

# The C-Band All-Sky Survey: total intensity point-source detection over the northern sky

R. D. P. Grumitt <sup>1,★</sup>, Angela C. Taylor,<sup>1</sup> Luke Jew <sup>1</sup>, Michael E. Jones <sup>1</sup>,  
C. Dickinson <sup>2,3</sup>, A. Barr,<sup>2</sup> R. Cepeda-Arroita <sup>2</sup>, H. C. Chiang,<sup>4</sup> S. E. Harper,<sup>2</sup>  
H. M. Heilgendorff,<sup>5</sup> J. L. Jonas,<sup>6,7</sup> J. P. Leahy,<sup>2</sup> J. Leech,<sup>1</sup> T. J. Pearson <sup>3</sup>,  
M. W. Peel <sup>8,9</sup>, A. C. S. Readhead<sup>3</sup> and J. Sievers <sup>4</sup>

<sup>1</sup>Sub-department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

<sup>2</sup>Jodrell Bank Centre for Astrophysics, Alan Turing Building, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

<sup>3</sup>Cahill Centre for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA

<sup>4</sup>Department of Physics, McGill University, 3600 Rue University, Montréal QC H3A 2T8, Canada

<sup>5</sup>Astrophysics & Cosmology Research Unit, School of Mathematics, Statistics & Computer Science, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa

<sup>6</sup>Department of Physics and Electronics, Rhodes University, Grahamstown 6139, South Africa

<sup>7</sup>South African Radio Astronomy Observatory, 2 Fir Road, Observatory, Cape Town 7925, South Africa

<sup>8</sup>Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

<sup>9</sup>Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206 La Laguna, Tenerife, Spain

Accepted 2020 May 29. Received 2020 May 28; in original form 2019 October 18

## ABSTRACT

We present a point-source detection algorithm that employs the second-order Spherical Mexican Hat wavelet filter (SMHW2), and use it on C-Band All-Sky Survey (C-BASS) northern intensity data to produce a catalogue of point sources. This catalogue allows us to cross-check the C-BASS flux-density scale against existing source surveys, and provides the basis for a source mask that will be used in subsequent C-BASS and cosmic microwave background (CMB) analyses. The SMHW2 allows us to filter the entire sky at once, avoiding complications from edge effects arising when filtering small sky patches. The algorithm is validated against a set of Monte Carlo simulations, consisting of diffuse emission, instrumental noise, and various point-source populations. The simulated source populations are successfully recovered. The SMHW2 detection algorithm is used to produce a 4.76 GHz northern sky source catalogue in total intensity, containing 1784 sources and covering declinations  $\delta \geq -10^\circ$ . The C-BASS catalogue is matched with the Green Bank 6 cm (GB6) and Parkes-MIT-NRAO (PMN) catalogues over their areas of common sky coverage. From this we estimate the 90 per cent completeness level to be approximately 610 mJy, with a corresponding reliability of 98 per cent, when masking the brightest 30 per cent of the diffuse emission in the C-BASS northern sky map. We find the C-BASS and GB6 flux-density scales to be consistent with one another to within approximately 4 per cent.

**Key words:** catalogues – surveys – methods: data analysis – cosmic background radiation – radio continuum: general – cosmology: observations.

## 1 INTRODUCTION

C-Band All-Sky Survey (C-BASS) is an experiment to observe the whole sky in total intensity and polarization at 4.76 GHz and at 45 arcmin angular resolution (Jones et al. 2018). The primary purpose of the experiment is to provide data on synchrotron emission

for cosmic microwave background (CMB) polarization experiments at low frequencies compared to the peak of the CMB, whilst avoiding significant de-polarization due to Faraday rotation. C-BASS will also provide improved measurements of low-frequency emission components, enabling a detailed study of the Galactic Magnetic Field. The northern sky survey has now been completed, with detailed analysis of these data currently underway. In this paper, we present a 4.76 GHz northern sky point-source catalogue, produced using a point-source detection algorithm employing the SMHW2.

\* E-mail: richard.grumitt@physics.ox.ac.uk

Extragalactic radio sources are a key contaminant in CMB studies, with their detection and removal being an important step in CMB component separation (Brandt et al. 1994; Taylor et al. 2001; Waldram et al. 2003, 2007; de Zotti et al. 2010). For future CMB experiments, it has been shown that point-source contamination has the potential to significantly complicate measurements of the  $B$ -mode polarization power spectrum on scales  $\ell \gtrsim 50$  (Mesa et al. 2002; Curto et al. 2013; Puglisi et al. 2018; Remazeilles et al. 2018). Given the large beam of C-BASS, the instrument is not well suited to point-source detection, with detections of all but the brightest sources ( $\gtrsim 1$  Jy) being limited by source confusion and diffuse emission. For the C-BASS analysis, performing our own dedicated source detection allows us to carry out important data quality checks on the C-BASS flux-density scale and pointing accuracy, by comparing the C-BASS catalogue with pre-existing catalogues around 4.76 GHz. The resulting catalogue will also allow us to construct an accurate mask of bright source emission in the C-BASS maps, which will be key to avoid biasing CMB component separation analyses with C-BASS. Fainter sources can either be subtracted using deeper, pre-existing source catalogues, or can be treated statistically at the power spectrum level. Assuming sources down to some flux density  $S_{\max}$  have been removed, and that sources are Poisson-distributed on the sky, the point-source contribution to the power spectrum is given by

$$C_{\ell}^{\text{PS}} = \int_0^{S_{\max}} \frac{dN}{dS} S^2 dS, \quad (1)$$

where  $S$  is the flux density and  $dN/dS$  is the differential source count, i.e. the number of sources per unit flux density, per unit steradian (Tegmark & Efstathiou 1996).

Numerous source-detection algorithms have been developed for CMB experiments (see e.g. Tegmark & de Oliveira-Costa 1998; Vielva et al. 2003; González-Nuevo et al. 2006; López-Cañiego et al. 2006; Argüeso et al. 2009; Herranz et al. 2009; Carvalho et al. 2012; Bennett et al. 2013; Planck Collaboration XXVIII 2014; Planck Collaboration XXVI 2016c). The Planck Catalogue of compact sources (PCCS) used an algorithm based on filtering small patches of sky using the flat-sky second-order Mexican Hat wavelet as an approximation to a matched filter, followed by threshold detection based on the signal-to-noise ratio (SNR). Here, we employ a similar scheme but using the equivalent spherical function on the whole sky at once, implemented in the Healpix pixelization scheme (Górski et al. 2005). Previous applications of the first-order SMHW (SMHW1) to searches for non-Gaussianity can be found in Cayón et al. (2001), Martínez-González et al. (2002), Curto, Martínez-González & Barreiro (2011), and to point-source detection in Vielva et al. (2003). The details of this implementation are discussed in Section 3.

The outline of this paper is as follows. In Section 2, we summarize pre-existing source catalogues directly relevant to our C-BASS analysis. In Section 3, we discuss the C-BASS point-source detection algorithm, giving an overview of the method and the Monte Carlo simulations used to validate the algorithm. In Section 4, we discuss the C-BASS total intensity, northern sky point-source catalogue obtained using our detection algorithm. We compare the C-BASS catalogue with the GB6 and PMN catalogues and calculate the differential source counts for the C-BASS sources as a cross-check on the statistical properties of the bright source population. We summarize our results in Section 5. Whilst analysis of polarized sources is important for future CMB polarization studies, this falls beyond the scope of this paper where we focus on the C-BASS total intensity results. In addition, due to the low level of source polarization, only a small number of polarized sources ( $\mathcal{O}(10)$ ) will be detected.

## 2 PRE-EXISTING SOURCE CATALOGUES

In order to construct an accurate template for masking out the point sources in the C-BASS maps, it is necessary to construct a point-source catalogue based on the C-BASS observations themselves and any useful information that can be gleaned from other more sensitive, higher resolution surveys. Relevant to our work in this paper are the GB6 (Gregory et al. 1996), PMN (Griffith & Wright 1993; Griffith et al. 1994, 1995; Wright et al. 1994, 1996a), Effelsberg S5 (Kuehr et al. 1981), RATAN-600 (Mingaliev et al. 2007), and Combined Radio All-Sky Targeted Eight-GHz Survey (CRATES; Healey et al. 2007) catalogues. These catalogues are primarily used in this paper to make comparisons with the C-BASS catalogue, for the purposes of data validation and estimation of the statistical properties of the C-BASS catalogue. However, the catalogues are also useful for characterizing the faint point-source population in any C-BASS analysis (i.e. sources with flux-densities below  $\sim 1$  Jy), where reliable extraction from the C-BASS map becomes more challenging. These radio surveys are summarized in Table 1.

The GB6 and PMN 4.85 GHz source catalogues cover declinations  $-87.5^\circ \leq \delta \leq 75^\circ$ . The GB6 catalogue was produced using the NRAO seven-beam receiver on the 91 m telescope, and the PMN catalogue was produced using the Parkes 64 m radio telescope. The GB6 catalogue has a flux-density limit of approximately 18 mJy, whilst the PMN catalogue has an average flux-density limit of approximately 35 mJy over the sky. The GB6 and PMN catalogues provide far deeper flux-density coverage than can be achieved with C-BASS, given the differing resolutions and hence confusion levels. However, it is still necessary for us to obtain our own source catalogue, such that we can construct an accurate mask for the brightest sources in the C-BASS maps, and account for source variability between the C-BASS and GB6/PMN surveys. The GB6 and PMN catalogues remain useful in accounting for fainter sources (below  $\sim 1$  Jy) in any C-BASS analysis.

Source catalogues covering the North Celestial Pole (NCP) region are more limited, with GB6 only covering declinations up to  $\delta = 75^\circ$ . The Effelsberg S5 catalogue covers declinations  $\delta \geq 70^\circ$  and is complete down to 250 mJy (Kuehr et al. 1981). In comparing to the S5 catalogue, there are significant issues from the variability of flat-spectrum sources, given the large separation in time between the S5 survey and the C-BASS survey. The RATAN-600 catalogue includes measurements at 4.8 GHz, and observed 504 sources in the NCP region with NRAO VLA Sky Survey (NVSS) flux-densities,  $S_{1.4\text{GHz}} \geq 200\text{mJy}$  (Mingaliev et al. 2001, 2007). The RATAN-600 catalogue was used in a previous analysis of diffuse emission in the NCP region with C-BASS in Dickinson et al. (2019). Whilst this catalogue provides deeper flux-density coverage than C-BASS, it was produced by pre-selecting sources for study from the NVSS catalogue at 1.4 GHz (Condon et al. 1998). This can potentially miss rising-spectrum sources that would otherwise be observable in the C-BASS catalogue. Given the C-BASS flux-density limit of approximately 500 mJy, we can use the NVSS source counts and the 1.4–4.85 GHz spectral index distributions in Tucci et al. (2011) to estimate that there may be  $\mathcal{O}(1)$  sources that could be observed by C-BASS, whilst being missed by RATAN-600.

It is also worth noting the CRATES catalogue (Healey et al. 2007). CRATES is an 8.4 GHz catalogue of flat-spectrum sources with flux-densities  $S_{\nu=4.85\text{GHz}} > 65\text{mJy}$ , covering Galactic latitudes  $|b| > 10^\circ$ . The catalogue therefore serves as a useful proxy for flat-spectrum sources in the C-BASS catalogue, which are the primary contributor to source variability. Healey et al. (2009) made additional observations of the NCP region at declinations,  $\delta >$

**Table 1.** Summary of radio surveys relevant to the work in this paper. In stating the sky coverage,  $\delta$  denotes declination, and  $b$  denotes Galactic latitude.

Survey name	Reference	Frequency (GHz)	Sky Coverage	FWHM (arcmin)	Flux Limit (mJy)	Number of sources
C-BASS	Jones et al. (2018)	4.76	$-90^\circ \leq \delta \leq 90^\circ$ <sup>a</sup>	45	$\sim 500$ <sup>b</sup>	1784
GB6	Gregory et al. (1996)	4.85	$0^\circ \leq \delta \leq 75^\circ$	3	$\sim 18$	75,162
PMN	Wright et al. (1996b)	4.85	$-87.5^\circ \leq \delta \leq 10^\circ$	5	$\sim 35$	50,814
Effelsberg S5	Kuehr et al. (1981)	4.9	$70^\circ \leq \delta \leq 90^\circ$	2.7	$\sim 250$	476
RATAN-600	Mingaliev et al. (2007)	4.8 (1.1 – 21.7) <sup>c</sup>	$75^\circ \leq \delta \leq 88^\circ$	$0.67 \times 6.6$ <sup>d</sup>	(200) <sup>e</sup>	504
CRATES	Healey et al. (2007)	8.4 (4.85) <sup>f</sup>	$ b  > 10^\circ$	2.4	$\sim 65$	11,131

<sup>a</sup>Here, we state the sky coverage of the whole C-BASS experiment. The results concerning point sources in this paper were obtained for the C-BASS northern intensity data, covering declinations  $-10^\circ \leq \delta \leq 90^\circ$ .

<sup>b</sup>For C-BASS, we state the value  $3.5\langle\sigma\rangle$ , where  $\langle\sigma\rangle$  is the mean background fluctuation level found across the map, applying the CG30 mask described in Section 3.5. Details on the estimation of background fluctuation levels are given in Section 3.1.

<sup>c</sup>The RATAN-600 catalogue covers six frequencies from 1.1 to 21.7 GHz, including 4.8 GHz.

<sup>d</sup>We state  $\text{FWHM}_{\text{RA}} \times \text{FWHM}_\delta$  for RATAN-600, determined from the values given in Kovalev et al. (1999).

<sup>e</sup>The RATAN-600 catalogue was produced by pre-selecting NVSS sources with  $S_{1.4\text{GHz}} > 200$  mJy.

<sup>f</sup>The CRATES catalogue is primarily at 8.4 GHz, with 4.85 GHz sources used as the basis for observations of 8.4 GHz counterparts. The properties of these 4.85 GHz sources are provided with the CRATES catalogue. Further observations were also made at 4.85 GHz to fill in gaps at  $\delta > 88^\circ$ .

$88^\circ$  to supplement the original CRATES catalogue. The purpose of this was to bring the flux-density limit in this region down to the CRATES flux-density limit of  $\sim 65$  mJy. Three sources were observed in this region at 4.85 GHz, with flux densities of 67, 58, and 142 mJy.

### 3 C-BASS SOURCE DETECTION ALGORITHM

To detect sources in a sky map we need to remove obscuring diffuse emission and noise. For a source with a known point spread function (PSF) embedded in additive noise, the matched filter (MF) is the optimal filter that can be applied to maximize the source SNR. The matched filter is given by

$$\Psi_{\text{MF}}(k) = \left[ 2\pi \int dk k \frac{\tau^2(k)}{P(k)} \right]^{-1} \frac{\tau(k)}{P(k)}, \quad (2)$$

where  $k$  is the Fourier wavenumber,  $\tau(k)$  is the source profile, and  $P(k)$  is the power spectrum of the unfiltered map (Tegmark & de Oliveira-Costa 1998; López-Cañiego et al. 2006). However, the calculation of the MF involves a number of complications. Chiefly, we are required to make a noisy estimate of the power spectrum from our unfiltered map, and integrate it. In constructing the PCCS, it was found that the Mexican Hat wavelet of the second kind (MHW2) achieved similar performance to the MF (López-Cañiego et al. 2006; Planck Collaboration XXVIII 2014; Planck Collaboration XXVI 2016c). For the present analysis, we adapt the *Planck* algorithm, using instead the SMHW2 in place of the flat-space MHW2. The SMHW2 is straightforward to calculate, enables us to filter the entire sky at once, and allows us to optimize a few free parameters as opposed to a noisy estimate of the full noise power spectrum.

#### 3.1 Source detection algorithm

Given a sky map consisting of point sources, diffuse emission and instrumental noise, we can enhance the SNR of point sources in the map by filtering with the SMHW2. We perform this filtering at a range of filter scales  $R$ , to maximize the number of sources we extract from the sky maps. In changing  $R$ , we effectively change the extent to which we down-weight large-scale  $\ell$ -modes that are dominated by diffuse emission, and also small scales dominated by instrumental noise. This is particularly important in regions close to the Galactic plane where diffuse emission is very strong, meaning

we must down-weight large-scale modes harshly in order to extract point sources. Conversely, in regions with little diffuse emission we may wish to be less extreme in our down-weighting, so that we do not excessively reduce point-source power and subsequently miss detection of fainter sources in these regions. We discuss the form of the SMHW2 filter and the effect of the filter scale,  $R$  in detail in Section 3.2.

The source-detection algorithm is described below, with the algorithm flowchart displayed in Fig. 1.

(i) We begin by taking the spherical harmonic transform of our sky map, obtaining the  $a_{\ell m}$  coefficients. We filter the map by weighting the  $a_{\ell m}$  coefficients with the SMHW2 window function,  $w_\ell^{S^2}(R)$  at some user-defined set of filter scales, i.e. we calculate

$$f_{\ell m}(R) = a_{\ell m} w_\ell^{S^2}(R). \quad (3)$$

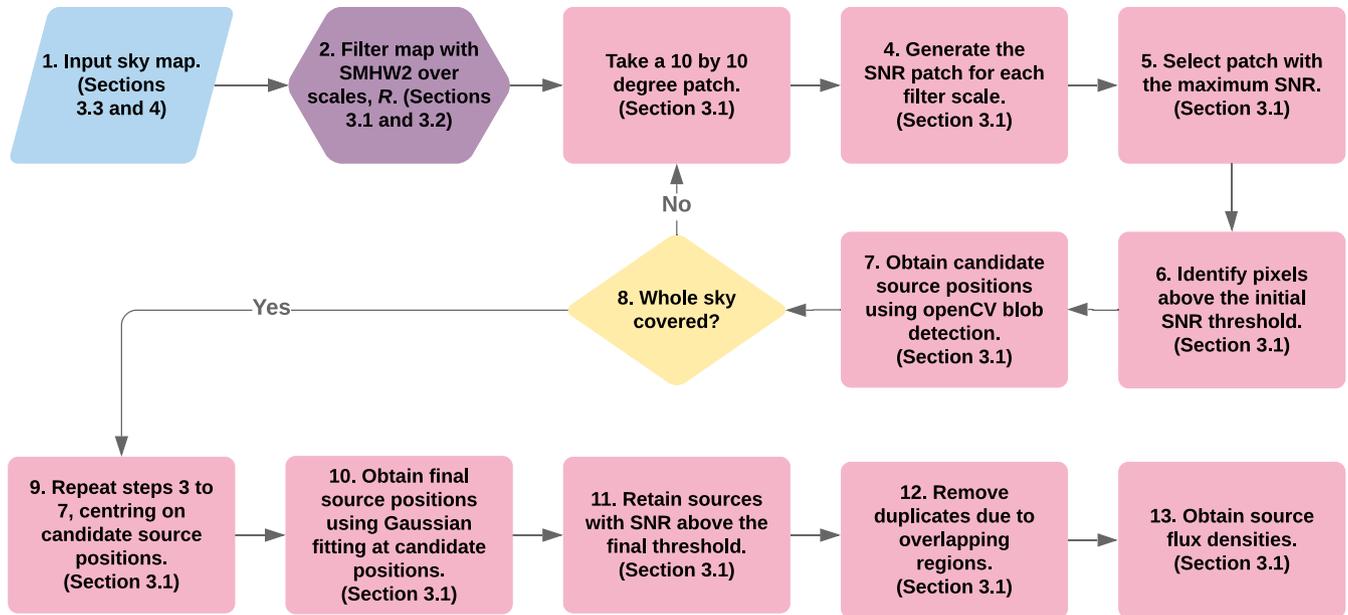
The effect of this filtering is demonstrated in Fig. 2. Here, we display a simulated sky consisting of diffuse emission, instrumental noise and point-sources, along with the corresponding filtered sky with scale  $R = 1$ . The details of the production of this simulated sky are discussed in Section 3.3. The filter has removed diffuse emission and enhanced the SNR of point-sources across the sky.

(ii) We break the maps, filtered at a range of  $R$  scales, into overlapping  $10^\circ \times 10^\circ$  patches on the sky and calculate the corresponding SNR patches. We used  $R$  scales in the range  $0.5 \leq R \leq 2$  in steps of 0.1. We choose  $10^\circ \times 10^\circ$  patches such that the flat-sky approximation holds, whilst capturing a sufficient sample of the sky background in the given sky region. We set the overlap to  $5^\circ$  in each direction to ensure that we obtain accurate positions for sources detected close to patch edges. The typical background fluctuation level in each patch is estimated by calculating the median absolute deviation (MAD) of the pixels, defined as

$$\text{MAD}(X) = \text{median}(|X_i - \text{median}(X)|), \quad (4)$$

where  $X$  is the set of pixel values. The MAD estimator avoids sensitivity to outlier pixel values, which can be caused by the presence of point sources in the sky patch. For each sky patch, we then select the filter scale that gives the maximal SNR for the patch i.e. the maximum of the peak pixel value divided by the MAD estimate.

(iii) Having selected the value of  $R$  that maximizes the patch SNR, an initial SNR threshold is applied to identify potential point sources. This produces a patch of tagged and un-tagged pixels. The tagging is performed on these small, overlapping regions so as to



**Figure 1.** Flowchart for the C-BASS point-source detection algorithm. The SMHW2 filter is applied over the entire sky, whilst identification of point sources is performed on  $10^\circ \times 10^\circ$  sky regions.

account for the varying background properties across the sky i.e. variations in residual diffuse emission and noise.

(iv) Given our patch of tagged pixels, we then use the `OPENCV Simple Blob Detector`<sup>1</sup> to obtain candidate source locations. The Simple Blob Detector algorithm works by applying thresholds to an input image and determining the centres of connected pixels, or sources. In addition to providing candidate source locations, this software allows for the filtering of detected sources according to their geometric properties i.e. their size, circularity, convexity, and inertia ratio. For our purposes, we set a maximum size limit to help prevent spurious tagging of bright diffuse emission and disallow detections within one beam FWHM of one another. For a detailed discussion of `OPENCV` and the Simple Blob Detector algorithm, we refer the reader to Kaehler & Bradski (2016). The processing steps involved in studying these small sky regions are illustrated in Fig. 3, where we display the typical output from the steps outlined above.

(v) Given a set of tentative source locations obtained over the whole sky, we now repeat steps (ii) and (iii), this time centring on the tentative source locations. In order to refine the source position estimates, we fit an elliptical Gaussian to a 20 arcmin square region of the wavelet-filtered map, centred on the candidate location obtained using `OPENCV`. Whilst the C-BASS beam and SMHW2 filter are not Gaussian, by fitting to this small region around the candidate location the Gaussian fit is sufficient to make the necessary refinements ( $\sim 1$  arcmin) to our estimates of source peak locations.

(vi) A final SNR thresholding is applied by dividing the peak value of the source in the wavelet filtered map by the MAD value of pixels in an annulus centred on the source, with an inner radius of  $3^\circ$  and outer radius of  $5^\circ$ . The annulus was chosen to avoid the first sidelobe in the C-BASS beam, whilst capturing the local background fluctuations around the source. Sources are retained if this new SNR estimate exceeds some final threshold. The SMHW2 is constructed so as to preserve point-source amplitudes after being

convolved with the C-BASS beam. As with the PCCS, we convert these wavelet amplitudes to Jy and report them as auxiliary flux-density (DETFLUX) estimates alongside the primary estimates from aperture photometry described in step (vii). The DETFLUX estimates are discussed in more detail in the Appendix.

(vii) The detected source locations from all of the filtered maps are combined and duplicates are removed from the catalogue, defined here to be any reported source positions that are within the beam FWHM of one another. For matched sources in our internal catalogue, we retain the source with the greatest detected SNR.

(viii) Given the final set of source positions, source flux densities are obtained. As its primary method, the detection pipeline obtains flux densities using aperture photometry, adapting the method from Planck Collaboration XXVIII (2014) and Génova-Santos et al. (2015). We begin by converting the map to  $\text{Jy pix}^{-1}$ , and define an aperture of radius 45 arcmin around the source position, and an annulus of inner radius  $3^\circ$  and outer radius  $5^\circ$ . The observed flux density is then given by

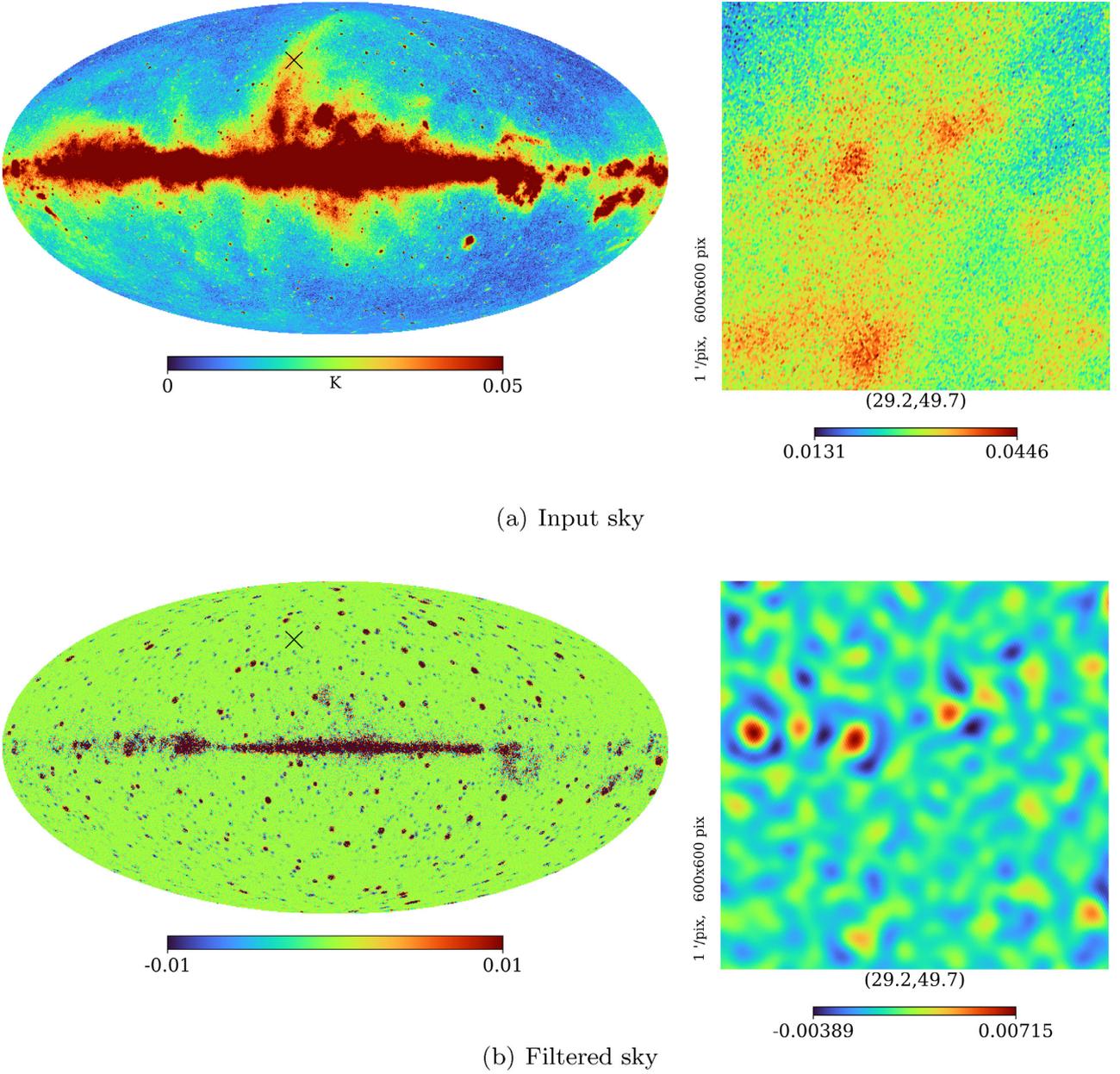
$$S_{\text{obs}} = \kappa (S_{\text{ap}} - \bar{S}_{\text{ann}} N_{\text{ap}}), \quad (5)$$

where  $\kappa \approx 1.34$  is a correction applied to account for flux density missing from the aperture (calculated using equation (A.2) of Planck Collaboration XXVIII 2014),  $S_{\text{ap}}$  is the total flux density in the aperture,  $\bar{S}_{\text{ann}}$  is the median flux density in the annulus and  $N_{\text{ap}}$  is the number of pixels in the aperture. The uncertainty was estimated as

$$\sigma(S_{\text{obs}}) \approx \text{MAD}(X_{\text{ann}}) \frac{N_{\text{ap}}}{\sqrt{N'_{\text{ap}}}}, \quad (6)$$

where  $\text{MAD}(X_{\text{ann}})$  is the MAD value of pixels in the annulus and  $N'_{\text{ap}}$  is the number of beams inside the aperture. The scaling applied to  $\text{MAD}(X_{\text{ann}})$  is to account for correlations in the background emission, approximately on the scale of the beam (Génova-Santos et al. 2015).

<sup>1</sup><https://opencv.org/>



(a) Input sky

(b) Filtered sky

**Figure 2.** Panel (a): An input 4.76 GHz simulated sky consisting of diffuse emission generated using PYSM, instrumental noise generated from a white-noise realization of the C-BASS sensitivity map, along with a point-source population generated by scattering the GB6, PMN, and RATAN-600 sources at random positions over the sky. Alongside this we display a zoom-in centred on the location of the black cross on the map. Panel (b): The same map after being convolved with the SMHW2 filter, removing the large scale diffuse emission and leaving behind a sky filled with point sources. As before, we display a zoom-in of the filtered map alongside it, centred on the location of the black cross. It is important to note here that not all the diffuse emission has been removed at low Galactic latitudes. In regions of strong diffuse emission the source detection algorithm is significantly less reliable. In panel (b), it is apparent that the SMHW2 filtering can still leave a significant sidelobe around bright sources. This can lead to issues with the spurious tagging of sidelobe peaks as sources, particularly for the brightest sources on the sky (i.e. those with flux-densities  $\gtrsim 20$  Jy). This problem can largely be side-stepped by allowing for larger exclusion zones around the brightest sources, as was the case with catalogues such as GB6, removing apparent detections in the first sidelobe of such sources.

### 3.2 The spherical Mexican Hat Wavelet

The Mexican Hat Wavelet (MHW) of the  $n$ th kind is defined in  $\mathbb{R}^2$  (i.e. two-dimensional Euclidean space) as

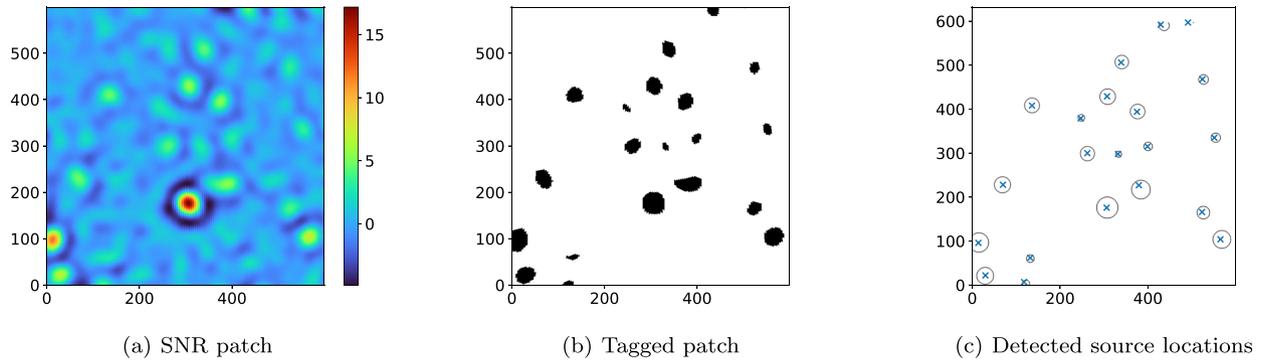
$$\Psi_n(\mathbf{x}, R) \propto \Delta^n \left\{ \exp\left(-\frac{|\mathbf{x}|^2}{2(R\sigma)^2}\right) \right\}, \quad (7)$$

where  $\mathbf{x}$  is the position in the image plane,  $\Delta$  is the usual Euclidean Laplacian operator,  $\sigma$  is the Gaussian standard deviation, and  $R$  is the filter scale factor. In changing  $R$ , we change the characteristic

scale of the filter and hence the extent to which the filter down-weights large- and small-scale modes. Applying this expression we can arrive at the real-space expression for the MHW2, given by

$$\Psi_2(\mathbf{x}, R) \propto (8(R\sigma)^4 - 8(R\sigma)^2|\mathbf{x}|^2 + |\mathbf{x}|^4) \exp\left(-\frac{|\mathbf{x}|^2}{2(R\sigma)^2}\right). \quad (8)$$

In applying the MHW2 to producing the PCCS, the sky was divided up into small patches, taking flat projections and applying



**Figure 3.** Images illustrating the pixel tagging process. Panel (a): An example  $10^\circ \times 10^\circ$  SNR patch, produced by dividing the filtered sky patch by the MAD value of the pixels in the patch. Panel (b): The corresponding patch of tagged and un-tagged pixels. The tagged pixels are shown in *black* and un-tagged pixels in *white*, obtained by applying a threshold of  $\text{SNR} \geq 3.5$  to the SNR patch. Panel (c): An image showing the detected source locations for this patch. We use the OPENCV blob detection algorithm to obtain preliminary source locations from this patch. These blob detections are shown as black circles centred on the preliminary source locations. These position estimates are then refined by fitting an elliptical Gaussian at the preliminary source locations, with the source positions obtained from the Gaussian fit shown as blue crosses.

the MHW2 filter to them. SNR thresholding was applied to the filtered patches to extract the source positions and flux densities were then obtained for the detected sources. Here, we employ the SMHW2, the equivalent of the MHW2 on  $S^2$  (i.e. the two-sphere). For a given filter scale, this allows us to filter the entire sky at once, avoiding complications from edge effects when filtering small patches. In dealing with the C-BASS northern sky map, we apply a  $5^\circ \cos^2$  apodization at the edge of our mask to mitigate edge effects.

In Antoine & Vanderghaynst (1999), it was shown the continuous wavelet transform on  $S^2$  can be constructed by taking the inverse stereographic projection of the  $\mathbb{R}^2$  wavelet. This preserves the properties of the  $\mathbb{R}^2$  wavelet and tends to the MHW in the small-angle limit. The spherical MHW of the  $n$ th kind is given by

$$\Psi_{S^2}^{(n)} \propto \frac{1}{\cos^4 \frac{\theta}{2}} \Psi_n \left( |x| \equiv 2 \tan \frac{\theta}{2}, R \right), \quad (9)$$

where  $\theta$  is the colatitude on the sphere (Cayón et al. 2001; Martínez-González et al. 2002; Vielva et al. 2003; Curto et al. 2011).

Substituting the MHW2 into equation (9), we obtain the SMHW2 as follows:

$$\Psi_{S^2}^{(2)}(\theta, R) \propto \frac{1}{\cos^4 \frac{\theta}{2}} \left[ (R\sigma)^4 - 4(R\sigma)^2 \tan^2 \frac{\theta}{2} + 2 \tan^4 \frac{\theta}{2} \right] \exp \left( -\frac{2 \tan^2 \frac{\theta}{2}}{(R\sigma)^2} \right). \quad (10)$$

We normalize the SMHW2 such that we preserve point-source amplitude after convolving with the filter, i.e.

$$\int_{S^2} \mathcal{B}(\theta) \Psi_{S^2}^{(2)}(\theta, R) \sin \theta \, d\theta \, d\phi = 1, \quad (11)$$

where  $\mathcal{B}(\theta)$  is the C-BASS beam profile.

Given the analytic expression for the SMHW2 in real space on  $S^2$ , we can obtain the SMHW2 window function by calculating the Legendre transform of equation (10), i.e. we calculate

$$w_\ell^{S^2}(R) = 2\pi \int_0^\pi \Psi_{S^2}^{(2)}(\theta, R) P_\ell(\cos \theta) \sin \theta \, d\theta, \quad (12)$$

where  $P_\ell(\cos \theta)$  is the Legendre polynomial of order  $\ell$ .

In Fig. 4, we show the real-space SMHW2 and the corresponding window function, calculated for a  $\sigma$  value corresponding to a beam

FWHM of 45 arcmin, with  $R = \{0.85, 0.95, 1.05, 1.15\}$ . The primary effect of the SMHW2 is to down-weight large scales, which are dominated by diffuse emission. In addition to this, the SMHW2 also acts to down-weight small-scale modes where instrumental noise becomes more significant. In the context of C-BASS, large-scale modes correspond to modes significantly larger than the C-BASS beam scale,  $\ell \lesssim 100$ , whilst small scales correspond to scales significantly smaller than the C-BASS beam scale,  $\ell \gtrsim 1000$ .

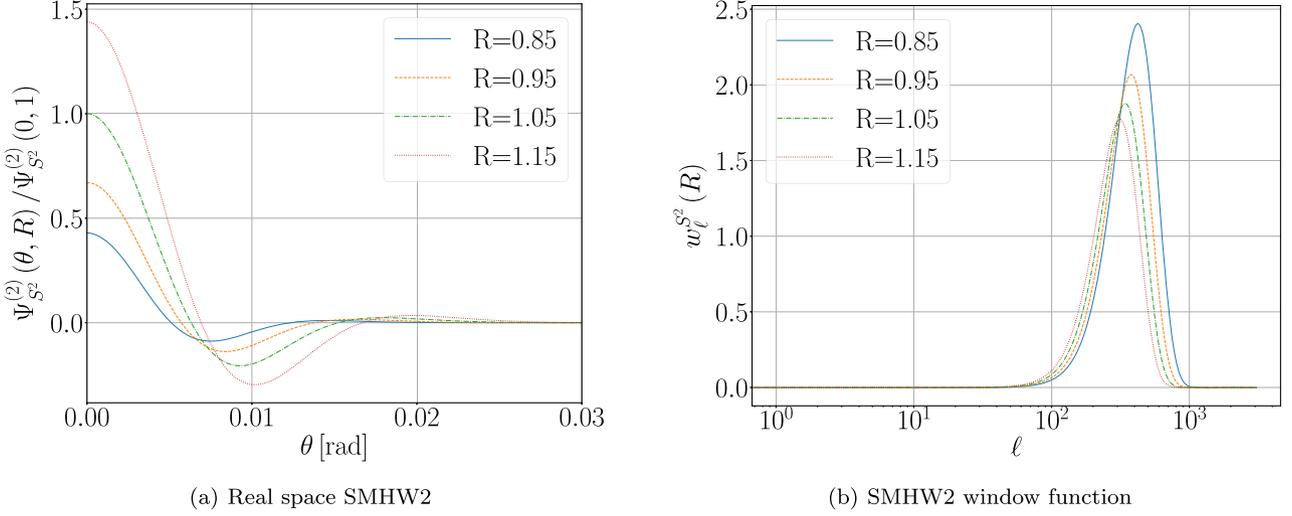
### 3.3 Algorithm validation

We validate the algorithm by running it over 100 simulations of the sky at 4.76 GHz. The simulated skies were generated at NSIDE = 1024, where the number of equal-area pixels on the Healpix sphere is given by  $12 \times \text{NSIDE}^2$ . This high resolution was chosen to allow for the precise determination of source positions, with the C-BASS PSF oversampled by a factor of  $\sim 6.5$ . The simulated sky includes diffuse emission generated by PYSM (Thorne et al. 2017) and an instrumental noise realization. The noise maps were created by generating white noise realizations of the C-BASS sensitivity map (including a simulated sensitivity map for the southern sky survey). This allows us to include the declination dependent effects from the C-BASS scan strategy.

To the diffuse emission simulation we add a source population to the sky. These are created by taking source flux-densities from the PMN, GB6 and RATAN-600 surveys, summarized in Section 2, and scattering the sources at random positions over the sky. Carrying this out 100 times we obtain 100 Monte Carlo simulations of the sky, consisting of the same diffuse background but with their source populations randomly distributed on the sky, and differing noise realizations. The simulated skies are all convolved with the C-BASS beam.

The diffuse components included in the simulation are as follows:

(i) Synchrotron radiation: This was generated using the de-sourced, re-processed 408 MHz Haslam map (Haslam et al. 1981, 1982; Remazeilles et al. 2015) as a template and scaled using a spectral index and curvature model. This corresponds to the PYSM s3 model. Pixel values at 4.76 GHz range from  $\sim 3000 \mu\text{K}$  to  $\sim 8 \times 10^5 \mu\text{K}$ , with a mean value of  $\sim 2 \times 10^4 \mu\text{K}$ .



**Figure 4.** Panel (a): The real space SMHW2, shown for  $\sigma$  corresponding to a Gaussian beam FWHM of 45 arcmin and  $R \in \{0.85, 0.95, 1.05, 1.15\}$ . Panel (b): The corresponding window functions. The SMHW2 down-weights large and small scales, with changes in  $R$  determining the severity with which we down-weight large- and small-scale emission. Whilst the C-BASS beam is not exactly Gaussian, by constructing the SMHW2 around these  $\sim 45$  arcmin scales we are able to filter the large- and small-scale modes around the characteristic beam scale, enhancing point-source SNR in the C-BASS sky map.

(ii) Free-free emission: This was generated with the analytic model used in the *Planck* COMMANDER analysis (Draine 2011; Planck Collaboration IX 2016a; Planck Collaboration X 2016b). The free-free template is scaled using a single power law. This corresponds to the PYSM f1 model. Pixel values at 4.76 GHz range from  $\sim 90 \mu\text{K}$  to  $\sim 5 \times 10^6 \mu\text{K}$ , with a mean value of  $\sim 1 \times 10^4 \mu\text{K}$ .

(iii) Anomalous microwave emission (AME): This was generated by modelling emission caused by the sum of two spinning-dust populations. Scaling is calculated using the SPDUST2 code (Ali-Haïmoud, Hirata & Dickinson 2009). This corresponds to the PYSM a2 model. Pixel values at 4.76 GHz range from  $\sim 5 \mu\text{K}$  to  $\sim 2 \times 10^5 \mu\text{K}$ , with a mean value of  $\sim 300 \mu\text{K}$ .

(iv) Thermal dust: This was generated by modelling the emission caused by a single modified blackbody. This corresponds to the PYSM d1 model (Planck Collaboration X 2016b). Pixel values at 4.76 GHz range from  $\sim 0.01$  to  $\sim 20 \mu\text{K}$ , with a mean value of  $\sim 0.4 \mu\text{K}$ .

(v) CMB: A lensed CMB realization was generated using the Taylens code (Næss & Louis 2013), as part of PYSM. This corresponds to the PYSM c1 model. Pixel values at 4.76 GHz range from  $\sim -300$  to  $\sim 300 \mu\text{K}$ , with a mean value of  $\sim 0.1 \mu\text{K}$ .

One of the simulated skies is shown in Fig. 2, along with the same sky filtered at the scale  $R = 1$ . After running our detection algorithm over the simulated skies, we compare the detected sources to our input catalogues. From this, we calculate the completeness level of our detected sources, averaged over all the simulations, and record the absolute deviations of detected sources from their true positions. The validation results are presented in Section 3.5, produced using an initial SNR threshold of 2.5 to tag candidate sources on the first loop over the sky, and retaining sources with  $\text{SNR} \geq 3.5$  after looping over the candidate sources. These thresholds were selected to give a catalogue reliability  $\gtrsim 90$  per cent away from the Galactic plane. Here, we define catalogue completeness as the fraction of sources above some given flux-density threshold that are recovered by the detection algorithm. The catalogue reliability is defined as

the fraction of detected sources that are matched with at least one source in the input catalogue.

### 3.4 Catalogue matching: likelihood ratios

The simulated validation catalogue is produced for a map with a resolution of 45 arcmin. Given the large number of sources used for our input catalogue, source confusion presents a major complication when trying to match the output catalogues with the corresponding input catalogues from our Monte Carlo simulations. One source in the output catalogue will likely result from the blending of multiple sources in the input catalogue, and a single source in the input catalogue can contribute to more than one source in the output catalogue. We face identical issues when matching the real C-BASS catalogue to higher resolution catalogues such as GB6 and PMN.

To find matches between our input and output catalogues we use a likelihood ratio test. This has previously been used in matching point-source catalogues, particularly when dealing with problems from source confusion (see e.g. Richter 1975; Sutherland & Saunders 1992; Mann et al. 1997; Rutledge et al. 2001; Chapin et al. 2011; Wang et al. 2014). The core of the likelihood ratio test consists in calculating the ratio of the probability of a match between an input and output source being a true match, to the probability of the match being a random association. The likelihood ratio is given by

$$\text{LR} = \frac{q(\Delta S_f) f(r)}{2\pi r \rho(\Delta S_f)}, \quad (13)$$

where  $q(\Delta S_f)$  is the probability density function (PDF) for a given match being a true match as a function of the fractional flux-density difference ( $\Delta S_f$ ) between the two sources,  $f(r)$  is the PDF for true matches as a function of the absolute positional offset between the two sources,  $r$ ,  $2\pi r$  is the PDF for random associations assuming a uniform spatial distribution of background sources, and  $\rho(\Delta S_f)$  is the PDF for random associations as a function of  $\Delta S_f$ .

For the positional PDF of true input-output matches, we assume a Gaussian distribution over the orthogonal coordinate axes. This gives a Rayleigh distribution for the PDF of true input-output

matches as a function of  $r$ , i.e.

$$f(r) = \frac{r}{\sigma_r} \exp\left(-\frac{r^2}{2\sigma_r^2}\right), \quad (14)$$

where  $\sigma_r$  is a scaling parameter. The position and flux-density PDF parameters are estimated by comparing our output source catalogues with the corresponding input catalogues. We determine  $f(r)$  by searching for all candidate matches between the input and output catalogues, using a search radius of 45 arcmin around each output source. Candidate matches will consist of true input–output matches, along with random background associations. Assuming a uniform distribution of background sources the random matches will follow a linear trend, scaling with the sky area enclosed by a given search radius. We therefore fit  $f(r)$  plus a linear trend to the histogram of absolute positional offsets for candidate input–output matches to obtain an estimate for  $f(r)$ .

To determine  $\rho(\Delta S_f)$ , we generate randomized output catalogues by randomly distributing our output sources over the sky and randomizing source flux-densities from the parent distribution. We then find candidate matches between input and randomized output sources, again using a 45 arcmin search radius. We bin the candidate matches as a function of the fractional flux-density difference between the input and output source, given by

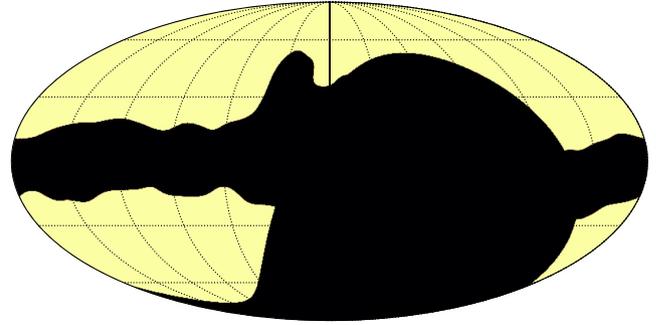
$$\Delta S_f = \frac{S_{\text{out}} - S_{\text{in}}}{S_{\text{in}}}, \quad (15)$$

where  $S_{\text{out}}$  is the output source flux density and  $S_{\text{in}}$  is the input source flux density. This gives us the number of random matches as a function of  $\Delta S_f$  and hence our estimate for  $\rho(\Delta S_f)$ . To obtain  $q(\Delta S_f)$  we carry out the same procedure as above, but with the real output catalogues. We then estimate  $q(\Delta S_f)$  from the difference between the number of matches as a function of  $\Delta S_f$  for the real output catalogues, and the number of matches for the randomized catalogues, i.e. from the excess matches in the real output catalogues.

To determine the LR threshold for declaring a true match, we calculate the likelihood ratios for all candidate matches between the input catalogues and randomized output catalogues. From this, we determine the value of LR at which we would declare 10 per cent of our randomized matches to be true matches. This value is used as the LR threshold at which we declare a true input–output match. This is sufficient for limiting the fraction of false identifications with the input catalogue, whilst ensuring a low probability of spuriously identifying a source in the output catalogue as a true positive, given that matched output sources are typically associated with multiple input sources.

### 3.5 Validation: completeness, position, and flux-density accuracy

We obtain the catalogue completeness and reliability, averaged over our Monte Carlo simulations, to confirm the ability of the source-detection algorithm to accurately recover an input source population. Throughout our analysis, we use the mask shown in Fig. 5 to characterize the catalogue properties away from the brightest diffuse emission along the Galactic plane. The mask was produced by smoothing the C-BASS northern sky map with a  $10^\circ$  FWHM Gaussian and masking the brightest 30 per cent of pixels. We also mask all pixels corresponding to declinations below  $\delta < -10^\circ$  in order to match the sky area observed by the C-BASS northern-sky survey. We refer to this mask as the CG30 mask.



**Figure 5.** The combined Galactic plane and C-BASS North mask used in our analysis of the C-BASS catalogue, shown here in Galactic coordinates. The mask was constructed to cover the brightest 30 per cent of the northern sky observed by C-BASS, along with declinations  $\delta < -10^\circ$ . We refer to this mask as the CG30 mask.

In determining the positional offset of the detected source from the input sources, we compare its position to the photocentre of matched input sources. That is, after matching to input sources using the LR test, we determine the weighted average position of the matched input sources, using the input source flux-densities as weights. In comparing flux densities, we sum flux densities for matched sources, weighted according to the C-BASS beam profile, i.e.

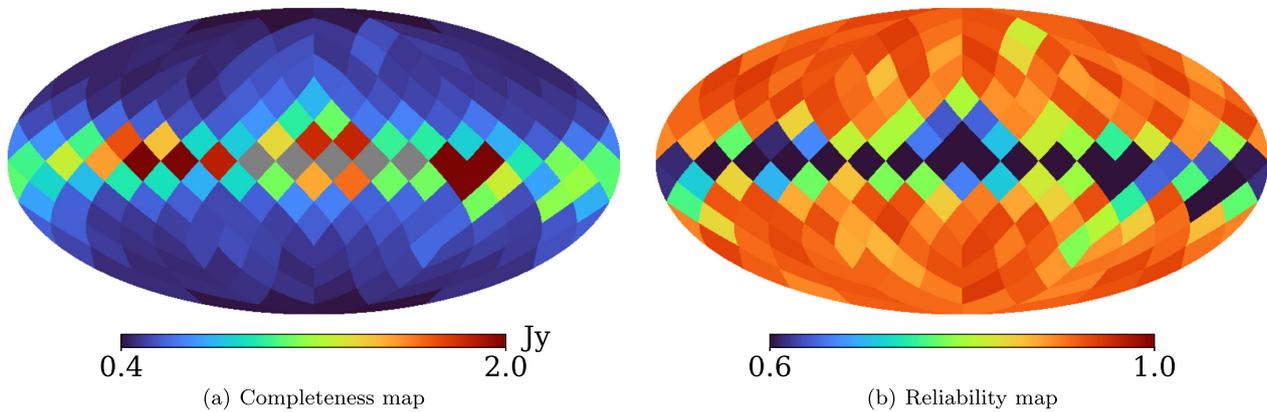
$$S_v^{\text{match}} = \sum_i S_v^{(i)} \mathcal{B}(|\mathbf{x}_i - \mathbf{x}_{\text{det}}|), \quad (16)$$

where the sum runs over matched sources in the input catalogues,  $S_v^{(i)}$  is the flux density of the  $i$ th matched source,  $\mathbf{x}_i$  is the position of the  $i$ th matched source,  $\mathbf{x}_{\text{det}}$  is the position of the detected source, and  $\mathcal{B}$  is the C-BASS beam profile.

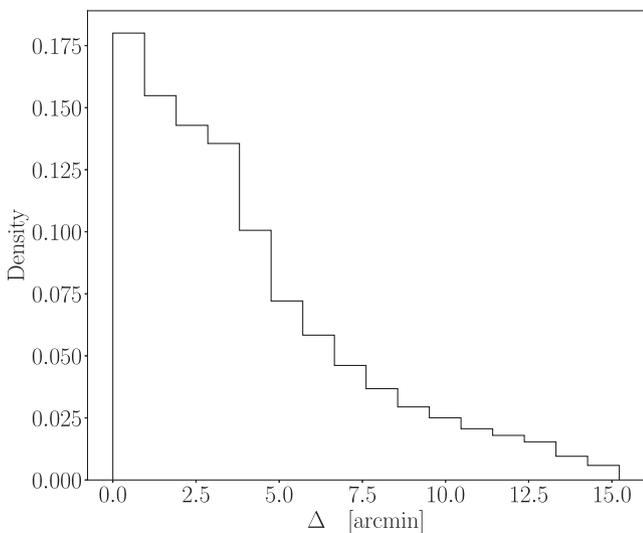
In Fig. 6, we show maps of the completeness and reliability in NSIDE = 4 pixels obtained for the validation simulations. We can see that close to the Galactic plane the 90 per cent completeness level is higher and the catalogue reliability is lower. Indeed, for some pixels along the Galactic plane 90 per cent completeness is never achieved. This is to be expected given the increased intensity of diffuse emission along the Galactic plane, which is not fully removed by the SMHW2 filter. Over regions of the sky left unmasked by the CG30 mask, we find a mean 90 per cent catalogue completeness of approximately 500 mJy, corresponding to a mean reliability of approximately 91 per cent.

From the validation simulations, we calculate the absolute positional offsets of detected sources from input sources. These cluster below 10 arcmin, peaking around zero, with a median offset of approximately 3.2 arcmin. Considering the 45 arcmin resolution of the maps used for source detection, and the significant confusion effects at this resolution, this is a small positional uncertainty for our source detection algorithm. Indeed, for the purposes of producing a point-source mask for the actual C-BASS maps this is more than sufficient. The histogram of absolute positional offsets of detected sources, applying the CG30 mask, is shown in Fig. 7.

We are able to recover input source flux-densities using aperture photometry, with a plot of detected flux densities against input flux densities being shown in Fig. 8. At flux densities greater than 1 Jy, we recover reasonable flux-density estimates for the majority of sources. A small fraction of sources are recovered with flux densities biased slightly high. This is to be expected given the 45 arcmin resolution of the C-BASS map, meaning that multiple input sources are associated with each detected source. The detected source flux



**Figure 6.** Panel (a): A map of the completeness estimates obtained from the validation simulations. Panel (b): A map of the corresponding reliability estimates. These estimates were calculated on NSIDE = 4 sky pixels and are shown in Galactic coordinates. In pixels outside of the CG30 mask, the mean 90 per cent completeness level is approximately 500 mJy, with a corresponding mean reliability of 91 per cent. It can be seen that along the Galactic plane the completeness and reliability is significantly lower than at higher Galactic latitudes. This is to be expected given the stronger diffuse emission that is not removed by the SMHW2 filter. Grey pixels in the completeness map denote pixels where 90 per cent completeness was never achieved.



**Figure 7.** The histogram of measured absolute positional offsets,  $\Delta$ , applying the CG30 mask, obtained for the 100 Monte Carlo simulations. Over the un-masked sky area we find the detected positional offsets peak around zero, with a median positional offset of 3.2 arcmin, which is small compared to the C-BASS beam of 45 arcmin. Larger positional offsets are predominantly driven by fainter detected sources below  $\sim 1$  Jy, and sources detected in regions of brighter diffuse emission. The positional offsets of all the detected sources will consist of samples drawn from Rayleigh distributions, which results in a tail of larger offsets beyond the peak of the positional offset distribution.

density will be the summed contribution from associated input sources. For flux densities below  $\sim 1$  Jy, the scatter in the recovered flux densities reaches the same magnitude as the input flux densities. This is again to be expected; below this level sources become much more heavily obscured by surrounding diffuse emission, making photometric estimates of source flux densities significantly more unreliable. We note here that there is a small cluster of recovered sources at these low flux densities where the recovered flux-density estimate is spuriously high. This is largely caused by sources being detected in regions of bright diffuse and highly extended emission, leading to errors in the background estimation.

Our validation simulations have demonstrated that our source-detection algorithm is able to accurately recover an underlying source population, from a sky consisting of additional contributions from diffuse emission and noise.

#### 4 THE C-BASS NORTHERN SKY CATALOGUE

In this section, we present the C-BASS northern sky, total-intensity point-source catalogue. The point-source catalogue was produced using data obtained from only night-time observations, combining data from the various elevation scans of the telescope. C-BASS data were calibrated against Tau A as a primary calibrator and Cas A as a secondary calibrator, with calibrator flux densities being calculated from the spectral forms in Weiland et al. (2011). For Tau A the spectral model takes the form

$$\log S_\nu(\text{Jy}) = 2.506 - 0.302 \log \left( \frac{\nu}{40 \text{ GHz}} \right), \quad (17)$$

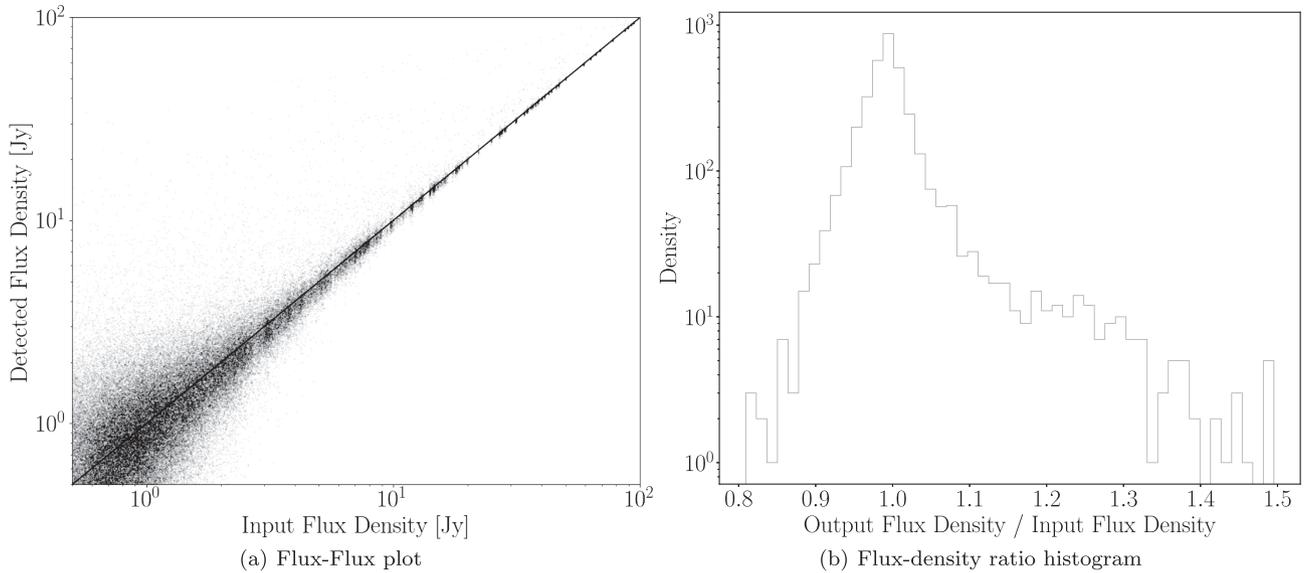
and for Cas A it takes the form

$$\log S_\nu(\text{Jy}) = 2.204 - 0.682 \log \left( \frac{\nu}{40 \text{ GHz}} \right) + 0.038 \log^2 \left( \frac{\nu}{40 \text{ GHz}} \right). \quad (18)$$

These spectral models give the flux density of Tau A at 4.76 GHz as  $S_\nu = 609.8 \pm 4.2$  Jy in epoch 2005, and the flux density of Cas A as  $S_\nu = 736.1 \pm 3.4$  Jy in epoch 2000. The secular decreases adopted for Tau A and Cas A were the same as those used for the model fits in Weiland et al. (2011). For Tau A this was  $-0.167$  per cent per year, which was taken from Macías-Pérez et al. (2010), and for Cas A the secular decrease was  $-0.53$  per cent per year. The C-BASS calibration is accurate to  $\sim 5$  per cent. The flux-temperature conversion is defined through

$$T_{\text{source}} = S_\nu(\text{Jy}) \frac{\pi}{2k_B \times 10^{26}} \left( \frac{6.1}{2} \right)^2 \alpha_{\text{eff}}, \quad (19)$$

where  $T_{\text{source}}$  is the source temperature,  $k_B$  is the Boltzmann constant, and  $\alpha_{\text{eff}} = 0.55$  is the theoretical aperture efficiency for C-BASS. Detailed discussions of the C-BASS data reduction and calibration will be given in the C-BASS survey and commissioning papers (Pearson et al. in preparation; Taylor et al., in preparation). We perform the source detection on the C-BASS map at NSIDE = 1024, using the high NSIDE map to allow for the precise determination of source positions.



**Figure 8.** Panel (a): The detected source flux densities against the matched input flux densities for the 100 Monte Carlo simulations, applying the CG30 mask. Panel (b): The histogram of the detected flux densities, divided by the matched input flux densities, for sources with input flux densities greater than 10 Jy. We can see that for the majority of bright sources we obtain reliable flux-density estimates. However, there is a small population ( $\sim 1$  per cent) of detected sources with spuriously high flux-density estimates, due to confusion effects.

The outline of the remainder of this section is as follows. In Section 4.1, we summarize the catalogue properties. In Section 4.2, we present the results from matching the C-BASS and GB6/PMN catalogues, using these results to quantify the C-BASS pointing accuracy and flux-density scale. In Section 4.3, we comment briefly on sources detected with outlier flux densities. In Section 4.4, we show the differential source counts for the C-BASS catalogue, and compare these to the GB6 source counts. Additional comparisons were also made with sources in the RATAN-600 catalogue over the NCP region. In this case, we recovered the expected bright source population, with the results obtained through these comparisons being consistent with those from comparisons with the GB6/PMN catalogues.

#### 4.1 Northern sky intensity catalogue: summary

For our default catalogue, we use the same SNR thresholds as for the validation analysis i.e.  $\text{SNR} \geq 2.5$  to tag candidate sources on the first loop over the sky, followed by a final detection threshold of  $\text{SNR} \geq 3.5$  after looping over the candidate source locations. To estimate the completeness and reliability of the real C-BASS catalogue, we compare it with the GB6 and PMN catalogues over their areas of common sky coverage. Applying the CG30 mask, we obtain a catalogue reliability of approximately 98 per cent, with a corresponding 90 per cent completeness level of 610 mJy, recovering the expected bright source population at 4.76 GHz. The confusion level in the C-BASS map is approximately 85 mJy (Jones et al. 2018). Below this level sources cannot be individually detected in our maps, acting as an additional noise contribution. Our analysis has resulted in a catalogue of 1784 point sources covering declinations  $\delta \geq -10^\circ$ . The catalogue properties are summarized in Table 2, with the detected source positions for the default catalogue shown in Fig. 9.

Beam sidelobes around bright sources can result in spurious source detections. The first sidelobe of the C-BASS beam has some azimuthal structure, and has a peak value of about 2.5 per cent of

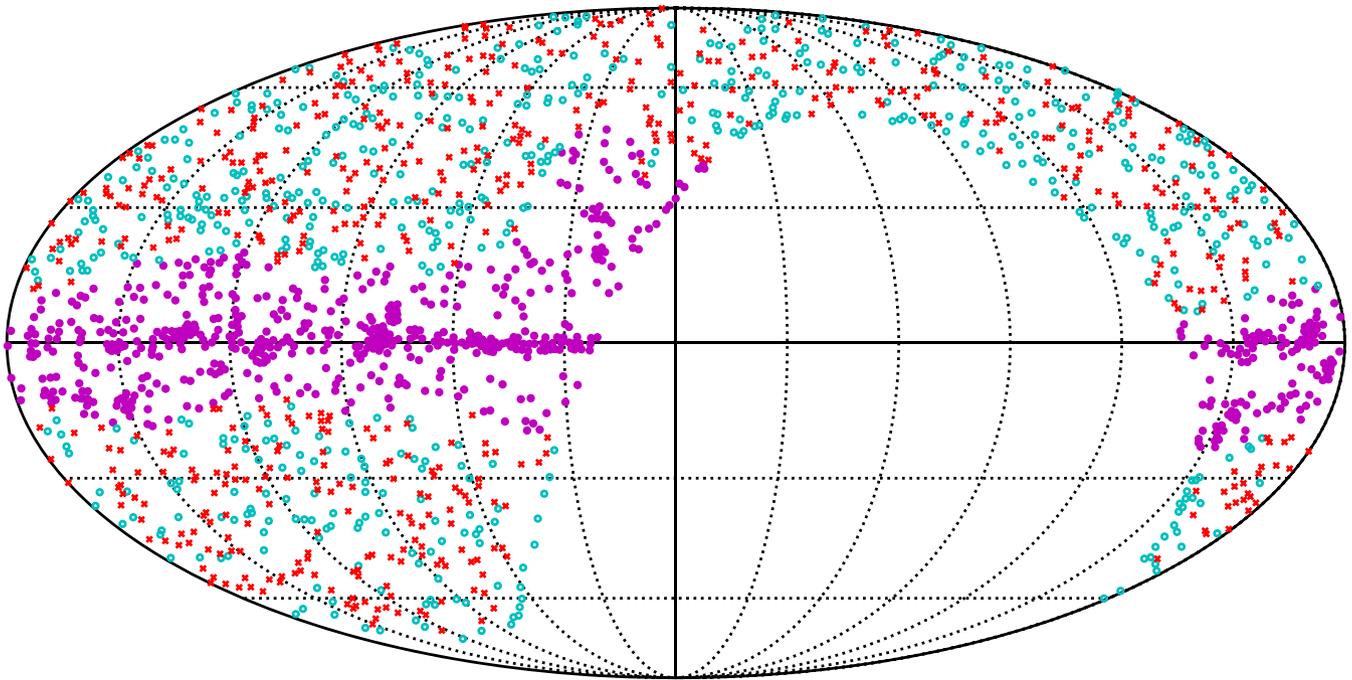
**Table 2.** C-BASS northern sky, total intensity catalogue summary. Position uncertainties include intrinsic uncertainty from the source detection algorithm, along with additional errors from the C-BASS pointing.

	C-BASS catalogue
Frequency	4.76 GHz
Map FWHM	45 arcmin
SNR threshold	3.5
Reliability	
Full Sky	90 per cent
CG30 mask	98 per cent
Number of sources	
Full sky	1784
CG30 mask	1136
90 per cent completeness	
Full sky	1000 mJy
CG30 mask	610 mJy
$N(\mathcal{S} > S_{90})^a$	
Full sky	793
CG30 mask	560
Position uncertainty	
Full sky	3.4 arcmin
CG30 mask	3.2 arcmin

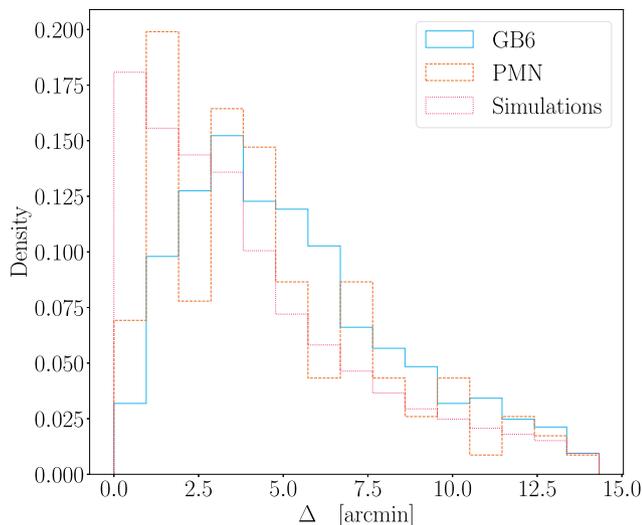
<sup>a</sup>This is the number of sources with measured flux densities greater than the 90 per cent completeness level.

the peak. Given the source detection level of around 500 mJy, we might expect spurious detections from sources brighter than 20 Jy. Outside of the CG30 mask there are three sources brighter than this, 3C 273, Virgo A and 3C 279. Therefore, we do not allow any additional detections within a  $2^\circ$  radius exclusion zone centred on these three sources. In constructing a point-source mask, any issues from spurious sidelobe detections can be further mitigated by masking larger areas around the brightest sources.

It is important to note here that, whilst there are a large number of source detections in the Galactic plane, the reliability of these



**Figure 9.** Positions of the northern sky, total intensity C-BASS point sources. Positions shown here are in Galactic coordinates. *Filled purple circles* denote sources that fall within the CG30 mask. *Open cyan circles* denote sources outside the mask, with a fitted flux density below the 90 per cent completeness level (610 mJy), and *red crosses* denote the remaining sources with fitted flux densities above the 90 per cent completeness level.



**Figure 10.** A histogram of the absolute positional offsets of C-BASS sources from matched GB6 and PMN sources, applying the CG30 mask, plotted alongside the corresponding histogram of positional offsets obtained from the validation simulations. We use the photocentres of matched GB6 and PMN sources for comparison. Positional offsets cluster below 10 arcmin, peaking at approximately 3.2 arcmin. We can see that the peak of the positional offsets obtained for the real C-BASS catalogue peak  $\sim 3$  arcmin higher than for the validation simulations. This is to be expected given there is some level of pointing jitter in the real C-BASS data that is not present in the simulations. We obtain similar distributions for the absolute positional offsets with both GB6 and PMN.

detections is significantly lower than at higher Galactic latitudes. This is a result of the more intense diffuse emission obscuring point sources, and highly extended emission in the Galaxy being spuriously identified as point sources. For the purposes of scientific

analyses with other CMB data sets, this is not a major issue, given that much of the Galactic plane will be masked out for such analyses.

#### 4.2 GB6 and PMN matching

To cross-check the C-BASS pointing and flux-density scale, we match the C-BASS point sources with the GB6 and PMN catalogues over their areas of common sky coverage. Similar to the situation encountered during the algorithm validation, the GB6 and PMN catalogues were produced using much higher resolution surveys than C-BASS. What may appear as a single source in the C-BASS maps could actually correspond to multiple sources in the GB6 and PMN catalogues. Matches between catalogues are found using the likelihood ratio technique described in Section 3.4. In characterizing the C-BASS pointing, we compare the C-BASS source position with the photocentre (defined in Section 3.4) of matched GB6 or PMN sources. In comparing flux densities, we use the beam-weighted sum of matched source flux densities, defined in equation (16).

In Fig. 10, we show a histogram of the absolute positional offsets of C-BASS sources from matched GB6 and PMN sources, using the photocentres of matched GB6 and PMN sources for comparison. The offsets are clustered below 10 arcmin, peaking at approximately 3.2 arcmin.<sup>2</sup> This is broadly consistent with the results obtained from our Monte Carlo simulations during the algorithm validation, with additional errors arising here from the C-BASS pointing. The distributions of the absolute positional offsets when matching to GB6 and PMN are also very similar. Larger positional offsets ( $\gtrsim 10$  arcmin) are predominantly driven by fainter detected sources below  $\sim 1$  Jy and sources detected in regions of brighter diffuse emission. Further, given a set of positional offsets drawn from

<sup>2</sup>Note that the position of the mode for a Rayleigh distribution is given by  $\sigma_r$ .

Rayleigh distributions, we expect the overall distribution to have a tail of larger offsets. The typical positional offsets determined using the C-BASS source catalogue are consistent with the estimated pointing accuracy of C-BASS (Pearson et al. in preparation).

We also plot the C-BASS source flux densities (determined using aperture photometry, as described in Section 3.1) against matched GB6 and PMN flux-densities, using equation (16) to determine the matched flux density, in Fig. 11. When no Galactic plane mask is applied, we can see that most sources cluster around the line  $S_v^{\text{GB6}} = S_v^{\text{CBASS}}$  ( $S_v^{\text{PMN}} = S_v^{\text{CBASS}}$ ). However, there is an apparent population of sources where the measured C-BASS flux density is significantly greater than the GB6 and PMN flux densities. This is largely due to the tagging of sources in the Galactic plane, where the additional presence of very bright, highly extended emission leads to spuriously high flux-density estimates. On applying a Galactic plane mask this population is largely removed. The small number of remaining sources where C-BASS significantly overestimates the flux densities may either result from sources sitting in regions of bright diffuse emission (e.g. within a Galactic spur) or from heavily confused sources in the C-BASS map. Most of the sources recovered with spuriously high densities are associated with low flux-density GB6/PMN sources ( $\lesssim 1$  Jy). At these flux densities, the uncertainties in our flux-density estimates from aperture photometry become much larger, consistent with the larger spread in C-BASS to matched flux densities at these levels.

To quantify the agreement between the C-BASS and GB6/PMN flux-density scales, we fit a straight line between the C-BASS and GB6/PMN flux-densities,

$$S_v^{\text{CBASS}} = b S_v^{\text{GB6/PMN}}, \quad (20)$$

where  $b$  is the gradient between the C-BASS and GB6/PMN flux densities. We use a measurement-error model (Stan Development Team 2012) to account for uncertainties in the C-BASS and GB6/PMN flux densities by treating the true value of the GB6/PMN flux densities as missing data and marginalizing over the distribution of the measured GB6/PMN flux densities. We also model the outlier distribution following Hogg, Bovy & Lang (2010). For this, we assume outliers are drawn from some broad Gaussian distribution specified by a mean  $\mu_{\text{out}}$  and standard deviation  $\sigma_{\text{out}}$ . The probability of a data point being an outlier is given by the parameter  $P_{\text{out}}$ . The full set of priors is given by

$$b \sim \mathcal{N}(\mu = 1, \sigma = 0.25), \quad (21)$$

$$\mu_{\text{out}} \sim \mathcal{N}(\mu = 1, \sigma = 100), \quad (22)$$

$$\sigma_{\text{out}} \sim \text{Exp}(\lambda = 1/100), \quad (23)$$

$$P_{\text{out}} \sim \text{Unif}(0, 1), \quad (24)$$

$$S_v^{x,i} \sim \text{Unif}(0, \infty), \quad (25)$$

$$S_v^{\text{GB6/PMN},i} \sim \mathcal{N}(\mu = S_v^{x,i}, \sigma = \sigma_{\text{GB6/PMN}}^i), \quad (26)$$

where  $S_v^{x,i}$  is the true GB6/PMN flux density for the  $i$ th matched source,  $S_v^{\text{GB6/PMN},i}$  is the measured GB6/PMN flux density for the  $i$ th matched source, and  $\sigma_{\text{GB6/PMN}}^i$  is the corresponding uncertainty

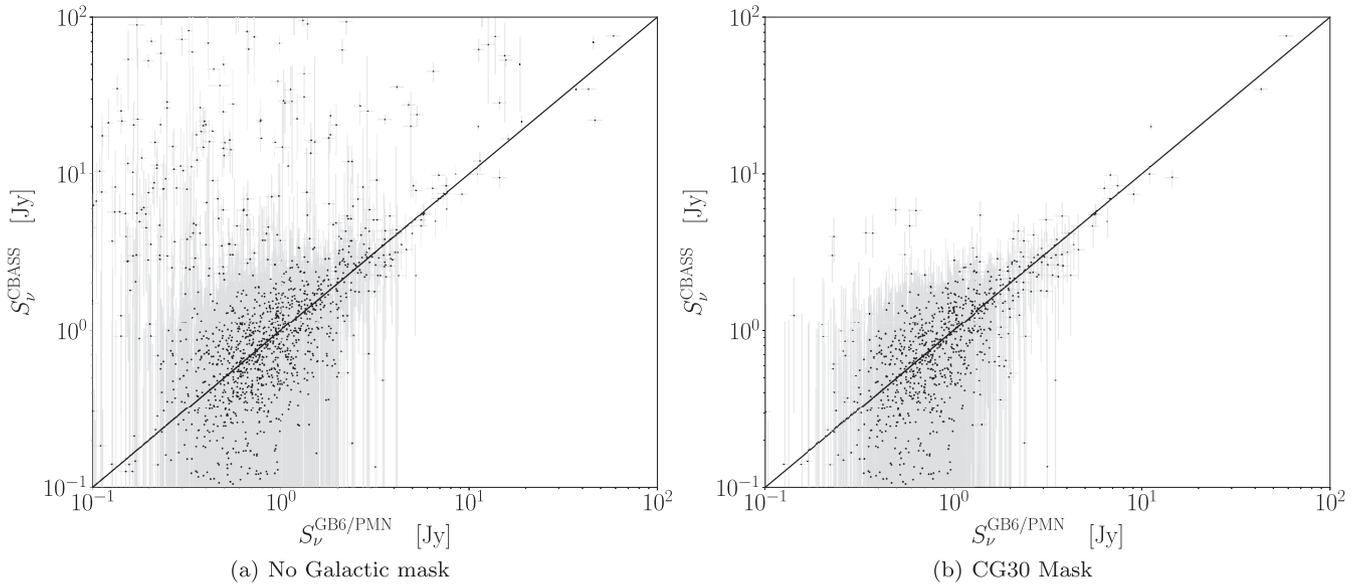
on the GB6/PMN flux density. The likelihood is given by

$$\mathcal{L} \propto \prod_{i=1}^{N_s} \left[ \frac{1 - P_{\text{out}}}{\sqrt{2\pi} (\sigma_{\text{CBASS}}^i)^2} \exp \left\{ -\frac{(S_v^{\text{CBASS},i} - b S_v^{x,i})^2}{2 (\sigma_{\text{CBASS}}^i)^2} \right\} + \frac{P_{\text{out}}}{\sqrt{2\pi} (\sigma_{\text{out}}^2 + (\sigma_{\text{CBASS}}^i)^2)} \exp \left\{ -\frac{(S_v^{\text{CBASS},i} - \mu_{\text{out}})^2}{2 (\sigma_{\text{out}}^2 + (\sigma_{\text{CBASS}}^i)^2)} \right\} \right], \quad (27)$$

where  $N_s$  is the number of matched sources,  $S_v^{\text{CBASS},i}$  is the measured C-BASS flux density for the  $i$ th matched source, and  $\sigma_{\text{CBASS}}^i$  is the corresponding uncertainty on the C-BASS flux density. We sample our model using the No-U-Turn Sampler (NUTS), implemented in PyMC3 (Hoffman & Gelman 2014; Salvatier, Wiecki & Fonnesbeck 2016). NUTS is a self-tuning variant of Hamiltonian Monte Carlo that allows for highly efficient exploration of the parameter space.

Pauliny-Toth et al. (1978) showed that approximately 60 per cent of strong radio sources at 5 GHz have flat spectra and are therefore likely to be compact, variable sources. Indeed, previous studies of the GB6 source catalogue have found a significant fraction of variable sources (Gregory et al. 1996; Gregory, Scott & Poller 1998; Gregory et al. 2001). As such, we would expect roughly half of the objects in 5 GHz surveys to be varying, meaning any comparison of individual flux densities observed at different epochs will be strongly contaminated by source variability. A comparison of flux densities over the whole catalogue remains useful for checking the overall consistency of the different flux-density scales. However, we can use the CRATES catalogue as a proxy for potentially variable sources in the C-BASS catalogue. By matching the C-BASS catalogue with the corresponding 4.85 GHz sources in the CRATES catalogue, we can remove any C-BASS sources that are positively identified with a CRATES source. In doing so, we obtain a set of CRATES cleaned C-BASS sources with the flat-spectrum sources largely removed, thereby removing the primary source of variability in the C-BASS catalogue. We perform the flux-density fits described above to both the full C-BASS catalogue matches, along with the CRATES cleaned matches. The latter set of fits will be better able to detect any instrumental or systematic errors in the C-BASS flux-density scale.

The results for our fitted values of  $b$ , comparing to GB6 and PMN, are given in Table 3. There is a slight discrepancy between the C-BASS flux-density scale and the PMN scale, when applying the CG30 mask and comparing to the full C-BASS catalogue. This discrepancy is removed when we remove any C-BASS sources associated with a CRATES source. We see a similar pattern for GB6, with the overall consistency being improved by cleaning the C-BASS catalogue of flat-spectrum sources. The C-BASS flux-density scale is consistent with the GB6 flux-density scale to within  $\sim 4$  per cent. For the PMN catalogue, there are fewer sources for comparison in the common overlap region, meaning consistency is only established in this case to  $\sim 10$  per cent. Differences between the flux-density scales will result from a number of complicating factors e.g. the fact that C-BASS sources are really blends of multiple GB6/PMN sources, differing pass-bands, source variability etc. Given this, and the additional  $\sim 5$  per cent uncertainties in the GB6 and PMN flux-density scales, we conclude that the C-BASS flux-density scale is consistent with these flux-density scales. We do not include any colour corrections in this analysis. The choice of 4.76 GHz as the effective frequency for C-BASS minimizes colour corrections to  $\sim 0.02$  per cent for a source with spectral index  $\alpha = 0$ , and  $\sim 0.2$  per cent for a source with spectral index  $\alpha = -1$ .



**Figure 11.** Flux–flux plots for matched C-BASS and GB6/PMN sources. Panel (a): A plot of C-BASS flux densities against matched GB6/PMN flux densities, with no mask applied. Panel (b): A plot of C-BASS flux densities against matched GB6/PMN flux densities, applying the CG30 mask. We also plot the line  $S_{\nu}^{\text{GB6/PMN}} = S_{\nu}^{\text{CBASS}}$  in each panel. When we do not apply a Galactic mask, we can see a group of C-BASS sources with significantly higher flux densities than the matched GB6/PMN sources. This can be attributed to errors in fitting for source flux densities in the Galactic plane, particularly in the presence of bright highly extended emission. On applying a Galactic plane mask, this group of sources is largely removed and sources cluster more tightly around  $S_{\nu}^{\text{GB6/PMN}} = S_{\nu}^{\text{CBASS}}$ . Apparent outliers in the flux-density matches can largely be attributed to variable, flat-spectrum sources. Removing any C-BASS sources identified with flat-spectrum sources in the CRATES catalogue, we remove most of these outliers and improve the consistency between the C-BASS and GB6/PMN flux-density scales.

**Table 3.** Fitted values for  $b$ , the ratio between C-BASS and GB6 or PMN flux densities. Values were obtained by comparing flux densities between C-BASS and GB6/PMN, for both the full C-BASS catalogue and the CRATES-cleaned C-BASS catalogue. Alongside the mean fitted value for  $b$ , we also give the 95 per cent highest posterior density (HPD) interval. We can see that by removing sources from the C-BASS catalogue associated with sources in the CRATES catalogue, the overall consistency between the C-BASS and GB6/PMN catalogues is improved.

Comparison Catalogue	CRATES Cleaned?	$\langle b \rangle$	95 per cent HPD[ $b$ ]
GB6 no mask	No	0.95	{0.90, 1.00}
GB6 CG30 mask	No	0.95	{0.91, 1.00}
GB6 no mask	Yes	1.04	{0.97, 1.11}
GB6 CG30 mask	Yes	1.04	{0.97, 1.11}
PMN no mask	No	0.95	{0.89, 1.00}
PMN CG30 mask	No	0.82	{0.68, 0.97}
PMN no mask	Yes	1.05	{0.93, 1.17}
PMN CG30 mask	Yes	1.01	{0.83, 1.19}

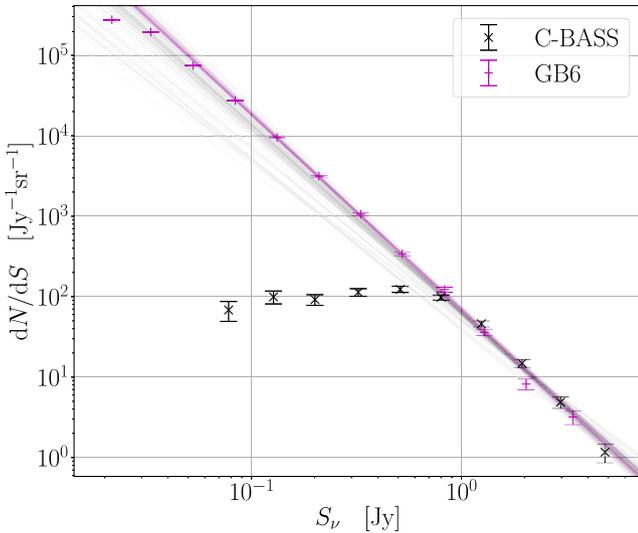
Colour correction for GB6/PMN are negligible due to their narrow bandwidths. Corrections due to flux-density extrapolation are also not considered here. For a typical source with spectral index,  $\alpha \approx -0.83$  the correction is  $\sim 1.5$  percent. Given the uncertainties on the fitted values of  $b$ , and the additional uncertainties on the GB6/PMN flux-density scales, any such corrections to the flux densities of C-BASS sources will not affect our overall conclusions here.

### 4.3 Outlier sources

In Fig. 11, we can see a number of C-BASS sources showing discrepancies with the GB6/PMN flux densities. As noted above, we

expect a significant number of sources with flux densities  $S_{\nu} \gtrsim 1$  Jy to be flat-spectrum, variable sources. If we extract sources from the CRATES cleaned C-BASS catalogue that show a  $\geq 3\sigma$  discrepancy between the C-BASS and GB6/PMN flux densities, we find six sources. In four of the cases, the sources are buried deep within surrounding diffuse emission, with the matched GB6 flux densities being below 500 Jy. In this situation, aperture photometry does not yield reliable point estimates of individual flux densities.

We identify QSO B0723-007 as having a discrepant flux density when matching to PMN ( $5.5 \pm 1.2$  Jy in C-BASS compared to  $1.4 \pm 0.07$  Jy in PMN). This is a BL Lacertae object, which are known to exhibit a high degree of radio variability (Nieppola et al. 2009), but is not present in the CRATES catalogue. However, it is important to note that CRATES pre-selected flat-spectrum sources for observation on the basis of spectral indices calculated between the PMN and NVSS catalogues in this region. Given the large separation in time between these surveys and the CRATES survey, we may expect CRATES to miss a small number of variable sources (Healey et al. 2007). The final extracted source was Virgo A. Whilst Virgo A is not a variable source, it does possess a large-scale halo that is not picked up the small beam of GB6, leading to the seemingly anomalously high value of the C-BASS flux density ( $75.9 \pm 1.08$  Jy). The C-BASS flux density for Virgo A is consistent with measurements in Baars et al. (1977). Whilst a detailed study of source variability between the C-BASS and GB6/PMN catalogues is beyond the scope of this paper, flat-spectrum sources do account for a significant degree of variability between these catalogues. This is evidenced by the improved consistency between the flux-density scales when matching to the CRATES-cleaned C-BASS catalogue, and the lack of a large additional population of variable sources beyond those identified with the CRATES catalogue.



**Figure 12.** Differential source counts calculated using the C-BASS northern sky total intensity catalogue, plotted alongside the differential source counts for the GB6 catalogue. Source counts were calculated applying the CG30 mask. We also plot as grey-scale the fitted model source counts corresponding to 100 posterior samples for the C-BASS and GB6 models. At low-flux densities, we note the expected lower levels in the C-BASS counts. This arises due to the low 45 arcmin resolution of the C-BASS map, leading to confusion issues for faint sources, along with fainter sources being more heavily obscured by diffuse emission and noise. Good agreement between GB6 and C-BASS differential source counts is found over intermediate flux densities ( $\sim 1\text{--}10\text{ Jy}$ ).

#### 4.4 Differential source counts

Differential source counts have been intensively studied (e.g. de Zotti et al. 2005; de Zotti et al. 2010; AMI Consortium et al. 2011), having important implications for models of cosmic evolution (e.g. Condon 1984; Massardi et al. 2010) and provided the first clear evidence against the Steady State Theory (Pooley & Ryle 1968). However, for the purposes of C-BASS analysis, we calculate the differential source counts to provide an additional cross-check on the statistical properties of our bright-source population.

Given that we are concerned with the statistical properties of our catalogue at flux-densities  $\gtrsim 1\text{ Jy}$ , where the C-BASS catalogue is largely complete, we do not consider correction factors to the differential source counts necessary at lower flux densities to account for catalogue incompleteness and varying flux-density sensitivities across the sky. Sources are put into flux-density bins of width 0.2 dex. For each bin, source counts are determined against a central flux density,  $S_c$ , taken to be the geometric mean of flux densities in that bin. The differential source counts for a given flux-density bin are then given by

$$\left. \frac{dN}{dS} \right|_{S=S_c} = n(S_c) = \frac{N_c}{A \Delta S}, \quad (28)$$

where  $N_c$  is the number of sources in the flux-density bin,  $A$  is the sky area to which the catalogue corresponds, and  $\Delta S$  is the bin width. Errors are estimated for the source counts by assuming Poisson statistics. We only consider bins with  $N_c \geq 9$  so as to maintain a  $\text{SNR} \geq 3$  for each bin.

The differential source counts for C-BASS North in total intensity, applying the CG30 mask, are plotted in Fig. 12, alongside the differential source counts for the GB6 catalogue. We can see that at low flux densities the C-BASS source counts are significantly lower.

This is to be expected from the increasing catalogue incompleteness at these lower flux densities. From the differential source counts, we can again see the catalogue completeness level from the point at which the source count turn-off ends and we have good agreement between the C-BASS and GB6 counts, i.e. over flux densities  $\sim 1\text{--}10\text{ Jy}$ . It is worth noting here that the errors obtained from assuming Poisson statistics are likely to be somewhat optimistic. These do not account for false positives in the source catalogues, or errors arising from incorrect binning due to uncertain flux-density estimates. These effects become significant at lower flux densities, where the proportion of false positives in the catalogue is greater, and our photometry methods become less reliable.

We fit the usual power-law model to the differential source counts, given by

$$\frac{dN}{dS} = n(S) = \Lambda \left( \frac{S}{S_0} \right)^\gamma, \quad (29)$$

where  $\Lambda$  is the count amplitude,  $S_0 = 3\text{ Jy}$  is the reference flux density, and  $\gamma$  is the power-law index for the source counts. For C-BASS sources, we fit to data points above the turn-off in the source counts ( $S_\nu > 750\text{ Jy}$ ), where the C-BASS catalogue is largely complete. For GB6, we fit to data points over the range  $S_\nu > 0.07\text{ Jy}$ , so as to avoid the need to apply correction factors to faint GB6 source counts. We assume a Gaussian likelihood, and assign the following weakly informative priors to  $\Lambda$  and  $\gamma$ ,

$$\Lambda \sim \mathcal{N}(\mu = 5, \sigma = 5), \quad (30)$$

$$\gamma \sim \mathcal{N}(\mu = -2.5, \sigma = 2.0). \quad (31)$$

To account for additional errors in the source counts arising from flux-density uncertainties and incorrect binning of sources, we marginalize over an additional percentage error,  $f_{\text{extra}}$ , so that the total error on the source count in the  $i$ th bin is given by

$$\sigma_i = \sqrt{\sigma_{\text{Poisson}}^2 + (f_{\text{extra}} n(S_i))^2}, \quad (32)$$

where  $\sigma_{\text{Poisson}}$  is the Poisson error on the source count. We assign an exponential prior to the additional percentage error,  $f_{\text{extra}} \sim \text{Exp}(\lambda = 20)$ , corresponding to a characteristic value of 5 per cent. We again explore our parameter space using the NUTS algorithm.

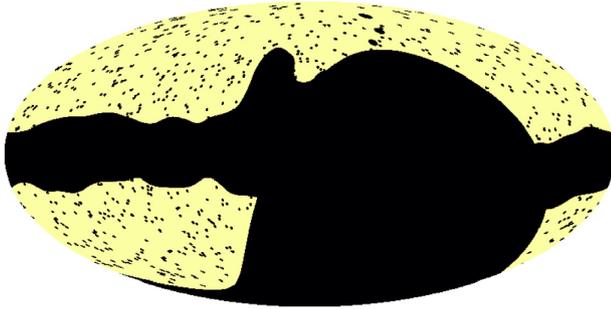
The results from these power-law fits are summarized in Table 4. When we do not apply any Galactic plane mask to the C-BASS catalogue, we fit a noticeably shallower spectral index than for the GB6 catalogue. This is driven by the excess in bright source counts in the C-BASS catalogue. This is caused by the spurious tagging of bright, highly extended emission in the Galactic plane, and the overestimation of source flux-densities in the presence of such emission, as discussed in Section 4.2. On applying the CG30 mask, the 95 per cent HPD intervals for C-BASS and GB6 overlap as expected. This can be seen in Fig. 12, where we plot 100 posterior samples from the C-BASS and GB6 models. When we apply the CG30 mask the posterior samples for the C-BASS and GB6 models overlap significantly.

## 5 CONCLUSIONS

We have produced a total-intensity catalogue of point sources at 4.76 GHz, covering declinations  $\delta \geq 10^\circ$ . The catalogue was produced by filtering the whole sky using the SMHW2 filter, and contains 1784 sources. The C-BASS source catalogue has been characterized through comparisons with the pre-existing GB6 and PMN catalogues over their common declination ranges. Through

**Table 4.** Fitted power-law parameters for the differential source counts. We quote the mean parameter values and the 95 per cent HPD intervals. Results are given for cases where we apply no mask and the CG30 mask.

Catalogue	$\langle\Lambda\rangle$	95 per cent HPD[ $\Lambda$ ]	$\langle\gamma\rangle$	95 per cent HPD[ $\gamma$ ]
C-BASS				
No mask	5.10	{4.43, 5.73}	-1.52	{-1.60, -1.43}
CG30 mask	4.73	{3.75, 5.70}	-2.37	{-2.59, -2.16}
GB6				
No mask	3.15	{2.75, 3.53}	-2.41	{-2.46, -2.36}
CG30 mask	4.55	{3.80, 5.28}	-2.45	{-2.52, -2.39}



**Figure 13.** An example point-source mask derived from the C-BASS total intensity northern sky catalogue, shown here in Galactic coordinates with masked pixels in black. In this example, we mask all sources with flux densities greater than 610 mJy i.e. above the 90 per cent completeness level obtained when applying the CG30 mask. We mask all pixels within 45 arcmin of a source, and further mask all pixels within  $2^\circ$  of sources with flux-densities greater than 10 Jy. In addition to masking around sources, we have masked all pixels within the CG30 mask.

these direct comparisons we estimate the C-BASS catalogue to have a 90 per cent completeness level of approximately 610 mJy, when applying the CG30 mask, with a corresponding reliability of 98 per cent. The catalogue is available on request to the corresponding author, and will be made publicly available with the C-BASS northern sky data release. A sample of the first ten sources from the C-BASS catalogue is given in the Appendix.

These comparisons with the GB6 and PMN catalogues also enabled a number of checks to be performed on the C-BASS data. We find a modal absolute positional offset of approximately 3.2 arcmin when applying the CG30 mask. The low level of additional scatter in positional offsets, compared to the positional uncertainty found during the algorithm validation, acts to confirm the accuracy of the C-BASS pointing. We compare the flux densities of C-BASS sources with matched sources in the GB6 and PMN catalogues, using a measurement error model to fit for the gradient between the flux-density scales. Applying the CG30 mask, and cleaning flat-spectrum sources from C-BASS using the CRATES catalogue, we find a mean gradient of  $\langle b \rangle = 1.04$  comparing to GB6, and  $\langle b \rangle = 1.01$  comparing to PMN. In both cases, the 95 per cent HPD intervals overlap with 1. Considering the  $\sim 5$  per cent uncertainties in the GB6 and PMN flux-density scales, we conclude that the C-BASS and GB6/PMN flux-density scales are consistent.

As an additional check on the statistical properties of the bright source population in the C-BASS catalogue, we calculate the differential source counts for C-BASS sources. Applying the CG30

mask, the C-BASS source counts show excellent agreement with the GB6 source counts over the flux-density range of  $\sim 1$ –10 Jy. At lower flux densities, we see the expected turn-off in the C-BASS source counts as the C-BASS catalogue becomes more incomplete.

For the purposes of C-BASS analysis, the primary products of this work are the checks on the data quality, and a mask of the bright point-source emission in the C-BASS map. In Fig. 13, we show an example point-source mask derived from the C-BASS catalogue, combined with the CG30 mask. Here, we mask all sources with flux densities greater than 610 mJy. We mask all pixels within 45 arcmin of a source, and further mask all pixels within  $2^\circ$  of sources brighter than 10 Jy. In future work, this will be extended to include a catalogue of polarized sources. The mask produced from this will be key for the analysis of C-BASS polarization data.

## ACKNOWLEDGEMENTS

The C-BASS project is a collaboration between Oxford and Manchester Universities in the UK, the California Institute of Technology in the USA, Rhodes University, UKZN and the South African Radio Observatory in South Africa, and the King Abdulaziz City for Science and Technology (KACST) in Saudi Arabia. It has been supported by the NSF awards AST-0607857, AST-1010024, AST-1212217, and AST-1616227, and NASA award NNX15AF06G, the University of Oxford, the Royal Society, STFC, and the other participating institutions. This research was also supported by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Technology. CD and SH acknowledge support from an STFC Consolidated grant no. (ST/P000649/1). CD acknowledges support from an ERC Starting (Consolidator) grant no. (307209). MWP acknowledges funding from a FAPESP Young Investigator fellowship, grant no. 2015/19936-1. We make use of the PYTHON MATPLOTLIB (Hunter 2007), NUMPY (Oliphant 2006), SCIPY (Jones et al. 2001), HEALPY (Górski et al. 2005; Zonca et al. 2019), ASTROPY (Astropy Collaboration et al. 2013, 2018), OPENCV (Kaehler & Bradski 2016), PYMC3 (Salvatier et al. 2016), and THEANO (Al-Rfou et al. 2016) packages.

## REFERENCES

- Al-Rfou R. et al., 2016, preprint ([arXiv:1605.02688](https://arxiv.org/abs/1605.02688))
- Ali-Haimoud Y., Hirata C. M., Dickinson C., 2009, *MNRAS*, 395, 1055
- AMI Consortium et al., 2011, *MNRAS*, 415, 2708
- Antoine J.-P., Vandergheynst P., 1999, *Appl. Comput. Harmon. Anal.*, 7, 262
- Argüeso F., Sanz J. L., Herranz D., López-Cañiego M., González-Nuevo J., 2009, *MNRAS*, 395, 649
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123

- Baars J. W. M., Genzel R., Pauliny-Toth I. I. K., Witzel A., 1977, *A&A*, 500, 135
- Bennett C. L. et al., 2013, *ApJS*, 208, 20
- Brandt W. N., Lawrence C. R., Readhead A. C. S., Pakianathan J. N., Fiola T. M., 1994, *ApJ*, 424, 1
- Carvalho P., Rocha G., Hobson M. P., Lasenby A., 2012, *MNRAS*, 427, 1384
- Cayón L., Sanz J. L., Martínez-González E., Banday A. J., Argüeso F., Gallegos J. E., Górski K. M., Hinshaw G., 2001, *MNRAS*, 326, 1243
- Chapin E. L. et al., 2011, *MNRAS*, 411, 505
- Condon J. J., 1984, *ApJ*, 287, 461
- 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
- Curto A., Martínez-González E., Barreiro R. B., 2011, *MNRAS*, 412, 1038
- Curto A., Tucci M., González-Nuevo J., Toffolatti L., Martínez-González E., Argüeso F., Lapi A., López-Caniego M., 2013, *MNRAS*, 432, 728
- de Zotti G., Ricci R., Mesa D., Silva L., Mazzotta P., Toffolatti L., González-Nuevo J., 2005, *A&A*, 431, 893
- de Zotti G., Massardi M., Negrello M., Wall J., 2010, *A&AR*, 18, 1
- Dickinson C. et al., 2019, *MNRAS*, 485, 2844
- Draine B. T., 2011, *Physics of the Interstellar and Intergalactic Medium*, Princeton University Press, USA
- Génova-Santos R. et al., 2015, *MNRAS*, 452, 4169
- González-Nuevo J., Argüeso F., López-Caniego M., Toffolatti L., Sanz J. L., Vielva P., Herranz D., 2006, *MNRAS*, 369, 1603
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
- Gregory P. C., Scott W. K., Douglas K., Condon J. J., 1996, *ApJS*, 103, 427
- Gregory P. C., Scott W. K., Poller B., 1998, in Zensus J. A., Taylor G. B., Wrobel J. M., eds, *ASP Conf. Ser.*, Vol. 144, IAU Colloq. 164: Radio Emission from Galactic and Extragalactic Compact Sources. Astron. Soc. Pac., San Francisco, p. 283
- Gregory P. C., Capak P., Gasson D., Scott W. K., 2001, in Schilizzi R. T., ed., *IAU Symp.* Vol. 205, *Galaxies and their Constituents at the Highest Angular Resolutions*. Astronomical Society of the Pacific, USA, p. 98
- Griffith M. R., Wright A. E., 1993, *AJ*, 105, 1666
- Griffith M. R., Wright A. E., Burke B. F., Ekers R. D., 1994, *ApJS*, 90, 179
- Griffith M. R., Wright A. E., Burke B. F., Ekers R. D., 1995, *ApJS*, 97, 347
- Haslam C. G. T., Klein U., Salter C. J., Stoffel H., Wilson W. E., Cleary M. N., Cooke D. J., Thomasson P., 1981, *A&A*, 100, 209
- Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, *A&AS*, 47, 1
- Healey S. E., Romani R. W., Taylor G. B., Sadler E. M., Ricci R., Murphy T., Ulvestad J. S., Winn J. N., 2007, *ApJS*, 171, 61
- Healey S. E., Fuhrmann L., Taylor G. B., Romani R. W., Readhead A. C. S., 2009, *AJ*, 138, 1032
- Herranz D., López-Caniego M., Sanz J. L., González-Nuevo J., 2009, *MNRAS*, 394, 510
- Hoffman M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593
- Hogg D. W., Bovy J., Lang D., 2010, preprint ([arXiv:1008.4686](https://arxiv.org/abs/1008.4686))
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jones E., et al., 2001, *SciPy: Open Source Scientific Tools for Python*, Available at: [https://www.researchgate.net/publication/213877848\\_SciPy\\_Open\\_Source\\_Scientific\\_Tools\\_for\\_Python#:~:text=SciPy%20is%20a%20Python%2Dbased](https://www.researchgate.net/publication/213877848_SciPy_Open_Source_Scientific_Tools_for_Python#:~:text=SciPy%20is%20a%20Python%2Dbased)
- Jones M. E. et al., 2018, *MNRAS*, 480, 3224
- Kaehler A., Bradski G., 2016, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, USA
- Kovalev Y. Y., Nizhelsky N. A., Kovalev Y. A., Berlin A. B., Zhekanis G. V., Mingaliev M. G., Bogdantsov A. V., 1999, *A&AS*, 139, 545
- Kuehr H., Pauliny-Toth I. I. K., Witzel A., Schmidt J., 1981, *AJ*, 86, 854
- López-Caniego M., Herranz D., González-Nuevo J., Sanz J. L., Barreiro R. B., Vielva P., Argüeso F., Toffolatti L., 2006, *MNRAS*, 370, 2047
- Macías-Pérez J. F., Mayet F., Aumont J., Désert F. X., 2010, *ApJ*, 711, 417
- Mann R. G. et al., 1997, *MNRAS*, 289, 482
- Martínez-González E., Gallegos J. E., Argüeso F., Cayón L., Sanz J. L., 2002, *MNRAS*, 336, 22
- Massardi M., Bonaldi A., Negrello M., Ricciardi S., Raccanelli A., de Zotti G., 2010, *MNRAS*, 404, 532
- Mesa D., Baccigalupi C., De Zotti G., Gregorini L., Mack K. H., Vigotti M., Klein U., 2002, *A&A*, 396, 463
- Mingaliev M. G., Stolyarov V. A., Davies R. D., Melhuish S. J., Bursov N. A., Zhekanis G. V., 2001, *A&A*, 370, 78
- Mingaliev M. G., Sotnikova Y. V., Bursov N. N., Kardashev N. S., Larionov M. G., 2007, *Astron. Rep.*, 51, 343
- Nieppola E., Hovatta T., Tornikoski M., Valtaoja E., Aller M. F., Aller H. D., 2009, *AJ*, 137, 5022
- Næss S. K., Louis T., 2013, *J. Cosmol. Astro-Part. Phys.*, 2013, 001
- Oliphant T., 2006, *NumPy: A Guide to NumPy*. Trelgol Publishing, USA
- Pauliny-Toth I. I. K., Witzel A., Preuss E., Kühr H., Kellermann K. L., Fomalont E. B., Davis M. M., 1978, *AJ*, 83, 451
- Planck Collaboration et al., 2014, *A&A*, 571, A28
- Planck CollaborationIX, 2016a, *A&A*, 594, A9
- Planck CollaborationX, 2016b, *A&A*, 594, A10
- Planck Collaboration et al., 2016c, *A&A*, 594, A26
- Pooley G. G., Ryle M., 1968, *MNRAS*, 139, 515
- Puglisi G. et al., 2018, *ApJ*, 858, 85
- Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M. A., Ghosh T., 2015, *MNRAS*, 451, 4311
- Remazeilles M. et al., 2018, *J. Cosmol. Astropart. Phys.*, 2018, 023
- Richter G. A., 1975, *Astron. Nachr.*, 296, 65
- Rutledge R. E., Bildsten L., Brown E. F., Pavlov G. G., Zavlin V. E., 2001, *ApJ*, 551, 921
- Salvater J., Wiecki T. V., Fannesbeck C., 2016, *PeerJ Computer Science*, 2, e55
- Stan Development Team, 2012, *Stan Modeling Language User's Guide and Reference Manual*, Version 1.0. Available at: <https://mc-stan.org/users/citations/>
- Sutherland W., Saunders W., 1992, *MNRAS*, 259, 413
- Taylor A. C., Grainge K., Jones M. E., Pooley G. G., Saunders R. D. E., Waldram E. M., 2001, *MNRAS*, 327, L1
- Tegmark M., de Oliveira-Costa A., 1998, *ApJ*, 500, L83
- Tegmark M., Efstathiou G., 1996, *MNRAS*, 281, 1297
- Thorne B., Dunkley J., Alonso D., Naess S., 2017, *MNRAS*, 469, 2821
- Tucci M., Toffolatti L., de Zotti G., Martínez-González E., 2011, *A&A*, 533, A57
- Vielva P., Martínez-González E., Gallegos J. E., Toffolatti L., Sanz J. L., 2003, *MNRAS*, 344, 89
- Waldram E. M., Pooley G. G., Grainge K. J. B., Jones M. E., Saunders R. D. E., Scott P. F., Taylor A. C., 2003, *MNRAS*, 342, 915
- Waldram E. M., Bolton R. C., Pooley G. G., Riley J. M., 2007, *MNRAS*, 379, 1442
- Wang L. et al., 2014, *MNRAS*, 444, 2870
- Weiland J. L. et al., 2011, *ApJS*, 192, 19
- Wright A. E., Griffith M. R., Burke B. F., Ekers R. D., 1994, *ApJS*, 91, 111
- Wright A. E., Griffith M. R., Hunt A. J., Troup E., Burke B. F., Ekers R. D., 1996a, *ApJS*, 103, 145
- Wright A. E., Griffith M. R., Burke B. F., Ekers R. D., 1996b, *ApJ*, 103, 145
- Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *J. Open Source Softw.*, 4, 1298

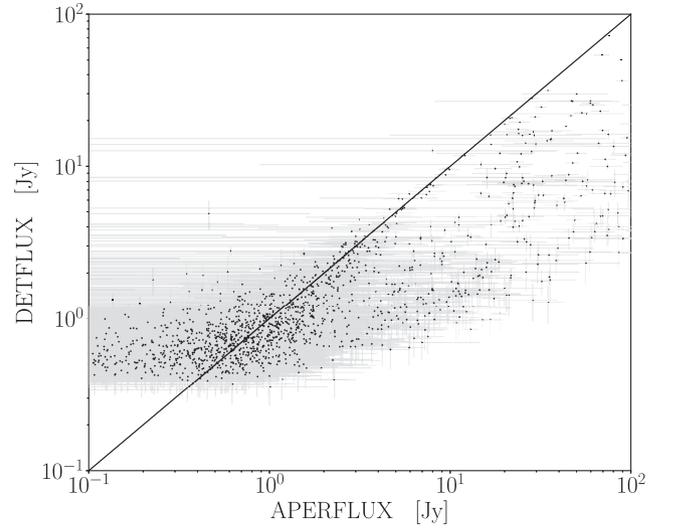
## APPENDIX: C-BASS CATALOGUE SAMPLE

In Table A1, we list the first 10 sources from the C-BASS catalogue. Columns are arranged as

- (i) **SourceID**: The C-BASS source identifier.
- (ii) **RA (deg)**: The source right ascension in degrees (J2000 coordinates).
- (iii) **Dec. (deg)**: The source declination in degrees (J2000 coordinates).
- (iv) ***l* (deg)**: The source Galactic longitude in degrees.
- (v) ***b* (deg)**: The source Galactic latitude in degrees.

**Table A1.** The first 10 sources from the C-BASS northern sky catalogue. In column 1, we give the C-BASS source identifier. In columns 2 and 3, we give the right ascension and declination in degrees, respectively (in J2000 coordinates). In columns 4 and 5, we give the Galactic longitude and latitude in degrees, respectively. In column 6, we give the APERFLUX estimate for the source flux density in Jy. In column 7, we give the uncertainty on the APERFLUX estimate in Jy. In column 8, we give the DETFLUX estimate in Jy, determined directly from the peak value of the source in the SMHW2 filtered map. In column 9, we give the uncertainty on the DETFLUX estimate in Jy. In column 10, we give the detected SNR of the source in the filtered map. This is equivalent to the DETFLUX estimate divided by its corresponding uncertainty. In column 11, we denote whether the C-BASS source was matched with sources in the GB6, PMN, RATAN-600, or CRATES catalogues. G means a source was matched with at least one GB6 source, P means a source was matched with at least one PMN source, R means a source was matched with at least one RATAN-600 source, and C means a source was matched with at least one CRATES source during the GB6/PMN comparisons. A 0 means the C-BASS source was not matched with any source in these ancillary catalogues. In column 12, we denote whether the C-BASS source was inside the CG30 mask. A 1 means the source was inside the CG30 mask and a 0 means the source was outside the CG30 mask.

SourceID	RA (deg) (J2000)	DEC (deg) (J2000)	<i>l</i> (deg)	<i>b</i> (deg)	APERFLUX (Jy)	APERFLUX Error (Jy)	DETFLUX (Jy)	DETFLUX Error (Jy)	SNR	Matches	CG30
CBASS_J000056 + 553448	0.23	+ 55.58	115.79	+ 6.59	0.43	5.23	0.72	0.19	3.8	G	1
CBASS_J000110 + 405055	0.29	+ 40.85	112.78	+ 21.03	0.11	0.93	0.66	0.13	5.2	G	0
CBASS_J000137 + 643717	0.40	+ 64.62	117.62	+ 2.26	6.30	3.84	2.41	0.44	5.4	0	1
CBASS_J000249 + 671313	0.71	+ 67.22	118.24	+ 4.79	127.65	4.29	40.92	0.31	133.2	0	1
CBASS_J000316 + 655840	0.82	+ 65.98	118.05	+ 3.56	32.00	4.03	1.75	0.35	5.1	0	1
CBASS_J000433 + 125050	1.14	+ 12.85	105.56	+ 48.44	1.69	1.20	1.04	0.14	7.3	G	0
CBASS_J000548 + 683650	1.45	+ 68.61	118.77	+ 6.11	42.21	4.16	6.21	0.26	24.1	0	1
CBASS_J000609 + 381223	1.54	+ 38.21	113.23	+ 23.81	1.11	0.93	0.86	0.14	6.0	GC	0
CBASS_J000618 + 722000	1.58	+ 72.33	119.48	+ 9.76	9.26	2.46	2.56	0.17	14.8	G	1
CBASS_J000622-062113	1.59	- 6.35	93.63	- 66.63	1.77	1.06	2.19	0.13	16.8	PC	0



**Figure A1.** DETFLUX estimates plotted against APERFLUX estimates for the C-BASS catalogue. We also plot the line DETFLUX = APERFLUX. For low flux densities ( $S_p \lesssim 500$  mJy), the DETFLUX estimates plateau at the C-BASS flux-density limit. At these low flux densities, the DETFLUX results typically provide more reliable flux-density estimates than the APERFLUX results. This is because sources are detected at greater SNR in the filtered maps, from which the DETFLUX peak estimates are extracted. However, for brighter sources the DETFLUX estimation procedure can miss extended emission, resulting in the flux-density estimates being systematically biased low compared to the APERFLUX estimates. This is more noticeable for sources close to the Galactic plane.

- (vi) APERFLUX (Jy): The source flux-density estimate obtained using aperture photometry.
- (vii) APERFLUX error (Jy): The uncertainty on the APERFLUX estimate.
- (viii) DETFLUX (Jy): The source flux-density estimate obtained from the peak value in the SMHW2 filtered map.
- (ix) DETFLUX Error (Jy): The uncertainty on the DETFLUX estimate.
- (x) SNR: The SNR at which the source was detected in the SMHW2 filtered map.
- (xi) Matched: Column denoting which catalogues contained at least one match to the C-BASS source. G denotes at least one GB6 match, P denotes at least one PMN match, R denotes at least one RATAN-600 match, and C denotes a source that was matched with a CRATES source during the GB6/PMN comparisons. A 0 means the source was not matched with any source in these ancillary catalogues. For example, GP means the source was matched with at least one GB6 source, and at least one PMN source.
- (xii) CG30: This column denotes whether the source was within the CG30 mask. A 1 means the source was inside the CG30 mask (i.e. it was masked for our analysis) and a 0 means the source was outside the CG30 mask.

In Fig. A1, we show the DETFLUX estimates plotted against the APERFLUX estimates for the C-BASS catalogue. The DETFLUX estimates typically provide more reliable flux-density estimates for fainter sources ( $S_p \lesssim 500$  mJy) and sources in regions of brighter diffuse emission. In this situation, the APERFLUX estimates become heavily contaminated by the diffuse background. This can result in the uncertainties on the APERFLUX estimates becoming larger than the APERFLUX estimates, or yield negative APERFLUX estimates due to excessive background subtraction. How-

ever, for brighter sources ( $S_\nu \gtrsim 1$  Jy) the DETFLUX estimates are systematically biased low compared to the APERFLUX estimates. This is due to the DETFLUX estimates missing any extended emission from brighter sources. It is for this reason that we use the APERFLUX estimates in our analysis of the catalogue flux-density scales, with the DETFLUX results provided as auxiliary flux-density estimates. These issues have previously been noted in Planck Collaboration XXVIII (2014) and Planck Collaboration XXVI (2016c), when constructing the PCCS.

We note here that we do not provide results from any Gaussian or PSF fitting at source locations in the C-BASS map. Due to the low

resolution of the C-BASS survey and the bright diffuse emission over much of the sky, fitting for detailed PSF parameters is a major challenge for all but the brightest sources ( $S_\nu \gtrsim 10$  Jy). For the vast majority of sources, it is extremely difficult to achieve convergence during the parameter fitting. We therefore leave this fitting to more detailed analyses of the brightest source emission.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.