks and relatively compact mosaics do not give rise to significant variations. However this process ensures that the mosaics can be delivered with restoring beam information in the header that is consistent across each of the constituent fields in a mosaic, which is beneficial for photometric accuracy (see Section 4.2), as well as image-plane stacking experiments and $P(D)$ measurements (Scheuer 1957; Condon et al. 2012).

The convolved image products for each pointing were corrected for primary beam attenuation by dividing them by a model of the Stokes I primary beam pattern, evaluated at 1284 MHz using the EIDOS (Asad et al. 2021) package. The main lobe of MeerKAT's primary beam is not circularly symmetric, resulting in a beam gain that varies with azimuthal angle by ∼5% at the nominal half-power point. We smear this variation out by azimuthally-averaging the beam model prior to the division.

Finally, the image products at both resolutions were mosaicked together using the MONTAGE[10] software, using the usual variance-weighted linear combination of pointings, and blanking pixels in each of the constituent images beyond the distance from the phase centre where the beam gain nominally drops below 0.3.

3 EARLY SCIENCE CONTINUUM DATA PRODUCTS

The observations listed in Table 1 were processed using the methods described in Section 2, and the resulting images presented here form the basis of the MIGHTEE Early Science continuum data. The subsections that follow describe the final total intensity images, as well as additional derived image and catalogue products.

3.1 Total intensity images

The total intensity MIGHTEE Early Science images for COSMOS and XMM-LSS are shown in Figures 2 and 3 respectively. These images are both the low resolution / high sensitivity variants, imaged with a robust weighting value of 0.0, with angular resolutions of 8″.6 and 8″.2 for COSMOS and XMM-LSS respectively.

We measure the thermal noise in these images by taking the RMS of the pixel values in clean, source-free regions away from the main lobe of the primary beam in the images prior to primary beam correction. In the robust 0.0 COSMOS image, the thermal noise is 1.7 μJy beam^{-1}. The deepest combined part of the robust 0.0 XMM-LSS mosaic reaches 1.5 μJy beam^{-1}. Both of these values exceed the design goals of the MIGHTEE survey, which was to reach a depth of 2 μJy beam^{-1}, based on simulations of the angular resolution of MeerKAT, its sensitivity, and the expected classical confusion limit.

The central regions of the robust 0.0 images are indeed limited by classical confusion rather than thermal noise. Classical confusion imposes a fundamental limit to the depth of an astronomical image, occuring when the surface density of discrete sources increases to the point where the angular resolution of the instrument is no longer sufficient to separate them. An often used definition of classical confusion is that it occurs when the number of synthesised beam solid angles per source falls below some value (typically 10–20). There is however no formal way to determine the level at which it affects an image. The classical confusion limit depends not only on the angular resolution of the image but its depth (coupled to the shape of the source counts function), as well as local astronomical clustering and sample variance effects (Heywood, Jarvis, & Condon 2013). For example, the RMS of the pixels in the central region of the (robust 0.0) residual COSMOS image is ∼5.5 μJy beam^{-1}, although this will contain some contribution from the sub-5.5σ sources that did not get included in the final cleaning mask, were not deconvolved, and thus are still present in the residual image. This is therefore likely an overestimate of the true confusion limit.

Figure 4 shows the inner 0.75×0.375 deg^2 of the COSMOS image, in order to contrast the image in Figure 2 with its higher angular resolution, less confused counterpart. The higher angular resolution (robust −1.2) images in COSMOS and XMM-LSS are not limited by classical confusion, and have 1σ noise levels of 5.5 and 6 μJy beam^{-1} respectively. The angular resolution of the robust −1.2 images is 5″ in both of the Early Science fields.

3.2 Effective frequency images

The antenna primary beam response causes the gain of a telescope to be a strong function of direction, but the pattern width is also inversely proportional to the observing frequency. Although MFS imaging techniques allow broadband interferometric data to be used to produce deep continuum images, the effective frequency at which each pixel samples the sky brightness distribution in the final image can differ significantly from the nominal band centre frequency. A source catalogue derived from a broadband image or mosaic will contain flux density and brightness measurements that are made at a range of effective observing frequencies. Here we determine the effective observing frequency as a function of position in order to provide consistent measurements in the source catalogues (see Section 3.3).

For each individual pointing, we calculate the effective frequency (for a spectrally flat source) ν_{eff} for each pixel (x,y) in the full-band image using the weighted mean

$$\nu_{\mathrm{eff}}(x,y) = \frac{\sum_i \nu_i \, A_i(x,y) \, \sigma_i^{-2}}{\sum_i A_i(x,y) \, \sigma_i^{-2}}, \qquad (3)$$

where $A_i(x,y)$ is the primary beam attenuation at pixel (x,y) for sub-band i, and σ_i is the 1σ noise level in sub-band i. The noise levels are measured directly from each of the eight sub-band images. It is assumed that these sub-band sensitivity measurements coarsely capture several effects, including the differing amount of RFI losses in each sub-band, the frequency-dependent system temperature, and the imaging weights. The effective frequency will also be sensitive to the latter, since the robust −1.2 images will give more weight to longer baselines, and RFI generally affects shorter spacings.

---

[10] http://montage.ipac.caltech.edu/
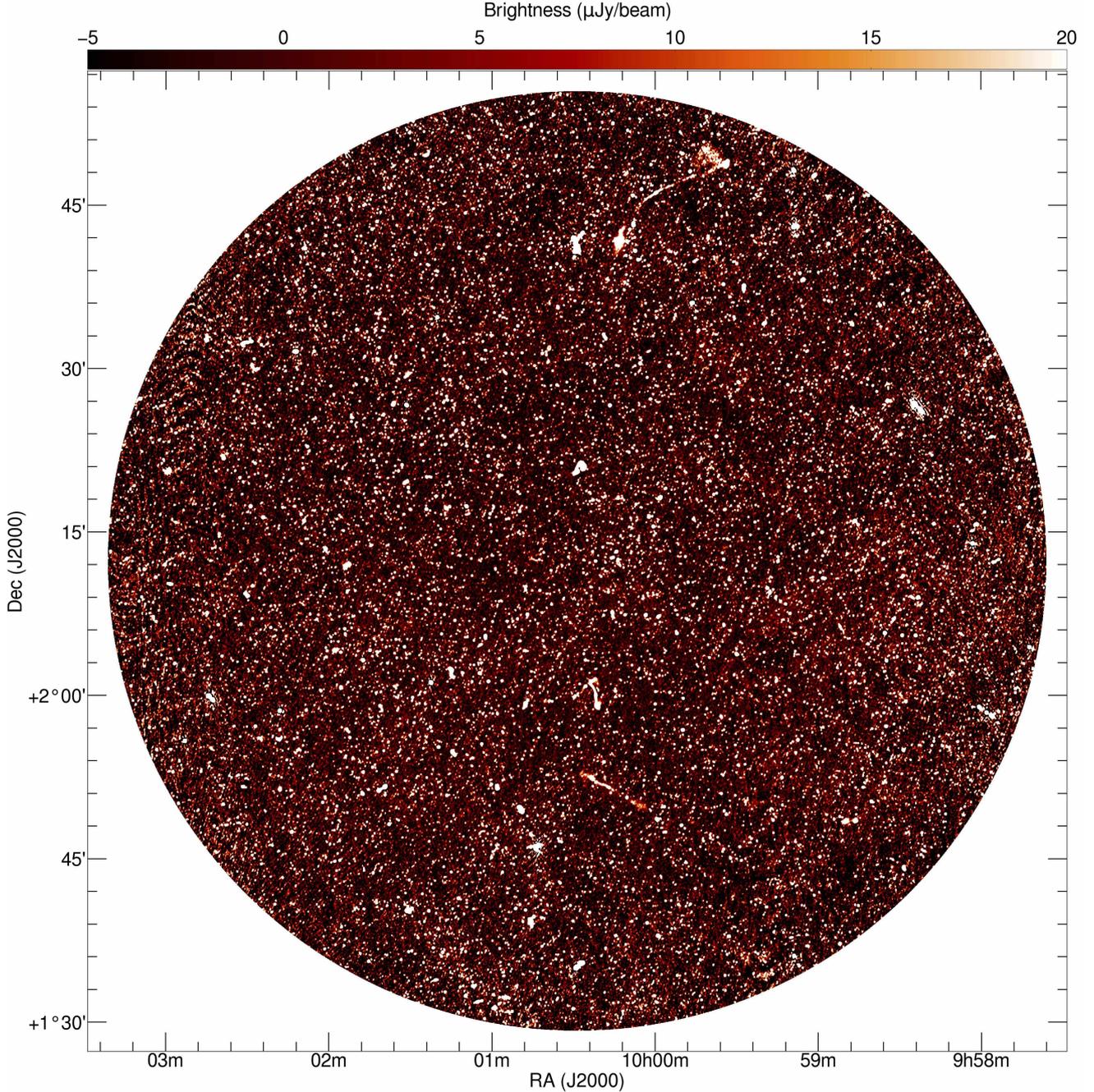
Figure 2. The MIGHTEE COSMOS Early Science image. The angular resolution is $8''.6$, and while the thermal noise in the data is 1.7 μJy beam^{-1} this map is limited by classical confusion at approximately 4.5 μJy beam^{-1}. The image covers 1.6 deg^2 and contains almost 10,000 radio components with peak brightnesses exceeding $5\sigma_{\mathrm{local}}$ (see Section 3.3 for details). The pair of giant radio galaxies reported by Delhaize et al. (2021) are visible towards the top and in the lower third of the image.

The relative contribution of RFI-prone regions will thus be decreased in images made with weighting that is closer to uniform.

For the XMM-LSS field, the three per-pointing effective frequency maps are mosaicked together using the same weighting scheme as was used for the total intensity mosaic shown in Figure 3. The suitability of using a narrowband primary beam correction factor in a broadband MFS image is a separate issue that we investigate in Section 4.5.

Figure 5 shows histograms of the effective frequency dis-

tribution as a function of area for both the COSMOS and XMM-LSS images, and for each of the two robust values that the data were imaged with. The weighting that delivers higher angular resolution has a lower effective frequency, consistent with the distribution of the RFI across MeerKAT's L-band, and it affecting primarily shorter spacings. The range of effective frequencies in the mosaic spans approximately 100 MHz in all cases, and deviates from the nominal band centre frequency by a maximum of 157 MHz (the lowest effective frequency in the high resolution XMM-LSS mosaic).
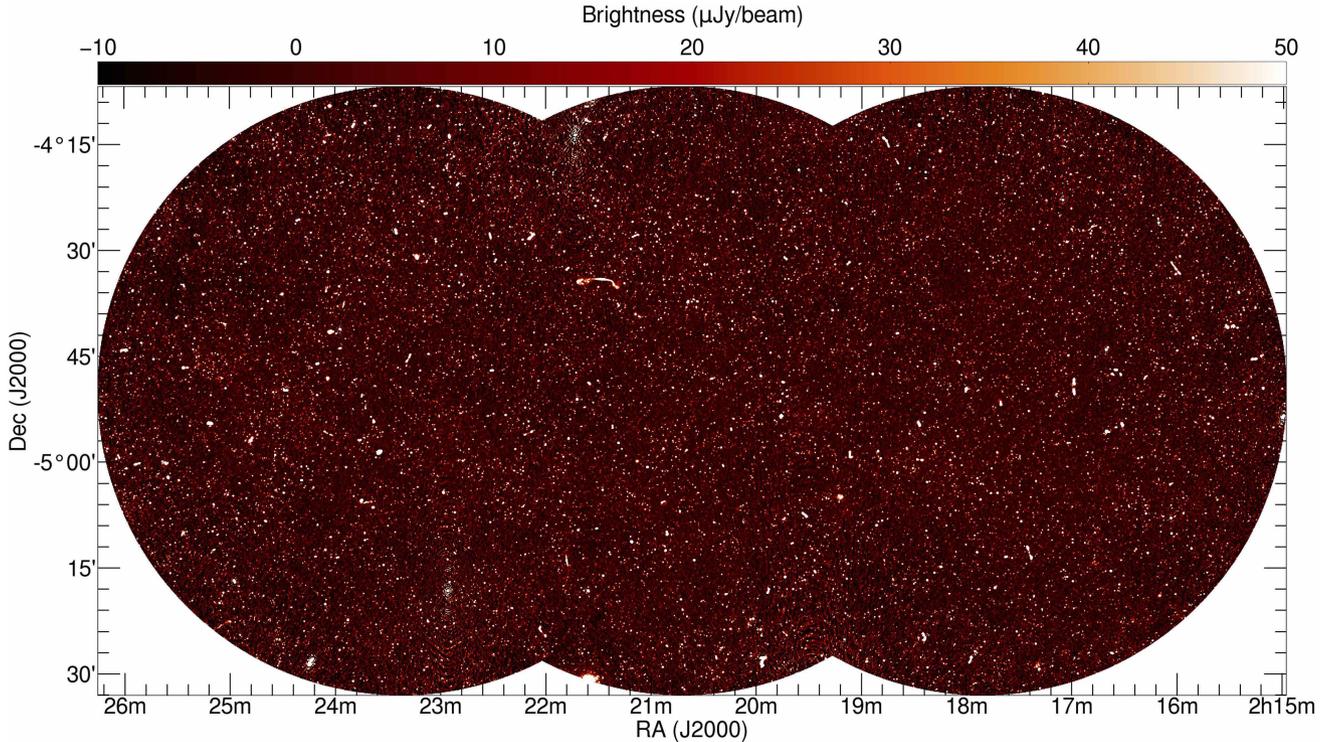
Figure 3. The three-pointing MIGHTEE Early Science mosaic covering 3.5 deg^2 of the XMM-LSS field, with an angular resolution of 8″.2.

This corresponds to a maximum flux density correction factor of 8.4 per cent for a source with a spectral index of -0.7. We return to this issue in the context of the Early Science source catalogues in Section 3.3.4.

3.3 Radio component catalogues

Here we describe the production of the catalogues for the MIGHTEE Early Science continuum data. This process makes use of an automated source finder to extract features from the image. The resulting database of components is then processed further, making use of some of the additional image products described above. We refer to the 'raw' output of the source finding software as the Level-0 catalogue, and the processed catalogue as the Level-1 catalogue, both of which are made available. The structure of the Level-1 catalogue is given in Appendix A. The cross-identification of the radio components with their optical / NIR counterparts (Level-2 catalogues) will be presented by Prescott et al. (*in prep.*).

3.3.1 Source finding

We used the PYBDSF source finder (Mohan & Rafferty 2015) to locate and characterise components in the primary beam corrected COSMOS image, and the XMM-LSS mosaic. Briefly, PYBDSF works by using a sliding box to estimate the local background noise ($\sigma_{\rm local}$) as a function of position in the image. It then locates pixels in the image whose brightness exceeds the local background noise by some factor (in this case $5\sigma_{\rm local}$). A flood-fill algorithm is then used to identify islands of contiguous emission down to some secondary threshold,

in this case $3\sigma_{\rm local}$. These islands are then iteratively fitted with point and 2D Gaussian components, and PYBDSF then attempts to group point and Gaussian components into sources, based on the brightness of the pixels between components in relation to the secondary threshold, and a separation criterion based on the measured sizes of the components. A catalogue describing the properties of the components and source groupings is then exported, and this raw output forms our Level-0 catalogue products. The Level-0 catalogues derived from the robust 0.0 images contain 9,915 and 20,397 components in COSMOS and XMM-LSS respectively. The steps in the sections that follow describe the modifications and additions that are made to produce the Level-1 catalogues.

3.3.2 Visual inspection

We visually examined the components by overlaying them on the FITS images for COSMOS and XMM-LSS. Spurious features around strong sources were removed. The nearby (z = 0.007651, see also Maddox et al. 2021) spiral galaxy NGC 895 (RA 02$^{\rm h}$21$^{\rm m}$36$^{\rm s}$47 -05°31′17″0) in XMM-LSS is also bisected by the edge of the mosaic where the primary beam cut-off was applied, so the incomplete set of components associated with this source was also removed from the catalogue. In total, 19 and 123 components were removed from COSMOS and XMM-LSS respectively, bringing their total numbers of components to 9,896 and 20,274. Note that although PYBDSF makes some attempt to group *components* together into *sources*, the results are generally not perfect. Users of the catalogue products who wish to detemine the properties of extended sources that consist of many components should
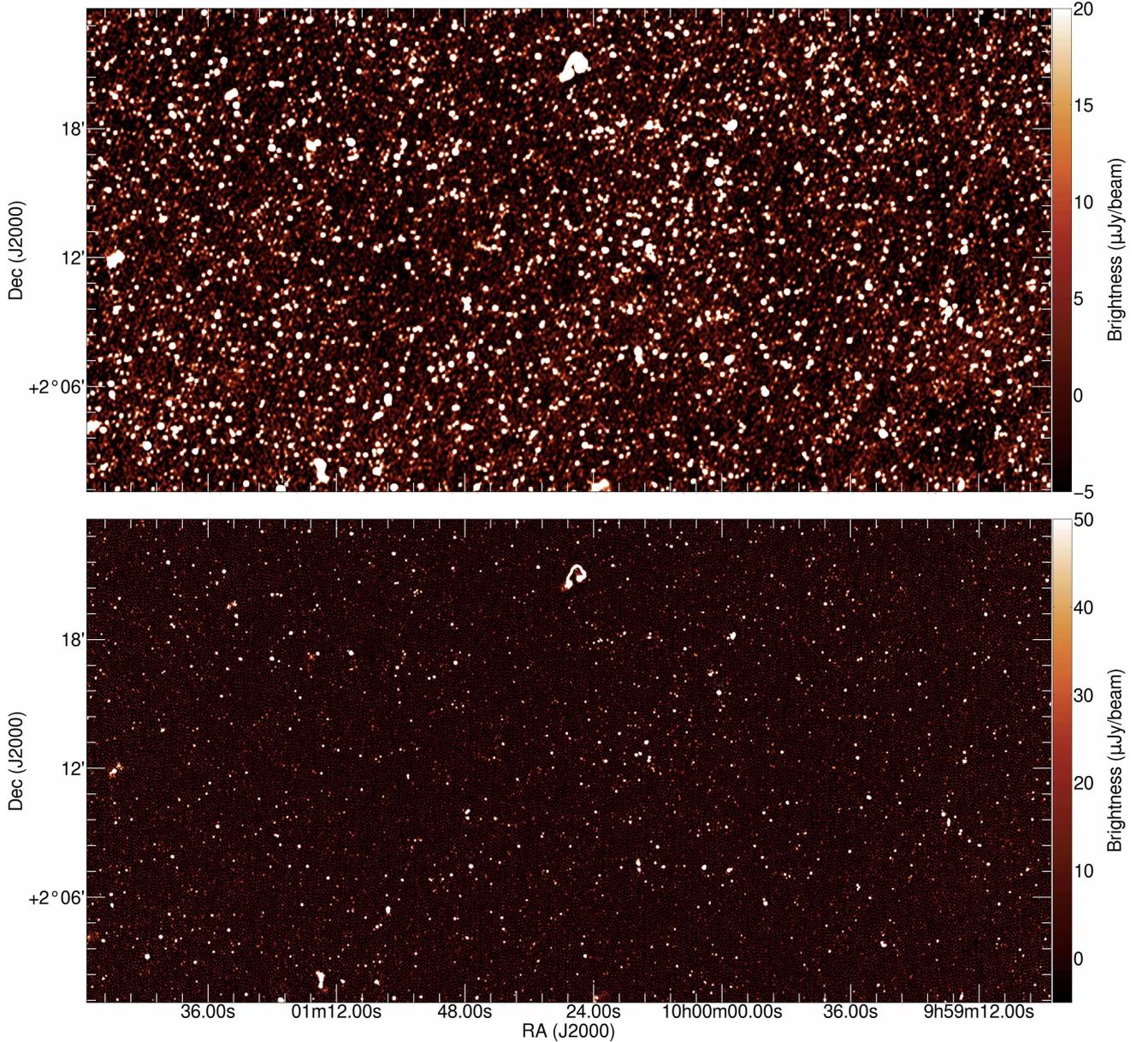
Figure 4. The central 0.75×0.375 deg^2 of the COSMOS Early Science image, showing the effects of the two different weighting schemes. The upper and lower images have angular resolutions of $8''.6$ and $5''$ respectively. The higher angular resolution image is noise limited rather than confusion limited, with a 1σ noise level of 5.5 μJy beam^{-1}.

perform this grouping manually. Prescott et al. (*in prep.*) will provide a Level-2 catalogue that groups the radio components of a particular source together, in addition to identifying the optical hosts of the MIGHTEE continuum detections.

3.3.3 Resolved sources

Following Murphy et al. (2017), the deconvolved size $(\theta_M \times \theta_m)$ of a Gaussian component with a fitted size of $(\phi_M \times \phi_m)$, where the M and m subscripts denote the major and minor axes respectively, is given by

$$\theta_M = \sqrt{\phi_M^2 - \theta_{\mathrm{beam}}^2} \qquad (4)$$

where θ_{beam} is the FWHM of the circular restoring beam (and similarly for θ_m and ϕ_m; see Wild 1970 for the equations for the generalised case involving an elliptical restoring beam). Dropping the M and m subscripts, the uncertainty in the deconvolved source size is calculated according to

$$\sigma_\theta = \sigma_\phi \left[1 - \left(\frac{\theta_{\mathrm{beam}}}{\phi} \right)^2 \right]^{-1/2}. \qquad (5)$$
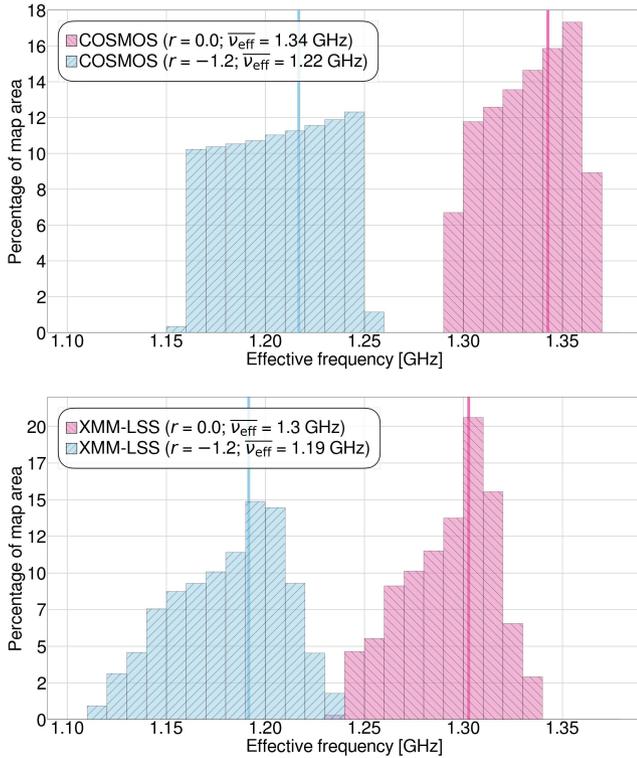
Figure 5. Histogram of the effective frequency values as a function of area for the COSMOS pointing (upper panel) and the XMM-LSS mosaic (lower panel). The two plots per panel show the distributions for the two different robust (*r*) values used in the imaging. The range of effective frequencies present in both images spans approximately 100 MHz. The vertical lines show the mean effective frequencies of each of the four images.

where σ_ϕ is the uncertainty in the fitted image component, as returned by PYBDSF. A source is deemed to be reliably resolved if its deconvolved major axis size exceeds the FWHM of the restoring beam by

$$\phi_M - \theta_{\text{beam}} \geq 2\sigma_{\phi_M}. \tag{6}$$

For each component in the Level-0 catalogue we compute the deconvolved source sizes and associated uncertainties according to Equations 4 and 5, and evaluate the relationship given in Equation 6 to flag sources that are deemed to be reliably resolved. The resulting values are provided in the Level-1 catalogues. For sources where PYBDSF returns unphysical fitted source sizes (i.e. the major or minor axes of the fitted components have extents that are smaller than that of the restoring beam), the Level-1 catalogues contain zeros for the deconvolved component sizes. The number of components that are flagged as resolved are 898 (10%) and 1376 (7%) in COSMOS and XMM-LSS respectively.

3.3.4 Effective frequencies

For each component in the catalogue we extracted the effective frequency (ν_{eff}) at its position from the maps described in Section 3.2, assuming a spectral index of -0.7, in order to correct the flux density and peak brightness measurements to a common frequency of 1.4 GHz. This frequency is selected as it is the commonly used reference frequency for previous

L-band continuum studies. As noted in Section 3.2 this correction factor will be at most 8.4% for the canonical synchrotron spectral index of -0.7. The mean correction factor across both Early Science fields is 4%, with a standard deviation of 1%.

4 DISCUSSION

Parts of the following subsections rely on the cross-matching of the component catalogues with additional catalogues, either from observations made with the VLA, or with those derived from alternative processing (see Section 4.5) of the MIGHTEE data. This is primarily for validation of the MIGHTEE continuum data products. In all cases where a cross-match is performed we conduct a nearest-neighbour match, and enforce the criteria that the radial separation of the two components must be less than $1''$, and the components in both catalogues must be described only by a single point or Gaussian component. The latter is to minimise any biases introduced by differing angular resolution, or sensitivity to extended emission. In the case where we make use of VLA data, rather than using the published catalogues, we obtain the radio images and run PYBDSF on them as described in Section 3.3.1 in order to further ensure that the catalogues being matched are as consistent as possible.

4.1 Astrometry verification

The absolute positional accuracy of a radio source is determined by a statistical component that is related to the signal-to-noise ratio (SNR) of the source and the angular resolution (e.g. Condon 1997), as well as a systematic component that is governed principally by the accuracy of the calibrator positions, and the transfer of the gain solutions derived from them. Atmospheric effects and primary beam phase gradients can also play a role in the wide field regime.

Determining the positional accuracy of the MIGHTEE data is important, as the potential of the data is fully realised when it is combined with observations of these fields from other facilities, including radio observations at other frequencies. Early observations made with MeerKAT had some potential issues that caused systematic errors in the measured positions of sources (Mauch et al. 2020, see also Knowles et al., *submitted*). The primary effects were: (i) timestamp offsets of 2 s (one correlator-beamformer interval), causing u, v, w coordinate errors that manifested as an apparent rotation about the phase centre; (ii) erroneous pointing (or catalogued) positions for calibrator sources, introducing positional offsets corresponding to that error into the target field if the calibrator was assumed to be a at the phase centre. A third consideration is the presence of significant structure in the emission of either the calibrator itself, or in the surrounding field. The effect of this has on the gain solutions is qualitatively similar to case (ii) above.

The first issue above was fixed by recomputing the u, v, w coordinates in the first stages of the data processing, and the MIGHTEE data were not ostensibly affected by the second issue. We investigate any remaining unforeseen astrometric issues by comparing the positions of the MIGHTEE sources to those measured from two surveys using the VLA, both of which have superior angular resolution to MeerKAT. In
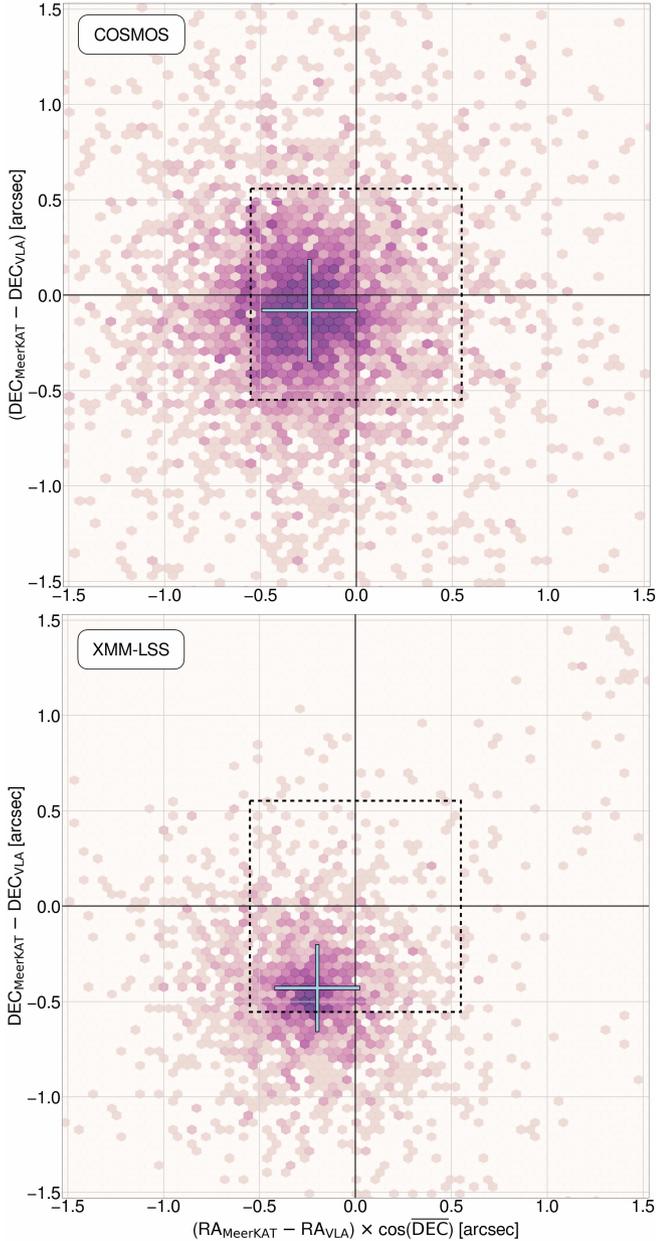
Figure 6. Differences between the MeerKAT and VLA positions for the matched components as described in Section 4.1. The figures show a 2D log-scale histogram of the distribution for COSMOS and XMM-LSS, and the cross shows the mean offset values, with the extent of the cross being 1σ in both directions. The mean (RA,Dec) values are $(-0''.27 \pm 0''.01, -0''.19 \pm 0''.01)$ and $(-0''.20 \pm 0''.01, -0''.43 \pm 0''.01)$ for COSMOS and XMM-LSS respectively, with the uncertainties quoted here being the standard error of the mean. The dashed box in the centre of the Figure shows the extent of a single map pixel $(1''.1 \times 1''.1)$ in the MIGHTEE data.

COSMOS we make use of the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017) with an angular resolution of $0''.75$, and in XMM-LSS the positional measurements of a 1.5 GHz survey with $4''.5$ angular resolution (Heywood et al. 2020a). Both of these VLA surveys are shallower than the robust 0.0 MIGHTEE Early Science images, however they cover slightly larger areas in both fields.

Figure 6 shows the differences in RA and Dec between the

MeerKAT and VLA components for COSMOS (upper panel) and XMM-LSS (lower panel). The mean offsets in (RA, Dec) are $(-0''.27 \pm 0''.01, -0''.19 \pm 0''.01)$ for COSMOS, and $(-0''.20 \pm 0''.01, -0''.43 \pm 0''.01)$ for XMM-LSS, with the uncertainties quoted being the standard error of the mean. The positions of the mean offsets are marked on the figure by the blue '+' symbol, the extent of which shows the standard deviation of the underlying offset distribution.

The forthcoming Level-2 catalogues (Prescott et al., *in prep*) contain host galaxy information from the visual cross-matching of the MIGHTEE Early Science data with the optical/NIR data (the latter being tied to the GAIA DR2 astrometric frame; Gordon et al. 2021). This process reveals mean offsets between the MIGHTEE positions and their optical/NIR hosts of $(-0''.03, 0''.01)$ in (RA, Dec) in the COSMOS field. But users of the MIGHTEE Early Science data should be aware that sub-pixel astrometric errors may be present between the MIGHTEE data and an external data set.

4.2 Photometry verification

Here we validate the integrated flux density measurements from the MIGHTEE Early Science catalogues (Section 3.3) by comparing them to matched components measured with the VLA. For the COSMOS field we use the 1.4 GHz measurements from the VLA-COSMOS Survey (Schinnerer et al. 2010) which was observed with a combination of the extended A and compact C configurations of the VLA. While this has significantly higher angular resolution ($1''.5$) than MIGHTEE, we opt for this over the deeper VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017) as the observing frequency is much closer to that of MIGHTEE and reduces the uncertainty introduced by spectral index corrections. Similarly, to compare the integrated flux density measurements in the XMM-LSS field we use a VLA mosaic at L-band made from observations in the more compact C and D configurations (Peters 2019). This mosaic uses the same pointing grid as the B-configuration mosaic used in Section 4.1 (Heywood et al. 2020a). Although the depth is shallower than the B-configuration mosaic, the angular resolution is lower, and the short spacing coverage is much better than that of the B-configuration mosaic.

The cross-matched VLA and MeerKAT components are shown on Figure 7 for the COSMOS and XMM-LSS fields. The increasing scatter with decreasing integrated flux density due to the increasing fractional noise contribution is evident. The diagonal line on the plot shows the 1:1 ratio. The (mean,median) MeerKAT:VLA ratios for the entire distribution are (0.96,0.95) for COSMOS and (0.99,1.0) for XMM-LSS, and are therefore consistent.

4.3 Reliability of resolved sources

We estimate the reliability of the resolved criterion described in Section 3.3.3 by means of a simulation. A true point source is injected at a random location into the COSMOS image, with peak brightness drawn from a uniform distribution with values between 10 μJy beam^{-1} and 1 mJy beam^{-1}. The PYBDSF source finder is run on the resulting image, and the properties of the fitted component at the location of the
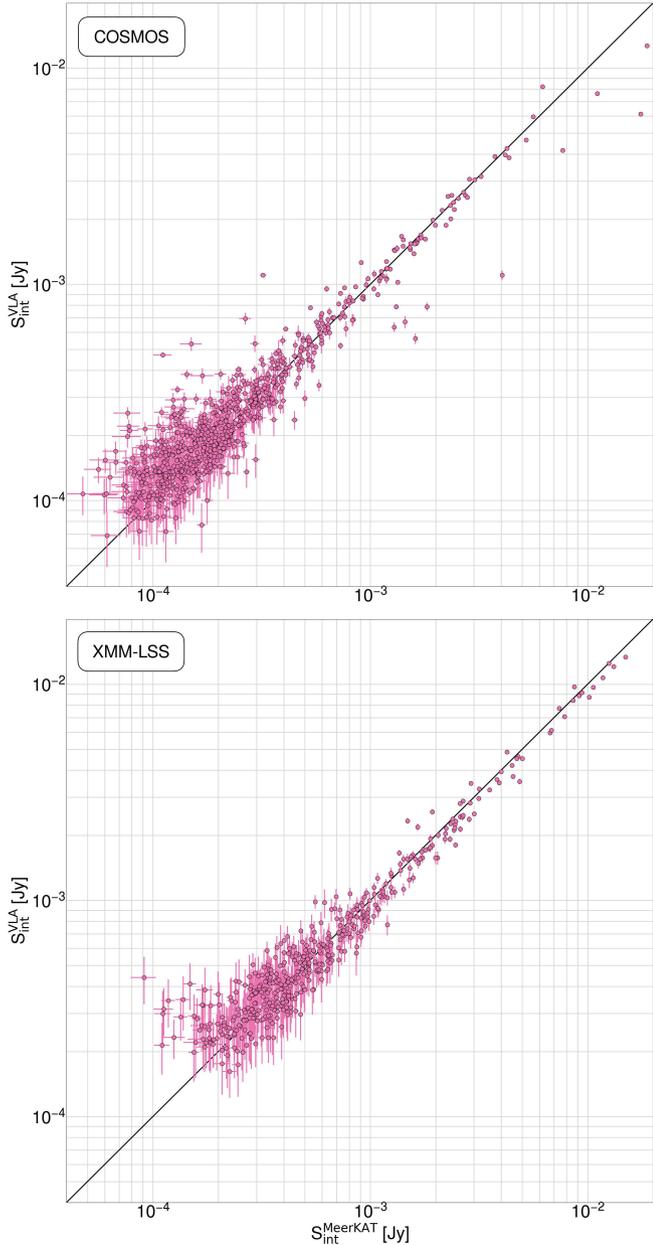
Figure 7. Integrated flux density measurements from the MIGHTEE catalogues plotted against matched components from VLA observations of the corresponding field. Details of the matching procedure is given in Section 4, and of the VLA observations used in Section 4.2. The 1:1 line is shown across the diagonal of the figure.

Figure 8. Percentage of sources that are erroneously marked as resolved as a function of distance from the phase centre in the COSMOS image. The results are presented in two peak brightness bins, as well as the total counts, as indicated on the legend. This plot was generated by injecting a true point source into the data and recovering its properties using PYBDSF, a process that we repeated to characterise 82,000 simulated sources.

injected source are compared. We repeated this process to compare the true vs recovered properties of 82,000 simulated sources.

Figure 8 shows the percentage of true point sources that are (erroneously) deemed to be resolved as a function of angular distance from the phase centre of the COSMOS image, and in two peak brightness bins ($0.01 < S_{peak} < 0.1$, and $0.1 < S_{peak} < 1.0$, with peak brightness values in mJy beam^{-1}). In the higher brightness bin the fraction of sources that are erroneously found to be resolved is 1–2 per cent. For the lower brightness bin the fraction reaches 10 per cent in the centre

of the map, with a gradual decline to 2 per cent as the angular distance increases. The thermal noise in the MIGHTEE Early Science images is below 2 μJy beam^{-1}, and the source finder was configured with a peak brightness threshold of $5\sigma_{local}$. Thus although the faintest sources in the simulation have peak brightnesses that are $\sim$5 times the thermal noise, in practice these sources are never recovered: classical confusion elevates the effective noise in the centre of the image, and the primary beam correction raises both the confusion and thermal contributions away from the field centre. The shape of the curve for the lowest brightness bin will be governed by these two effects. Although the simulated source brightness was drawn from a uniform distribution between the two limits, the brighter sources dominate the counts in the simulation. This explains why the fractions for the total simulated population shown on Figure 8 is closer to those of the higher brightness bin. We investigate the issue of completeness in the next section.

Since the goal for the final MIGHTEE continuum survey is to achieve uniform sensitivity by means of close-packed mosaics, the innermost angular separation bin of Figure 8 is likely to be representative for the fraction of reliably resolved sources. In summary, in regions where the survey reaches its target depth, the fraction is conservatively 3 per cent on average, but may rise to 10 per cent for the fainter source population.

4.4 Completeness

We make use of the simulation described in Section 4.3 to examine catalogue completeness, as shown in Figure 9. To generate this figure we simply count sources that are injected into the data but not recovered by PYBDSF. In the full-depth (confused) regions, MIGHTEE's continuum data is 96 per cent complete at 1 mJy beam^{-1}, 90 per cent complete at 300 μJy beam^{-1}, and 60 per cent complete at 50 μJy beam^{-1}.

A companion paper (Hale et al., *in prep.*) presents an analysis of the differential source counts and sky background temperature, and includes a thorough analysis of the bias
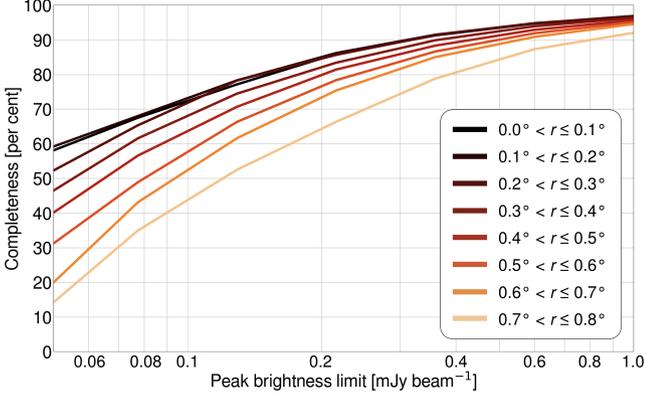
Figure 9. Point source completeness (expressed as a percentage) as a function of peak brightness in the COSMOS image. The completeness percentages are determined in concentric annuli with angular separation ranges from the phase centre indicated on the figure legend. The uppermost curves are representative of the point source completeness for the maximum depth (confused) regions of the COSMOS image (or mosaicked image products).

corrections required to account for completeness and other effects in the source counts derived from the Early Science catalogues.

4.5 Primary beam correction in broadband, widefield imaging

The 'traditional' approach for primary beam correction is to divide the final continuum image by an image of the primary beam pattern. For modern broadband observations the primary beam image is usually evaluated at the nominal central frequency of the observing band. This approach is potentially problematic for several reasons. The main issue, already touched upon in Section 3.2, is the large fractional bandwidth. In this case a single-frequency primary beam model becomes inaccurate for off-axis sources in a way that depends on both the bandwidth (including gaps due to RFI) and the intrinsic spectra of the sources. The second issue is that due to the telescope mount the primary beam is time dependent as well as frequency and direction dependent, and effects such as pointing errors impart an effective primary beam that differs from antenna to antenna. For anything other than a snapshot observation, an image-based primary beam model will also be a time (and thus direction) averaged product.

The two main alternatives to the traditional approach for primary beam correction are A-projection, and facet-based corrections. In the former approach a primary beam correction is applied to each visibility in the form of a convolution kernel during gridding (Bhatnagar, Rau, & Golap 2013). The advantage of this approach is that in principle the frequency, time, polarisation, and direction dependencies of the primary beam can be fully modelled, however the computational cost is significant. At the time of writing, an A-projection implementation for general use with MeerKAT is forthcoming (Sekhar et al. 2021), so this is not tested here. Facet-based primary beam corrections can be applied by DDFACET (Tasse et al. 2018) in much the same way as the direction-dependent gain corrections that are ap-

plied on a per-tessellation basis as described in Section 2.2. The beam model in this case consists of a multi-dimensional FITS model that contains a full 2×2 Jones matrix model of the MeerKAT beam as a function of direction, evaluated at several frequency intervals. The sky image is divided up into a regular grid of facets, and a piecewise primary beam correction is evaluated for the direction corresponding to the facet centre. The beam model for each facet is averaged in time according to the parallactic angle interval of the observation, and frequency dependence is captured by evaluating and applying the beam model for each of the (de-)gridding subbands. The beam corrections can also be optionally smoothed in the spatial dimensions. The advantage of this approach is that the sky model during deconvolution is generated in an intrinsic rather than apparent form, allowing for constraints on the source spectra that are astrophysically rather than instrumentally motivated, as well as fewer degrees of freedom when subsequently solving for direction-dependent effects.

In this section we compare the traditional and facet-based approaches, making use of additional MIGHTEE pointings in the COSMOS field, in order to match attenuated and corrected sources with counterparts that are on-axis in the additional pointings, and thus represent the intrinsic properties.

4.5.1 Production of additional images and catalogues

The MIGHTEE COSMOS observations form a close-packed mosaic of 15 individually deep pointings, the arrangement of which can be seen in Figure 1. The inner regions of the numbered pointings where the primary beam response is relatively flat can be used to provide intrinsic integrated flux density and peak brightness measurements for ensembles of sources that lie off-axis to varying degrees in the single central Early Science pointing. Thus the apparent and beam-corrected intrinsic measurements from the single pointing can be compared to a set of true intrinsic measurements derived from the additional pointings to test the validity of the primary beam model. For the true intrinsic measurements we restrict source extraction to the inner 0.12 degree radius of the 14 additional pointings. In this region the nominal primary beam gain does not fall below 97%.

Each of the 14 additional pointings were processed using the methods described in Section 2, up to and including the correction for DDEs (although no peeling step was used for any COSMOS pointings). The model and residuals for each pointing were also convolved to a common resolution prior to mosaicking. No primary beam correction was applied to the individual pointings, however each of them was masked to the circular regions shown on Figure 1 prior to mosaicking with MONTAGE in order to produce a single FITS image from which to extract a catalogue. The PYBDSF source finder was run on the resulting image. Components that were within 18″ of the edge of a pointing were excluded in order to avoid comparisons between sources that were partially missing due to the radial cuts. The resulting catalogue (hereafter the 'intrinsic' catalogue) contains 4,828 point and Gaussian components.

The 2GC-calibrated (Section 2.3) Early Science COSMOS data was imaged again with DDFACET, however for this run a model of the primary beam was used. The model was applied to the data over 1024 facets covering the 3.13×3.13 deg^2 area shown in Figure 1, although the beam model is interpolated and smoothed by DDFACET when producing the final image,

with an independent beam correction evaluated in each of 10 sub-bands. The beam model was generated using the EIDOS package (Asad et al. 2021). The real and imaginary part of each correlation product (XX, XY, YX and YY) are captured via a FITS cube (for a total of eight cubes). Each cube covers 6×6 deg^2 by means of 257×257 spatial pixels, and the full MeerKAT L-band is captured in 95 frequency channels.

Deconvolution of the data with the beam model enabled used the sub-space deconvolution mode (SSD2), with a second-order log-space polynomial capturing the (intrinsic) sky frequency dependence. Five major cycles were performed, with 60,000 iterations, and making use of the same post-2GC cleaning mask that was employed to produce the images described in Section 2.4.

The final primary beam corrected image was convolved to the common resolution using the same method as described in Section 2.5, and masked to the same circular region as the 'traditional' Early Science COSMOS image. Once again the PYBDSF source finder was used, and the resulting 'facet' catalogue contains 10,675 point and Gaussian components. The excess sources are mainly artefacts around strong sources that remain present despite the use of the beam model. Figure 10 illustrates this, showing cutouts of two bright double radio sources from the COSMOS field to show the differing levels of artefacts for different processing stages / methods. Each of these sources is about half a degree from the pointing centre. Another cause of the increased artefacts in the image with the faceted beam model could be the second-order polynomial used to fit the spectra of the clean components. While likely to be 'astrophysically' suitable given typical source spectra, a higher order fit may allow more flexibility to compensate for any deficiencies in the beam model for sources this far away from the pointing centre.

To produce the plots that follow, the cross-matching between the traditional (apparent and primary beam corrected), mosaic, and facet catalogues was performed using the same criteria described in Section 4. We further impose the additional constraint that the SNR of the components must exceed 20.

4.5.2 Apparent over intrinsic

An obvious check of the suitability of the primary beam model is to compare the apparent (no primary beam correction applied) flux density measurements from the MIGHTEE COSMOS Early Science image with the assumed-intrinsic measurements from the 'mosaic' catalogue. Such a comparison is shown in Figure 11, where the apparent to intrinsic ratio is plotted as a function of the angular distance of the source from the phase centre. The difference between the single frequency primary beam model (the solid black line) and the median measurements of the ratio in seven bins is shown on the figure, and is ∼1%.

The ratio shown by the black points exhibit significant scatter, and we investigate the expected level of this scatter by means of a Monte Carlo simulation. A simulated source is assigned a radial offset in the range [0,0.7] and its true flux density is assigned by selecting a random value from the list of 'mosaic' flux densities. A random spectral index value is assigned drawn from a normal distribution with a mean of -0.7 and a standard deviation of 0.1, consistent with the 1.4 GHz - 4.86 GHz spectral index distribution of the ex-
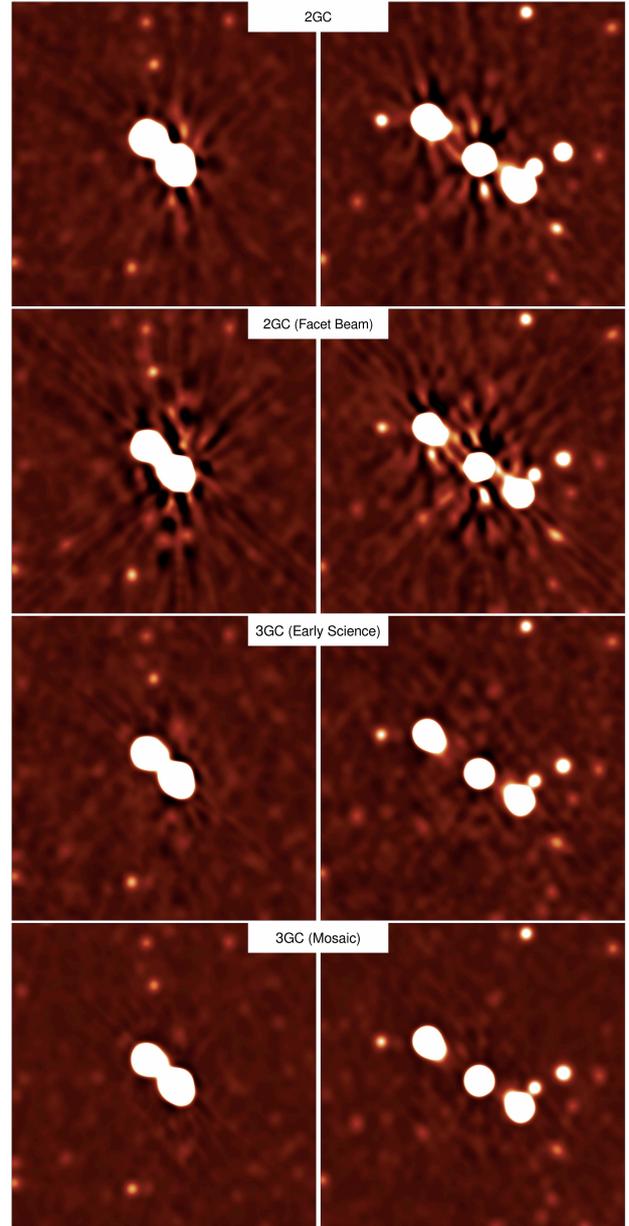
Figure 10. Thumbnails spanning $4'$, showing two bright double radio galaxies in the COSMOS field. Each of these sources is approximately half a degree from the pointing centre, and they are on opposite sides of the image. From top to bottom the panels show the imaging quality (i) after the 2GC step (Section 2.3); (ii) after the 2GC step but using SSD clean and a primary beam model (Section 4.5.1); (iii) after the 3GC step (Section 2.4); (iv) the full COSMOS mosaic, for which the increase in depth is apparent due the contributions of multiple overlapping pointings. The colour scale is the same as that used in Figures 2, 3, and 4, and runs from -50 to 150 μJy beam^{-1}.

tragalactic sky simulation presented by Wilman et al. (2008). Using the spectral index and the true flux density, an eight-point spectrum for the simulated source is computed, representing the eight sub-bands across the L-band in which the MeerKAT data is deconvolved. This true spectrum is then perturbed with frequency-dependent primary beam attenuations, as well as thermal noise that is appropriate for each
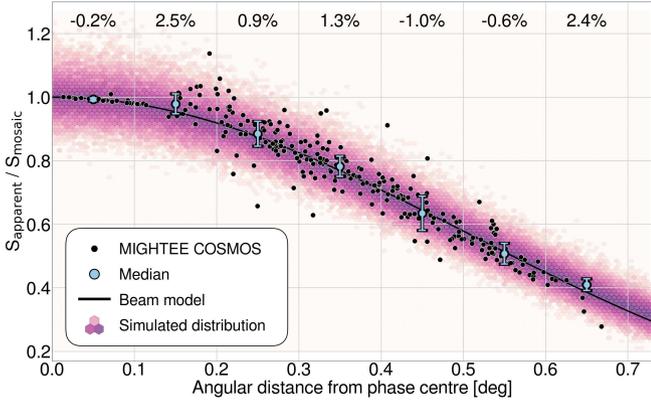
Figure 11. The apparent source flux densities from the Early Science COSMOS image divided by the presumed-intrinsic measurements produced for the 'mosaic' catalogue (see Section 4.5.1), as a function of the angular distance of the component from the phase centre (black points). The nominal primary beam model is shown by the solid black line, and the median (and median absolute deviation) is shown in seven bins by the heavy points with error bars. A Monte Carlo simulation is used to show the scatter on the points that is introduced by the source spectral indices and the differing noise properties of the two measurements. The results of simulating 150,000 components is shown in the background of this figure as a 2D log-scale histogram. Please refer to Section 4.5.2 for details. The percentages are the difference between the median points and the primary beam model.

sub-band, based on the measured values that were used in Section 3.2. The mean value of this corrupted spectrum is then taken to be the apparent measured flux density. The simulated ratio is produced by dividing this value by the intrinsic flux density of the source perturbed by appropriate Gaussian noise. This process is repeated 150,000 times, and the resulting distribution is plotted as the 2D log-scale histogram in the background of Figure 11. Note also that the additional COSMOS pointings are significantly shallower (and therefore noisier) than the central Early Science pointing (as can be see in the on-source time column of Table B1). In the final COSMOS mosaic the depth will be recovered by the co-addition of the many neighbouring pointings, however that is not done for this analysis in order to minimise the effects of the antenna primary beam on the flux density measurements. The increased depth provided by the co-addition of neighbouring pointings can be seen in the lowest row of Figure 10.

The distribution of the measured points is well represented by the simulation. The principal cause of the scatter is the differing noise properties of the two images (illustrated further by the innermost radii where the mosaic and the Early Science image share the same data), however the spectral index of the source also plays a part via the colour correction issues mentioned in Section 3.2, as evidenced by the deviation between the distribution of the simulated measurements and the primary beam model at high angular separations. The distribution of spectral indices based on the Wilman et al. (2008) simulation may be somewhat simplistic, e.g. flatter spectrum AGN cores may be under-represented at these depths (Whittam et al. 2017).

Despite these issues and the use of the single-frequency primary beam image for correcting broadband data, the me-

dian ratios between the apparent and intrinsic flux densities are typically ∼1% and in all radius bins are better than 3%.

4.5.3 Comparison of the traditional and facet-based methods

In this final section we make a comparison between the traditional and facet-based approaches for primary beam correction, as well as comparing both of these methods to the assumed-intrinsic 'mosaic' measurements. As in Section 4.5.2 these comparisons are made by determining the ratio of the flux densities of matched components as a function of their separation from the phase centre in order to determine the level of any systematic biases.

The results are presented in Figure 12, which shows such plots for traditional/mosaic (upper panel), facet/mosaic (middle panel), and facet/traditional (lower panel). Since all the flux density measurements that go into these plots are intrinsic, the unity line represents the point where there is no difference between the methods. The level of scatter in these plots is (as expected) consistent with the scatter evident in Figure 11, and as simulated in Section 4.5.2.

The median (and median absolute deviation) of the flux density ratios is plotted in seven bins, and the deviation from unity is expressed as a percentage label in each bin on the plots. The traditional correction method (upper panel) is in all bins consistent with unity. The medians of the facet/mosaic points (central panel) have hints of a sinusoidal structure, although the traditional/mosaic points also show a slight excess in the 0.1 − 0.4 deg bins. The facet/traditional distribution also begins to deviate from unity beyond a radius of 0.4, so the real differences could mainly be in the region beyond the half power point of the primary beam. Within this region there is essentially no clear difference between the traditional and facet based methods.

A further difference between the two methods arises via the deconvolution. As mentioned earlier, the DDFACET SSD approach is to fit for the intrinsic properties of the sky model. When producing the final image, this model is evaluated at the reference frequency and then restored into the residual. This means that the deconvolved part of the sky will in principle not be subjected to the effective frequency issues touched upon in Sections 3.2 and 4.2. The images used in the traditional approach however, are restored with components that are the weighted mean of the MFS model, which are then restored into the MFS-weighted residual. Clearly both of these approaches have their advantages; the former avoids the primary beam and position-dependent effective frequency effects, and while the latter method is subject to these effects, the deconvolved and residual portions of the image are at least consistent. As is often the case in radio astronomy one size does not fit all, and the choice of method often depends on the science goals and the observational parameters. For survey observations it is essential to conduct (and continue to conduct) tests such as this to investigate such systematics.

5 CONCLUSIONS

We have described in detail the total intensity continuum processing strategy for the MeerKAT MIGHTEE survey, and validated the suitability of this approach. This has been achieved by producing the continuum Early Science data products,

Figure 12. Flux density ratios for matched components as a function of their separation from the phase centre for (top to bottom) the traditional/mosaic, facet/mosaic, and facet/tradtional scenarios. The scatter in these plots is discussed in Section 4.5.2, and the fact that the additional COSMOS pointings that form the mosaic are shallower than the Early Science pointing is evident. The median ratios (and median absolute deviations) are shown in seven bins by the points with error bars, and the deviation from unity of these points expressed as a percentage is also noted in each bin.

totalling 5.1 deg^2 in the COSMOS and XMM-LSS fields, and performing both self-consistency checks and comparisons with existing radio observations in these fields. The imaging data is presented with two different depth / angular resolution combinations, and the images are decomposed into per-field catalogues, the combined component count of which is 30,174.

Corrections for direction-dependent effects (DDEs) are essential for delivering performance-limited images from MeerKAT, a goal which is highly desirable for surveys such as MIGHTEE where uniformity of the data is key, and for maximising their legacy value when not all science goals can

be foreseen. The use of a primary beam model during imaging does not fully correct for the DDEs in the image. Assuming no major deficiencies in the beam model, this suggests that the dominant causes are stochastic processes (mechanical pointing errors and possibly ionospheric effects at lower frequencies) that cannot be known *a priori* and must be solved for via self-calibration techniques.

Despite the major role the antenna primary beam patterns play in these (and all other) broadband, wide-field observations, we demonstrate that the traditional method of primary beam correction (namely the division of the final science image by a single frequency beam model at the band centre) is likely to be good enough for observations of this type. It does not appear to significantly bias the photometry extracted from the images, even at significant radii from the phase centre. Flux densities extracted from the primary beam corrected COSMOS image are consistent with those extracted from the central regions of 14 additional offset pointings at the ~1% level. This level is on par with other effects that introduce photometric uncertainties, such as absolute flux-scale calibration, and the uncertain colour corrections for sources that do not have a well-constrained spectral index measurement.

Several studies have made use of these data during the development and validation phase (e.g. Pasini et al. 2020; Delhaize et al. 2021; Delvecchio et al. 2021), however a public release of the image and catalogue products that form the Early Science MIGHTEE continuum data now accompanies this article. The properties of the Early Science continuum data meet or exceed the design specifications of the MIGHTEE survey, meaning that the continuum science goals of the full survey can also be met as planned. A set of sky footprints for the MIGHTEE survey that supersede those presented along with the science overview by Jarvis et al. (2016) are shown in Appendix C, with machine readable pointing centres available as supplementary material.

DATA AVAILABILITY

ACKNOWLEDGEMENTS

References

Asad K. M. B., Girard J. N., de Villiers M., Ansah-Narh T., Iheanetu K., Smirnov O., Santos M. G., et al., 2021, MNRAS, 502, 2970. doi:10.1093/mnras/stab104

Astropy Collaboration, Robitaille T. P., Tollerud E. J., Greenfield P., Droettboom M., Bray E., Aldcroft T., et al., 2013, A&A, 558, A33. doi:10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan A. M., Sipőcz B. M., Günther H. M., Lim P. L., Crawford S. M., Conseil S., et al., 2018, AJ, 156, 123. doi:10.3847/1538-3881/aabc4f

Baker A. J., Blyth S., Holwerda B. W., LADUMA Team, 2018, AAS

Bhatnagar S., Rau U., Golap K., 2013, ApJ, 770, 91. doi:10.1088/0004-637X/770/2/91

Boucaud A., Bocchio M., Abergel A., Orieux F., Dole H., Hadj-Youcef M. A., 2016, A&A, 596, A63. doi:10.1051/0004-6361/201629080

Briggs D. S., 1995, AAS

Condon J. J., 1992, ARA&A, 30, 575. doi:10.1146/annurev.aa.30.090192.003043

Condon J. J., 1997, PASP, 109, 166. doi:10.1086/133871

Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693. doi:10.1086/300337

Condon J. J., Cotton W. D., Fomalont E. B., Kellermann K. I., Miller N., Perley R. A., Scott D., et al., 2012, ApJ, 758, 23. doi:10.1088/0004-637X/758/1/23

Delhaize J., Heywood I., Prescott M., Jarvis M. J., Delvecchio I., Whittam I. H., White S. V., et al., 2021, MNRAS, 501, 3833. doi:10.1093/mnras/staa3837

Delvecchio I., Daddi E., Sargent M. T., Jarvis M. J., Elbaz D., Jin S., Liu D., et al., 2021, A&A, 647, A123. doi:10.1051/0004-6361/202039647

Fanaroff B., Lal D. V., Venturi T., Smirnov O., Bondi M., Thorat K., Bester L., et al., 2021, arXiv, arXiv:2105.11695

Fernández X., van Gorkom J. H., Hess K. M., Pisano D. J., Kreckel K., Momjian E., Popping A., et al., 2013, ApJL, 770, L29. doi:10.1088/2041-8205/770/2/L29

Gordon Y. A., Boyce M. M., O'Dea C. P., Rudnick L., Andernach H., Vantyghem A. N., Baum S. A., et al., 2021, arXiv, arXiv:2102.11753

Grobler T. L., Nunhokee C. D., Smirnov O. M., van Zyl A. J., de Bruyn A. G., 2014, MNRAS, 439, 4030. doi:10.1093/mnras/stu268

Hale C. L., McConnell D., Thomson A. J. M., Lenc E., Heald G. H., Hotan A. W., Leung J. K., et al., 2021, arXiv, arXiv:2109.00956

Heywood I., Jarvis M. J., Condon J. J., 2013, MNRAS, 432, 2625. doi:10.1093/mnras/stt843

Heywood I., Jarvis M. J., Baker A. J., Bannister K. W., Carvalho C. S., Hardcastle M., Hilton M., et al., 2016, MNRAS, 460, 4433. doi:10.1093/mnras/stw1250

Heywood I., Hale C. L., Jarvis M. J., Makhathini S., Peters J. A., Sebokolodi M. L. L., Smirnov O. M., 2020, MNRAS, 496, 3469. doi:10.1093/mnras/staa1770

Heywood I., Lenc E., Serra P., Hugo B., Bannister K. W., Bell M. E., Chippendale A., et al., 2020, MNRAS, 494, 5018. doi:10.1093/mnras/staa941

Heywood I., 2020, ascl.soft. ascl:2009.003

Heywood I., Murphy E. J., Jiménez-Andrade E. F., Armus L., Cotton W. D., DeCoursey C., Dickinson M., et al., 2021, ApJ, 910, 105. doi:10.3847/1538-4357/abdf61

Högbom J. A., 1974, A&AS, 15, 417

Hönig S. F., Watson D., Kishimoto M., Gandhi P., Goad M., Horne K., Shankar F., et al., 2017, MNRAS, 464, 1693. doi:10.1093/mnras/stw2484

Hotan A. W., Bunton J. D., Chippendale A. P., Whiting M., Tuthill J., Moss V. A., McConnell D., et al., 2021, PASA, 38, e009. doi:10.1017/pasa.2021.1

Hurley-Walker N., Callingham J. R., Hancock P. J., Franzen T. M. O., Hindson L., Kapińska A. D., Morgan J., et al., 2017, MNRAS, 464, 1146. doi:10.1093/mnras/stw2337

Jarvis M. J., Smith D. J. B., Bonfield D. G., Hardcastle M. J., Falder J. T., Stevens J. A., Ivison R. J., et al., 2010, MNRAS, 409, 92. doi:10.1111/j.1365-2966.2010.17772.x

Jarvis M. J., Bonfield D. G., Bruce V. A., Geach J. E., McAlpine K., McLure R. J., González-Solares E., et al., 2013, MNRAS, 428, 1281. doi:10.1093/mnras/sts118

Jarvis M., Taylor R., Agudo I., Allison J. R., Deane R. P., Frank B., Gupta N., et al., 2016, in *MeerKAT Science: On the pathway to the SKA*, PoS, 6

Jonas J., MeerKAT Team, 2016, in *MeerKAT Science: On the pathway to the SKA*, PoS, 1

Kemball, A. J. & Wieringa, M. H., 2000, CASA Memo 1, https://casa.nrao.edu/casadocs/casa-6.1.0/memo-series/casa-memos

Kenyon J. S., Smirnov O. M., Grobler T. L., Perkins S. J., 2018, MNRAS, 478, 2399. doi:10.1093/mnras/sty1221

Lacy M., Baum S. A., Chandler C. J., Chatterjee S., Clarke T. E., Deustua S., English J., et al., 2020, PASP, 132, 035001. doi:10.1088/1538-3873/ab63eb

Laing R. A., Guidetti D., Bridle A. H., Parma P., Bondi M., 2011, MNRAS, 417, 2789. doi:10.1111/j.1365-2966.2011.19436.x

Lindegren L., Hernández J., Bombrun A., Klioner S., Bastian U., Ramos-Lerate M., de Torres A., et al., 2018, A&A, 616, A2. doi:10.1051/0004-6361/201832727

Lonsdale C. J., Cappallo R. J., Morales M. F., Briggs F. H., Benkevitch L., Bowman J. D., Bunton J. D., et al., 2009, IEEEP, 97, 1497. doi:10.1109/JPROC.2009.2017564

Maddox N., Frank B. S., Ponomareva A. A., Jarvis M. J., Adams E. A. K., Davé R., Oosterloo T. A., et al., 2021, A&A, 646, A35. doi:10.1051/0004-6361/202039655

Matthews A. M., Condon J. J., Cotton W. D., Mauch T., 2021, arXiv, arXiv:2104.11756

Mauch T., Klöckner H.-R., Rawlings S., Jarvis M., Hardcastle M. J., Obreschkow D., Saikia D. J., et al., 2013, MNRAS, 435, 650. doi:10.1093/mnras/stt1323

Mauch T., Cotton W. D., Condon J. J., Matthews A. M., Abbott T. D., Adam R. M., Aldera M. A., et al., 2020, ApJ, 888, 61. doi:10.3847/1538-4357/ab5d2d

McConnell D., Hale C. L., Lenc E., Banfield J. K., Heald G., Hotan A. W., Leung J. K., et al., 2020, PASA, 37, e048. doi:10.1017/pasa.2020.41

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, Astronomical Data Analysis Software and Systems XVI, 376, 127

Mohan, N., & Rafferty, D. 2015, Astrophysics Source Code Library, ascl:1502.007

Mooley K. P., Hallinan G., Bourke S., Horesh A., Myers S. T., Frail D. A., Kulkarni S. R., et al., 2016, ApJ, 818, 105. doi:10.3847/0004-637X/818/2/105

Mouri Sardarabadi A., Koopmans L. V. E., 2019, MNRAS, 483, 5480. doi:10.1093/mnras/sty3444

Murphy E. J., Momjian E., Condon J. J., Chary R.-R., Dickinson M., Inami H., Taylor A. R., et al., 2017, ApJ, 839, 35. doi:10.3847/1538-4357/aa62fd

Muxlow T. W. B., Thomson A. P., Radcliffe J. F., Wrigley N. H., Beswick R. J., Smail I., McHardy I. M., et al., 2020, MNRAS, 495, 1188. doi:10.1093/mnras/staa1279

Noordam J. E., Smirnov O. M., 2010, A&A, 524, A61. doi:10.1051/0004-6361/201015013

Norris R., 2015, fers.conf, 5

Offringa A. R., McKinley B., Hurley-Walker N., Briggs F. H., Wayth R. B., Kaplan D. L., Bell M. E., et al., 2014, MNRAS, 444, 606. doi:10.1093/mnras/stu1368

Offringa A. R., Smirnov O., 2017, MNRAS, 471, 301. doi:10.1093/mnras/stx1547

Pasini T., Brüggen M., de Gasperin F., Bîrzan L., O'Sullivan E., Finoguenov A., Jarvis M., et al., 2020, MNRAS, 497, 2163. doi:10.1093/mnras/staa2049

Peters, J. A., 2019, PhD thesis, Univ. Oxford https://ora.ox.ac.uk/objects/uuid:9e9df12c-c9ec-4bfc-88f8-a4299f5c384d

Petrov L., Kovalev Y. Y., Fomalont E. B., Gordon D., 2011, AJ, 142, 35. doi:10.1088/0004-6256/142/2/35

Prandoni I., Guglielmino G., Morganti R., Vaccari M., Maini A., Röttgering H. J. A., Jarvis M. J., et al., 2018, MNRAS, 481, 4548. doi:10.1093/mnras/sty2521

Reynolds, J., 1994, ATNF Technical Memos, 39.3 040, http://www.atnf.csiro.au/observers/memos/d96783~1.pdf

Sekhar S., Jagannathan P., Kirk B., Bhatnagar S., Taylor R., 2021, arXiv, arXiv:2107.10009

Scheuer P. A. G., 1957, PCPS, 53, 764. doi:10.1017/S0305004100032825

Schinnerer E., Sargent M. T., Bondi M., Smolčić V., Datta A., Carilli C. L., Bertoldi F., et al., 2010, ApJS, 188, 384. doi:10.1088/0067-0049/188/2/384

Shimwell T. W., Röttgering H. J. A., Best P. N., Williams W. L., Dijkema T. J., de Gasperin F., Hardcastle M. J., et al., 2017, A&A, 598, A104. doi:10.1051/0004-6361/201629313

Smirnov O. M., 2011, A&A, 527, A106. doi:10.1051/0004-6361/201016082

Smirnov O. M., 2011, A&A, 527, A107. doi:10.1051/0004-6361/201116434

Smirnov O. M., 2011, A&A, 527, A108. doi:10.1051/0004-6361/201116435

Smirnov O. M., Tasse C., 2015, MNRAS, 449, 2668. doi:10.1093/mnras/stv418

Smolčić V., Novak M., Bondi M., Ciliegi P., Mooley K. P., Schinnerer E., Zamorani G., et al., 2017, A&A, 602, A1. doi:10.1051/0004-6361/201628704

Tasse C., Hugo B., Mirmont M., Smirnov O., Atemkeng M., Bester L., Hardcastle M. J., et al., 2018, A&A, 611, A87. doi:10.1051/0004-6361/201731474

Tasse C., Shimwell T., Hardcastle M. J., O'Sullivan S. P., van Weeren R., Best P. N., Bester L., et al., 2020, arXiv, arXiv:2011.08328

van Cappellen W. A., Oosterloo T. A., Verheijen M. A. W., Adams E. A. K., Adebahr B., Braun R., Hess K. M., et al., 2021, arXiv, arXiv:2109.14234

van der Vlugt D., Algera H. S. B., Hodge J. A., Novak M., Radcliffe J. F., Riechers D. A., Röttgering H., et al., 2021, ApJ, 907, 5. doi:10.3847/1538-4357/abcaa3

van Haarlem M. P., Wise M. W., Gunst A. W., Heald G., McKean J. P., Hessels J. W. T., de Bruyn A. G., et al., 2013, A&A, 556, A2. doi:10.1051/0004-6361/201220873

van Weeren R. J., de Gasperin F., Akamatsu H., Brüggen M., Feretti L., Kang H., Stroe A., et al., 2019, SSRv, 215, 16. doi:10.1007/s11214-019-0584-z

White S. V., Jarvis M. J., Häußler B., Maddox N., 2015, MNRAS, 448, 2665. doi:10.1093/mnras/stv134

White S. V., Jarvis M. J., Kalfountzou E., Hardcastle M. J., Verma A., Cao Orjales J. M., Stevens J., 2017, MNRAS, 468, 217. doi:10.1093/mnras/stx284

Whittam I. H., Riley J. M., Green D. A., Jarvis M. J., 2016, MNRAS, 462, 2122. doi:10.1093/mnras/stw1725

Whittam I. H., Jarvis M. J., Green D. A., Heywood I., Riley J. M., 2017, MNRAS, 471, 908. doi:10.1093/mnras/stx1564

Wild J. P., 1970, AuJPh, 23, 113. doi:10.1071/PH700113

Wilman R. J., Miller L., Jarvis M. J., Mauch T., Levrier F., Abdalla F. B., Rawlings S., et al., 2008, MNRAS, 388, 1335. doi:10.1111/j.1365-2966.2008.13486.x

Author Affiliations

[1]Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, UK

[2]Centre for Radio Astronomy Techniques and Technologies, Department of Physics and Electronics, Rhodes University, PO Box 94, Makhanda, 6140, South Africa

[3]South African Radio Astronomy Observatory, 2 Fir Street, Black River Park, Observatory, Cape Town, 7925, South Africa

[4]Physics Department, University of the Western Cape, Private Bag X17, Bellville, 7535, South Africa

[5]Institute for Astronomy, Royal Observatory Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ

[6]The Inter-University Institute for Data Intensive Astronomy (IDIA), Department of Astronomy, University of Cape Town,

Private Bag X3, Rondebosch, 7701, South Africa

[7]GEPI, Observatoire de Paris, CNRS, Université Paris Diderot, 5 Place Jules Janssen, 92190 Meudon, France

[8]Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, PT1349-018 Lisboa, Portugal

[9]Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisbon, Portugal

[10]SUPA, Institute for Astronomy, Royal Observatory Edinburgh, EH9 3HJ, UK

[11]The Inter-University Institute for Data Intensive Astronomy (IDIA), Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa

[12]School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia

[13]CSIRO Astronomy and Space Science, PO Box 1130, Bentley, WA, 6102, Australia

[14]Wits Centre for Astrophysics, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, Johannesburg 2050, South Africa

[15]Department of Physics, University of Pretoria, Hatfield, Pretoria 0028, South Africa

[16]Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

[17]Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK

[18]Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 Munich, Germany

[19]National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

[20]INAF-IRA, Via P. Gobetti 101, I-40129, Italy

[21]South African Astronomical Observatory, P.O. Box 9, Observatory 7935, Cape Town, South Africa

[22]Southern African Large Telescope, P.O. Box 9, Observatory 7935, Cape Town, South Africa

[23]A&A, Department of Physics, Faculty of Sciences, University of Antananarivo, B.P. 906, Antananarivo 101, Madagascar

[24]National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801, USA

[25]School of Astronomy, Institute for Research in Fundamental Sciences, P.O. Box 131, Tehran, Iran

[26]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

APPENDIX A: STRUCTURE OF THE LEVEL-1 SOURCE CATALOGUE

The Level-1 Early Science catalogues for COSMOS and XMM-LSS contain a subset of the columns from the raw PYBDSF output (the Level-0) catalogues, as well as some additional derivative columns. In addition they have been filtered for artefacts and incomplete sources that are cut off by the primary beam correction and mosaicking (see Section 3.3 for full details). Below we describe the set of columns that form the Level-1 catalogues. The first ten rows of the COSMOS catalogue are presented in Table A1.

(0): An IAU-style identifier of the form JHHMMSS.SS+/-DDMMSS.S.

(1), (2): The J2000 Right Ascension of the component in degrees, and its associated 1σ positional uncertainty.

(3), (4): The J2000 Declination of the component in degrees, and its associated 1σ positional uncertainty. Note that columns (2) and (4) are the statistical uncertainties derived from the component fitting by PYBDSF and thus do not include astrometric errors. Note also that columns (1) and (3) have not been corrected

for any systematic offsets. Please refer to Section 4.1 for further details.

(5), (6): The integrated flux density and its associated 1σ uncertainty in Jy.

(7), (8): The peak brightness and its associated 1σ uncertainty in Jy beam^{-1}.

(9): The effective frequency in Hz at which the source is observed assuming a spectral index of -0.7 (see Section 3.2 for details).

(10), (11): The integrated flux density and its associated 1σ uncertainty in Jy, corrected to 1.4 GHz assuming the effective frequency given in column **(9)** and a spectral index of -0.7.

(12), (13): The peak brightness and its associated 1σ uncertainty in Jy beam^{-1}, corrected to 1.4 GHz assuming the effective frequency given in column **(9)** and a spectral index of -0.7.

(14), (15): The major axis and associated 1σ uncertainty of the 2D Gaussian fitted to the source by PYBDSF, in degrees.

(16), (17): The minor axis and associated 1σ uncertainty of the 2D Gaussian fitted to the source by PYBDSF, in degrees.

(18), (19): The position angle (measured east of north) and associated 1σ uncertainty of the 2D Gaussian fitted to the source by PYBDSF, in degrees.

(20), (21): The major axis and associated 1σ uncertainty of the deconvolved source size, in degrees, evaluated according to Equations 4 and 5, see Section 3.3.3 for details. A zero in these columns means that the source was fitted with an unphysical size by PYBDSF and is assumed to be unresolved along the major axis.

(22), (23): The minor axis and associated 1σ uncertainty of the deconvolved source size, in degrees, evaluated as per columns **(19)** and **(20)**. A zero in these columns means that the source was fitted with an unphysical size by PYBDSF and is assumed to be unresolved along the minor axis.

(24): A Boolean flag indicating if the source is reliably resolved, i.e. if it satisfies the criterion given in Equation 6.

(25): The estimate from PYBDSF of the background RMS around the component, in Jy beam^{-1}.

(26): A unique integer identifier for the Gaussian component from the raw PYBDSF output.

(27): A unique integer identifier for the source from the raw PYBDSF output.

(28): A unique integer identifier for the island from the raw PYBDSF output. Columns 25, 26 and 27 are provided in the Level 1 catalogues to preserve the source finder's original groupings of Gaussian components into sources.

APPENDIX B: ADDITIONAL OBSERVATIONS IN COSMOS

Table B1 lists the properties of the 14 additional COSMOS observations (totalling 111 h) that we made use of for the analysis in Section 4.5. The table columns are the same as Table 1, and the

Table A1. The first ten rows from the COSMOS Level 1 Early Science catalogue. Please refer to the text in Appendix A for full column descriptions.

	Name (0)	RA (deg) (1)	$\sigma_{\rm RA}$ (deg) (2)	Dec (deg) (3)	$\sigma_{\rm Dec}$ (deg) (4)	$S_{\rm int}$ (Jy) (5)	$\sigma_{\rm local}$ (Jy) (6)	$S_{\rm peak}$ (Jy beam^{-1}) (7)	$\sigma_{\rm peak}$ (Jy beam^{-1}) (8)
1	J100320.73+020931.7	150.83641	0.00029	2.15882	0.00030	0.0001038	0.0000289	0.0000555	0.0000107
2	J100320.26+021312.7	150.83442	0.00003	2.22022	0.00008	0.0003704	0.0000235	0.0002480	0.0000101
3	J100319.37+021447.2	150.83073	0.00011	2.24645	0.00013	0.0000749	0.0000159	0.0000785	0.0000095
4	J100318.49+021000.5	150.82705	0.00016	2.16682	0.00021	0.0000608	0.0000179	0.0000582	0.0000099
5	J100317.73+021606.7	150.82388	0.00007	2.26855	0.00008	0.0001638	0.0000186	0.0001419	0.0000097
6	J100317.32+022016.0	150.82219	0.00007	2.33780	0.00009	0.0001212	0.0000163	0.0001200	0.0000093
7	J100316.70+020457.0	150.81962	0.00012	2.08250	0.00024	0.0000833	0.0000203	0.0000683	0.0000100
8	J100316.52+021207.5	150.81886	0.00011	2.20209	0.00014	0.0001001	0.0000189	0.0000876	0.0000099
9	J100316.52+021810.2	150.81884	0.00011	2.30284	0.00016	0.0000695	0.0000157	0.0000708	0.0000091
10	J100316.10+021700.1	150.81712	0.00003	2.28339	0.00003	0.0003463	0.0000160	0.0003498	0.0000093

	$\nu_{\rm eff}$ (Hz) (9)	$S_{\rm int}^{1.4{\rm GHz}}$ (Jy) (10)	$\sigma_{\rm int}^{1.4{\rm GHz}}$ (Jy) (11)	$S_{\rm peak}^{1.4{\rm GHz}}$ (Jy beam^{-1}) (12)	$\sigma_{\rm peak}^{1.4{\rm GHz}}$ (Jy beam^{-1}) (13)	ϕ_M (deg) (14)	σ_ϕ^M (deg) (15)	ϕ_m (deg) (16)	σ_ϕ^m (deg) (17)
1	1304066432	0.0000987	0.0000275	0.0000528	0.0000101	0.00397	0.00087	0.00269	0.00045
2	1304924800	0.0003526	0.0000224	0.0002361	0.0000096	0.00364	0.00018	0.00234	0.00008
3	1305698816	0.0000713	0.0000152	0.0000748	0.0000090	0.00246	0.00031	0.00222	0.00025
4	1306541312	0.0000579	0.0000170	0.0000555	0.0000094	0.00276	0.00053	0.00216	0.00033
5	1306921600	0.0001561	0.0000177	0.0001352	0.0000093	0.00261	0.00018	0.00253	0.00017
6	1305002624	0.0001154	0.0000155	0.0001143	0.0000089	0.00256	0.00021	0.00225	0.00016
7	1306006144	0.0000794	0.0000193	0.0000651	0.0000095	0.00325	0.00058	0.00214	0.00026
8	1308797184	0.0000955	0.0000180	0.0000835	0.0000095	0.00279	0.00034	0.00234	0.00024
9	1307299840	0.0000662	0.0000150	0.0000675	0.0000087	0.00267	0.00039	0.00210	0.00024
10	1308201600	0.0003303	0.0000152	0.0003336	0.0000089	0.00238	0.00006	0.00237	0.00006

	$\phi_{\rm PA}$ (deg) (18)	$\sigma_\phi^{\rm PA}$ (deg) (19)	θ_M (deg) (20)	σ_θ^M (deg) (21)	θ_m (deg) (22)	σ_θ^m (deg) (23)	Resolved (24)	$\sigma_{\rm Isl}$ (Jy beam^{-1}) (25)	$\rm ID_{Gaus}$ (26)	$\rm ID_{Src}$ (27)	$\rm ID_{Isl}$ (28)
1	137.6	25.1	0.00317	0.00109	0.00124	0.00098	0	0.0000100	0	0	0
2	179.6	4.7	0.00274	0.00023	0.00000	0.00000	1	0.0000097	1	1	1
3	178.8	50.2	0.00057	0.00135	0.00000	0.00000	0	0.0000096	2	3	2
4	153.7	31.5	0.00138	0.00106	0.00000	0.00000	0	0.0000099	3	4	3
5	178.8	89.3	0.00104	0.00045	0.00082	0.00053	0	0.0000094	4	5	4
6	12.7	25.8	0.00092	0.00059	0.00000	0.00000	0	0.0000093	5	6	5
7	165.1	17.3	0.00221	0.00085	0.00000	0.00000	0	0.0000099	6	7	6
8	20.6	28.3	0.00144	0.00066	0.00000	0.00000	0	0.0000097	7	8	7
9	167.3	24.1	0.00119	0.00087	0.00000	0.00000	0	0.0000092	8	9	8
10	10.9	0.0	0.00000	0.00000	0.00000	0.00000	0	0.0000093	9	10	9

data were also processed using the methods described in Section 2.

APPENDIX C: UPDATED MOSAIC LAYOUTS

The footprint of the MIGHTEE survey is principally determined by those of existing and planned optical and near-infrared surveys in the four target fields. A secondary consideration is the placement of the pointing centres of the radio observations within these regions. The pointing strategy has been modified since the layout presented in Figure 4 by Jarvis et al. (2016), with the above considerations being joined by our experience with the first batch of MeerKAT observations in each of the four fields. The modifications aim to increase the uniformity of the final survey products, and improve the off-axis polarimetry, mainly by increasing the density of pointings with a corresponding reduction on the time spent on each pointing.

For XMM-LSS and E-CDFS most of the area is covered by a standard close-packed hexagonal mosaic pattern, which is known to efficiently deliver uniform coverage (e.g. Condon et al. 1998). In these two fields each pointing is typically covered by either 2 × 4

h track, or a single 8 h track (both including overheads), however there will be additional depth in certain areas due to the addition of the initial batch of pointings that were taken as part of the commissioning and Early Science phases. Such additional pointings are marked on the relevant figures. COSMOS and ELAIS-S1 have smaller areas but go deeper due to either having a closer packed mosaic, having 2 × 8 h tracks on some pointings, or both. The pointing centres for COSMOS are not provided in this appendix as they are already presented in this paper via Table B1, and shown in Figure 1.

Figure C1 shows the proposed final survey coverage for (left to right) E-CDFS, ELAIS-S1, and XMM-LSS. The blue rectangles show the coverage of the VIDEO (Jarvis et al. 2013) or VEILS (Hönig et al. 2017) infrared surveys, and are generally representative of the regions that contain the deepest ancillary data from numerous facilities. The labelled pointing centres represent the positions of the final pointing grid, with the circles showing the nominal half-power point of the MeerKAT L-band primary beam at 1284 MHz. The pink markers represent the positions of commissioning or Early Science pointings, the latter featuring in this article, all of which will be used for the final survey products.

Note that for E-CDFS the MIGHTEE pointing grid is snapped

Table B1. Details of the additional pointings in the COSMOS field, as plotted in Figure 1, and discussed in Section 4.5.

Date	ID	Field	RA	Dec	Track (h)	On-source (h)	N_{chan}	N_{ant}	Primary	Secondary
2019-07-16	1563267356	COSMOS_1	$09^h59^m46^s$	$+02°01'44''6$	7	6.33	4096	59	J0408-6545	3C237
2019-07-27	1564215117	COSMOS_2	$09^h59^m46^s$	$+02°22'57''4$	7.95	6.98	4096	61	J0408-6545	3C237
2019-07-28	1564301832	COSMOS_3	$10^h01^m11^s$	$+02°01'44''6$	7.96	6.97	4096	60	J0408-6545	3C237
2019-08-16	1565939836	COSMOS_4	$10^h01^m11^s$	$+02°22'57''4$	7.99	6.97	4096	58	J0408-6545	3C237
2020-03-28	1585413022	COSMOS_5	$09^h59^m04^s$	$+02°12'21''0$	8	6.25	32768	59	J0408-6545	3C237
2020-03-29	1585498873	COSMOS_6	$10^h01^m54^s$	$+02°12'21''0$	8	6.25	32768	59	J0408-6545	3C237
2020-03-31	1585671638	COSMOS_7	$10^h00^m29^s$	$+01°51'08''2$	8	6.25	32768	60	J0408-6545	3C237
2020-04-02	1585844155	COSMOS_8	$10^h00^m29^s$	$+02°33'33''8$	8	6.25	32768	60	J0408-6545	3C237
2020-04-30	1585928757	COSMOS_9	$10^h01^m54^s$	$+02°33'33''8$	8	6.25	32768	60	J0408-6545	3C237
2020-04-04	1586016787	COSMOS_10	$09^h59^m04^s$	$+02°33'33''8$	8.03	6.25	32768	60	J0408-6545	3C237
2020-04-06	1586188138	COSMOS_11	$09^h58^m21^s$	$+02°22'57''4$	8	6.25	32768	59	J0408-6545	3C237
2020-04-07	1586274966	COSMOS_12	$09^h58^m21^s$	$+02°01'44''6$	8	6.25	32768	60	J0408-6545	3C237
2020-04-12	1586705155	COSMOS_13	$09^h59^m04^s$	$+01°51'08''2$	8	6.25	32768	59	J0408-6545	3C237
2020-04-13	1586791316	COSMOS_14	$10^h01^m53^s$	$+01°51'08''2$	8	6.25	32768	60	J0408-6545	3C237
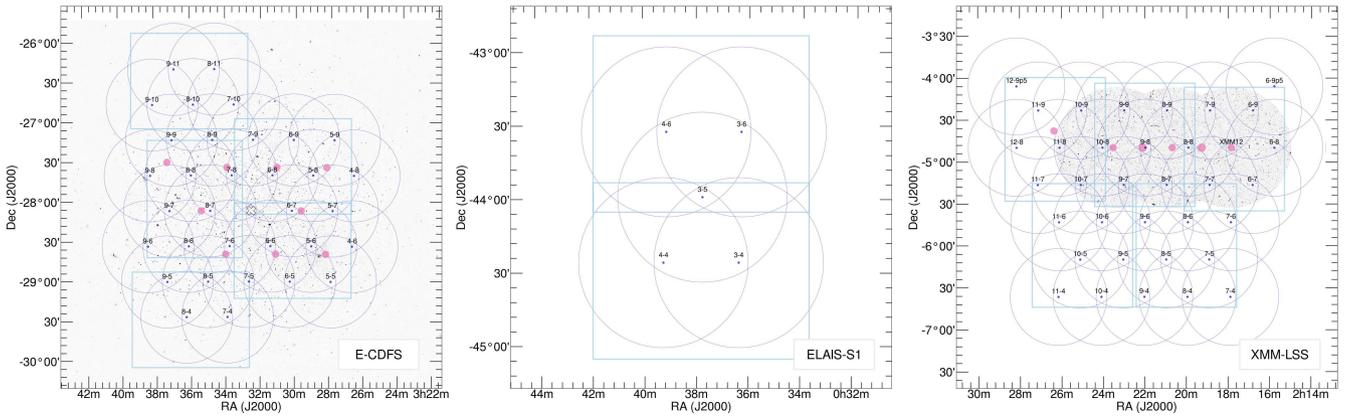
Figure C1. The final pointing layouts (and thus eventual sky coverage) for the MIGHTEE observations of E-CDFS, ELAIS-S1 and XMM-LSS. Please refer to Appendix C for further details, and Figure 1 and Appendix B for the COSMOS layout.

to the position of the pointing centre for the ultra-deep LADUMA survey (Baker et al. 2018), marked on Figure C1 by the central cross marker. The continuum imaging from LADUMA will be included in the final MIGHTEE mosaic, and since most of LADUMA's observations will be conducted with MeerKAT's UHF receivers, there will also be a lower frequency ($\nu_{centre} = 860$ MHz) continuum image for that central pointing available for combination with the L-band MIGHTEE data. The background image of the E-CDFS panel on Figure C1 shows the continuum image from a single LADUMA UHF track, without any primary beam correction applied, demonstrating the ~ 2 degree field of view of the UHF imaging (MIGHTEE collaboration, *priv. comm.*). The XMM-LSS Early Science image forms the background to the lower panel of Figure C1, showing the signficant increase in sky area (and cosmological volume) that the full survey will provide. Note that the XMM-LSS pointings are snapped to the original XMMLSS_12 pointing (see Table 1), which has already achieved full depth with the 32,768 channel mode necessary for the spectral line component of MIGHTEE. For planning purposes the pointing centres of the final MIGHTEE survey are provided online as supplementary material.

This paper has been typeset from a TEX/LATEX file prepared by the author.