

## THE COORDINATED RADIO AND INFRARED SURVEY FOR HIGH-MASS STAR FORMATION. II. SOURCE CATALOG

C. R. PURCELL<sup>1,2,3</sup>, M. G. HOARE<sup>1</sup>, W. D. COTTON<sup>4</sup>, S. L. LUMSDEN<sup>1</sup>, J. S. URQUHART<sup>1,5,6</sup>, C. CHANDLER<sup>7</sup>,  
E. B. CHURCHWELL<sup>8</sup>, P. DIAMOND<sup>2,5</sup>, S. M. DOUGHERTY<sup>9</sup>, R. P. FENDER<sup>10</sup>, G. FULLER<sup>2</sup>, S. T. GARRINGTON<sup>2</sup>, T. M. GLEDHILL<sup>11</sup>,  
P. F. GOLDSMITH<sup>12</sup>, L. HINDSON<sup>5,11</sup>, J. M. JACKSON<sup>13</sup>, S. E. KURTZ<sup>14</sup>, J. MARTÍ<sup>15</sup>, T. J. T. MOORE<sup>16</sup>, L. G. MUNDY<sup>17</sup>,  
T. W. B. MUXLOW<sup>2</sup>, R. D. OUDMAIJER<sup>1</sup>, J. D. PANDIAN<sup>18</sup>, J. M. PAREDES<sup>19</sup>, D. S. SHEPHERD<sup>7,20</sup>, S. SMETHURST<sup>2</sup>,  
R. E. SPENCER<sup>2</sup>, M. A. THOMPSON<sup>11</sup>, G. UMANA<sup>21</sup>, AND A. A. ZIJLSTRA<sup>2</sup>

<sup>1</sup> School of Physics & Astronomy, E.C. Stoner Building, University of Leeds, Leeds LS2 9JT, UK; C.R.Purcell@leeds.ac.uk

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

<sup>3</sup> Sydney Institute for Astronomy (SiFA), School of Physics, The University of Sydney, NSW 2006, Australia

<sup>4</sup> National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903-2475, USA

<sup>5</sup> CSIRO Astronomy and Space Science, P.O. Box 76, Epping, NSW 1710, Australia

<sup>6</sup> Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany

<sup>7</sup> National Radio Astronomy Observatory, Array Operations Center, P.O. Box O, 1003 Lopezville Road, Socorro, NM 87801-0387, USA

<sup>8</sup> The University of Wisconsin, Department of Astronomy, 475 North Charter Street, Madison, WI 53706, USA

<sup>9</sup> National Research Council of Canada, Herzberg Institute for Astrophysics, Dominion Radio Astrophysical Observatory,  
P.O. Box 248, Penticton, British Columbia V2A 6J9, Canada

<sup>10</sup> School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK

<sup>11</sup> Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

<sup>12</sup> Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

<sup>13</sup> Astronomy Department, Boston University, 725 Commonwealth Avenue, Boston, MA 02215, USA

<sup>14</sup> Centro de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México - Morelia, Apartado Postal 3-72, C.P. 58090 Morelia, Michoacan, Mexico

<sup>15</sup> Departamento de Física, EPSJ, Universidad de Jaén, Campus Las Lagunillas s/n, Edif. A3, E-23071 Jaén, Spain

<sup>16</sup> Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD, UK

<sup>17</sup> Department of Astronomy, University of Maryland College Park, MD 20742-2421, USA

<sup>18</sup> Institute for Astronomy, 2680 Woodlawn Drive, Honolulu, HI 96822-1839, USA

<sup>19</sup> Departament d'Astronomia i Meteorologia and Institut de Ciències del Cosmos (ICC),

Universitat de Barcelona (UB/IEEC), Martí Franquès 1, E-08028 Barcelona, Spain

<sup>20</sup> Square Kilometre Array - South Africa, 3rd floor, The Park, Park Road, Pinelands, Cape Town, 7405 Western Cape, South Africa

<sup>21</sup> INAF Osservatorio Astrofisico di Catania, via S. Sofia 78, I-95123 Catania, Italy

Received 2012 January 5; accepted 2012 November 29; published 2013 February 20

### ABSTRACT

The CORNISH project is the highest resolution radio continuum survey of the Galactic plane to date. It is the 5 GHz radio continuum part of a series of multi-wavelength surveys that focus on the northern GLIMPSE region ( $10^\circ < l < 65^\circ$ ), observed by the *Spitzer* satellite in the mid-infrared. Observations with the Very Large Array in B and BnA configurations have yielded a  $1''.5$  resolution Stokes  $I$  map with a root mean square noise level better than  $0.4 \text{ mJy beam}^{-1}$ . Here we describe the data-processing methods and data characteristics, and present a new, uniform catalog of compact radio emission. This includes an implementation of automatic deconvolution that provides much more reliable imaging than standard CLEANing. A rigorous investigation of the noise characteristics and reliability of source detection has been carried out. We show that the survey is optimized to detect emission on size scales up to  $14''$  and for unresolved sources the catalog is more than 90% complete at a flux density of  $3.9 \text{ mJy}$ . We have detected 3062 sources above a  $7\sigma$  detection limit and present their ensemble properties. The catalog is highly reliable away from regions containing poorly sampled extended emission, which comprise less than 2% of the survey area. Imaging problems have been mitigated by down-weighting the shortest spacings and potential artifacts flagged via a rigorous manual inspection with reference to the *Spitzer* infrared data. We present images of the most common source types found: H II regions, planetary nebulae, and radio galaxies. The CORNISH data and catalog are available online at <http://cornish.leeds.ac.uk>.

*Key words:* catalogs – H II regions – radio continuum: general – radio continuum: ISM – surveys – techniques: image processing

*Online-only material:* color figures

### 1. INTRODUCTION

The observed progression of massive star formation, from cold collapsing core to young OB clusters, is largely understood via observations of discrete examples that have been ordered into an evolutionary sequence. Key to separating objects of different age and type are measurements of their spectral energy distributions (SEDs) at submillimeter, infrared, and radio wavelengths.

The *Spitzer* GLIMPSE (Galactic Legacy Infrared Mid-Plane Survey Extraordinaire) programme is the first of a number of sensitive infrared surveys covering the inner Galactic plane at high resolution and in an unbiased manner (Churchwell et al. 2009). The northern half of GLIMPSE covers the region  $10^\circ < l < 65^\circ$ ,  $|b| < 1^\circ$  at wavelengths spanning  $3.6 \mu\text{m}$ – $8.0 \mu\text{m}$ , which preferentially selects warm and dusty embedded sources. The companion *Spitzer* MIPS GAL survey (Carey et al. 2009) has imaged the same region at  $24 \mu\text{m}$  and  $70 \mu\text{m}$  (where

**Table 1**  
Details of the CORNISH Observational Epochs

Epoch	Dates	Decl. Range	$l$ Range	Config.	Notes
I	2006 Jul 12 → 2006 Sep 16	$-10^{\circ}.5 \rightarrow +14^{\circ}.2$	$21^{\circ}.1 \rightarrow 48^{\circ}.9$	B	VLA antennas only, storms
II	2007 Sep 28 → 2007 Oct 6	$-20^{\circ}.8 \rightarrow -14^{\circ}.9$	$10^{\circ}.0 \rightarrow 16^{\circ}.1$	BnA	VLA + EVLA antennas, low Decl.
IIIa	2007 Oct 27 → 2008 Feb 4	$-14^{\circ}.9 \rightarrow -10^{\circ}.5$	$16^{\circ}.1 \rightarrow 21^{\circ}.1$	B	VLA + EVLA antennas
IIIb	2007 Oct 27 → 2008 Feb 4	$+14^{\circ}.2 \rightarrow +29^{\circ}.1$	$48^{\circ}.9 \rightarrow 65^{\circ}.5$	B	VLA + EVLA antennas

**Notes.** The properties of the data differ in the combination of antenna types included in the array, the configuration of the array, the weather experienced and the declination range observed. Unless otherwise noted the weather during the observations was reasonable.

the bulk of the energy from massive young stellar objects is emitted) and is hence sensitive to cooler and more deeply embedded young stellar objects. Most recently, the *Herschel* Infrared Galactic Plane survey (Hi-GAL; Molinari et al. 2010) is delivering the most comprehensive survey of embedded objects to date. With observations in six far-infrared bands between  $70\ \mu\text{m}$  and  $500\ \mu\text{m}$ , Hi-GAL samples the peak of the star-forming SED and covers the northern GLIMPSE region out to  $l = 60$ . Completing the infrared picture of Galactic star formation is the UKIDSS<sup>22</sup> project (UK IR Deep Sky Survey; Lawrence et al. 2007). A subset of UKIDSS, the Galactic Plane Survey (Lucas et al. 2008), has observed the northern GLIMPSE region in the near-infrared  $J$ ,  $H$ , and  $K$  bands and is sensitive to objects down to 18th magnitude. The combined data from these surveys are driving the detailed characterization of the Galactic population via their infrared colors (e.g., Robitaille et al. 2007; Arvidsson et al. 2010; Smith et al. 2010; Wright et al. 2010; Mottram et al. 2011). A complementary picture of the molecular and atomic interstellar medium is being provided by the BU-FCRAO Galactic Ring Survey for CO (Jackson et al. 2006) and the VLA Galactic Plane Survey (VGPS) for H I (Stil et al. 2006). Similarly, the ongoing Isaac Newton Telescope Photometric Survey of the Northern Galactic Plane (IPHAS; Drew et al. 2005) probes  $H\alpha$  in emission toward nebulae, and in both absorption and emission toward stars. The UKIRT Wide Field Infrared Survey for  $H_2$  (UWISH2; Froebrich et al. 2011) also covers the same GLIMPSE region in molecular hydrogen ( $2.122\ \mu\text{m}$  line) highlighting regions of shocked or fluorescently excited molecular gas ( $T \approx 2000\ \text{K}$ ,  $n_{H_2} > 10^3\ \text{cm}^{-3}$ ).

Conspicuous by its absence is a comparable radio continuum survey for compact ionized gas. Previous surveys are either targeted at individual sources selected via infrared colors (e.g., Wood & Churchwell 1989; Kurtz et al. 1994; Urquhart et al. 2009) or are limited in their resolution and sky-coverage (e.g., Becker et al. 1994; White et al. 2005). From a star formation perspective, the presence or absence of free-free emission is vital to distinguish the more evolved ultra-compact H II (UCH II) regions from their younger counterparts with similar thermal SEDs (Urquhart et al. 2009, 2011). The sheer number density of sources in the near and mid-infrared surveys necessitates complementary data at similarly high resolution to enable the full science potential to be fulfilled. This is particularly true in highly clustered star forming regions. It is important that any radio continuum survey for UCH II regions be carried out at relatively high frequencies ( $\geq 5\ \text{GHz}$ ) where thermal free-free emission is optically thin with a spectral index of  $S_\nu \propto \nu^{-0.1}$ . At lower frequencies the spectrum becomes optically thick with  $S_\nu \propto \nu^2$ . High-frequency observations hence confer a signal-to-noise advantage and probe the structure of the ionized gas at all depths in UCH II regions. We note that even at  $\nu = 5\ \text{GHz}$

we will be insensitive to a population of young and compact H II regions: the so-called hyper-compact H II (HCH II) regions (see Sewilo et al. 2011 and references therein). These objects have greater emission measures than UCH II regions and the turnover frequency from optically thick to thin occurs at high frequencies.

No previous radio survey of the Galactic plane has similar resolution and coverage to the *Spitzer* GLIMPSE survey. A number of single dish surveys have been conducted at 5 GHz (e.g., Altenhoff et al. 1979), however, their arcminute resolution is quite low compared to the arcsecond resolution of *Spitzer*. Most interferometric surveys have been carried out at a frequency of 1.4 GHz (e.g., the NRAO VLA Sky Survey; Condon et al. 1998) except for the catalogs of Becker et al. (1994), Gieveon et al. (2005), and White et al. (2005), who surveyed the inner Galactic plane ( $-10^\circ < l < 42^\circ$ ,  $|b| < 0.4^\circ$ ) at 5 GHz. These three surveys are published as the Multi-Array Galactic Plane Imaging Survey (MAGPIS).<sup>23</sup> They used the Very Large Array (VLA) in C and D configurations, which deliver a relatively large beam ( $4'' \times 9''$ ), and the total survey area only covers 26% of the northern GLIMPSE region.

The CORNISH (Coordinated Radio “N” Infrared Survey for High-mass star formation) project delivers a uniform, sensitive, and high-resolution radio survey of the northern GLIMPSE region to address key questions in high-mass star formation as well as many other areas of astrophysics. In addition to UCH II regions, the CORNISH survey detects many other radio-bright objects, including planetary nebulae, ionized winds from evolved massive stars, non-thermal emission from active stars, active galactic nuclei (AGNs), and radio galaxies. The full rationale behind the survey design and the scientific motivation are presented in an accompanying paper (Hoare et al. 2012).

## 2. OBSERVATIONS

CORNISH covers the  $110\ \text{deg}^2$  of the northern GLIMPSE region ( $10^\circ < l < 65^\circ$ ,  $|b| < 1^\circ$ ) using the VLA in B and BnA configurations at 5 GHz. The combination of array configuration and observing frequency results in a  $\sim 1''.5$  synthesized beam within a  $8.9$  field of view, corresponding to the FWHM primary beam. With a total integration time of 80 s per pointing, the root mean square (rms) noise in the images is better than  $0.4\ \text{mJy beam}^{-1}$ —sufficient to detect an unresolved UCH II region around a B0 star on the far edge of the Galaxy (16 kpc; Kurtz et al. 1994).

CORNISH observations of the northern GLIMPSE region were conducted using the VLA during the 2006 and 2007/2008 observing seasons. The observations fall naturally into the epochs presented in Table 1, which are distinguished by the combinations of array configuration used, inclusion or exclusion

<sup>22</sup> <http://www.ukidss.org>

<sup>23</sup> <http://third.ucllnl.org/gps/>

of upgraded VLA antennas, declination ranges observed, and weather conditions experienced. We show later that data from each epoch have unique properties.

To facilitate scheduling the target area was divided into 42 blocks, each corresponding to 8 hr of observations per day. Blocks contain between 180 and 220 fields arranged in rows of equal right ascension on a hexagonal pointing grid. Individual fields were observed as two 45 s “snapshots” separated by  $\sim 4$  hr in time, maximizing the  $uv$  coverage and minimizing the elongation of the synthesized beam. The telescope was advanced along each row ( $\sim 20$  fields) integrating for 45 s on each pointing position, before observing a secondary calibrator (one of 1832–105, 1856+061 or 1925+211) for two minutes and then continuing to the next row. Including overheads, the secondary calibrators were observed with a cadence of 20 minutes. Fields at declinations greater than  $-15^\circ$  were observed using the VLA’s B configuration while fields at lower declinations were observed using the BnA configuration, which is designed to compensate for beam distortion at low elevations.

To allow imaging of the widest possible field of view without bandwidth-smearing the observations were carried out in pseudo-spectral line mode. The two 25 MHz wide spectral windows (also known as intermediate frequencies or IFs) of the VLA correlator were tuned to adjoining frequency bands centered on 5 GHz. Each window was sampled by eight 3.1 MHz channels, degrading the peak response by only a few percent at the edge of the 8.9 primary beam. Due to hardware limitations only the RR and LL polarizations were recorded, meaning that linear polarization information is not available in the CORNISH data.

During both CORNISH observing seasons significant engineering works were underway to upgrade the VLA to the next generation instrument: the Karl G. Jansky VLA (JVLA). In 2006 between two and six antennas were missing from the array as they were being refurbished with new receivers and electronics to convert them to the JVLA design. By the start of the second season of CORNISH observations (2007 September), almost half the array was comprised of JVLA antennas and the instrument was operating in a transition mode. Over the season VLA antennas were progressively removed from the active array and substituted with JVLA antennas. The JVLA antennas conferred the advantage of enhanced sensitivity, but at the same time were untested and prone to software and hardware problems. Special care was needed to properly calibrate VLA–JVLA baselines and to ensure that the JVLA data were properly flagged. As part of the upgrade, the venerable Modcomp-based VLA control systems were also replaced in mid-2007 with new software running under Linux. Taken as a whole, the CORNISH data required close inspection and vigilance during post-processing.

### 3. THE DATA REDUCTION PIPELINE

The raw CORNISH data set consists of 9349 pointing positions, each of which was observed twice. A manually guided data-reduction procedure was considered too labor-intensive to use on such a large volume of data, hence a semi-automatic pipeline was developed with the control parameters tuned to the average observation. This approach has the advantage of applying uniform processing over the majority of the survey area, while still allowing manual intervention in a minority of special cases (e.g., fields with complicated emission structures, or very bright sources).

The pipeline was implemented in the *python* language and made use of the ObitTalk module to interface directly with

the NRAO<sup>24</sup> AIPS<sup>25</sup> and OBIT<sup>26</sup> data-reduction packages. The CORNISH pipeline utilized a MySQL database to record metadata and perform bookkeeping operations during the reduction procedure.

Figure 1 illustrates the pipeline logic, which is broken up into calibration and imaging stages. In the following sections we describe each of the stages in detail.

#### 3.1. Calibration and Flagging

Raw data from the telescope were corrected for atmospheric opacity using phase monitor data and written to an AIPS-format  $uv$  file in spectral-line mode by the AIPS task FILLM. Each 8 hr block of observations was first inspected by eye and  $uv$  visibilities with large phase scatter, errant amplitudes, or system-temperature spikes were flagged out of the data set. Gross errors in the data, such as bad antennas, IFs, or polarizations were also flagged out at this stage. It was necessary to edit out the first 5 s of data from each pointing to allow for antenna settling time, reducing the on-source integration time from 45 s to 40 s. All manual flagging parameters were written to a master flag list, which was automatically applied upon restarting the pipeline. Care was taken here that the primary flux calibrators contained only good data.

The shapes of the VLA and JVLA pass-bands are different enough that a 6% closure error has been measured on JVLA–VLA baselines in continuum modes using 50 MHz bandwidths.<sup>27</sup> This error is expected to be larger at narrower bandwidths. Because we are operating in pseudo-spectral line mode the issue was mitigated by performing bandpass calibration (phase and amplitude) immediately after the initial flagging and before any further calibration. Solutions for the atmospherically and electronically induced changes in phase and amplitude were then calculated using the standard AIPS CALIB task, operating on one of the three secondary calibrators. The data were bootstrapped on to an absolute flux scale by comparing observations of the quasars 1331+305 (3C286) or 0137+331 (3C48) to their C-band model in AIPS. A global calibration table was produced, which could then be applied to the whole block.

After a first pass at calibration the OBIT flagging tasks AutoFlag and MednFlag were applied to each IF, polarization, and channel of the secondary calibrator observations. AutoFlag edits out bad visibilities based on an absolute maximum allowed value in Stokes  $I$  or  $V$ . Radio-frequency interference (RFI), e.g., from commercial broadcasting, is often highly polarized and all visibilities with Stokes  $V$  amplitude greater than 2 Jy were flagged as bad. MednFlag applies a rolling median filter to each IF and spectral channel. Visibilities were edited out if they had Stokes  $I$  values greater than five standard deviations from the median, calculated in a  $\sim 50$  s time-window. A second pass at calibration was then performed before applying a similar flagging procedure to the calibrated science observations on a field-by-field basis. Finally, all calibration and flagging tables were applied to the data and the individual forty-second pointings were split into  $uv$  FITS format files.

Metadata associated with each observation (e.g., the pointing center coordinates and number of flagged visibilities) were automatically saved in the MySQL database. The imaging

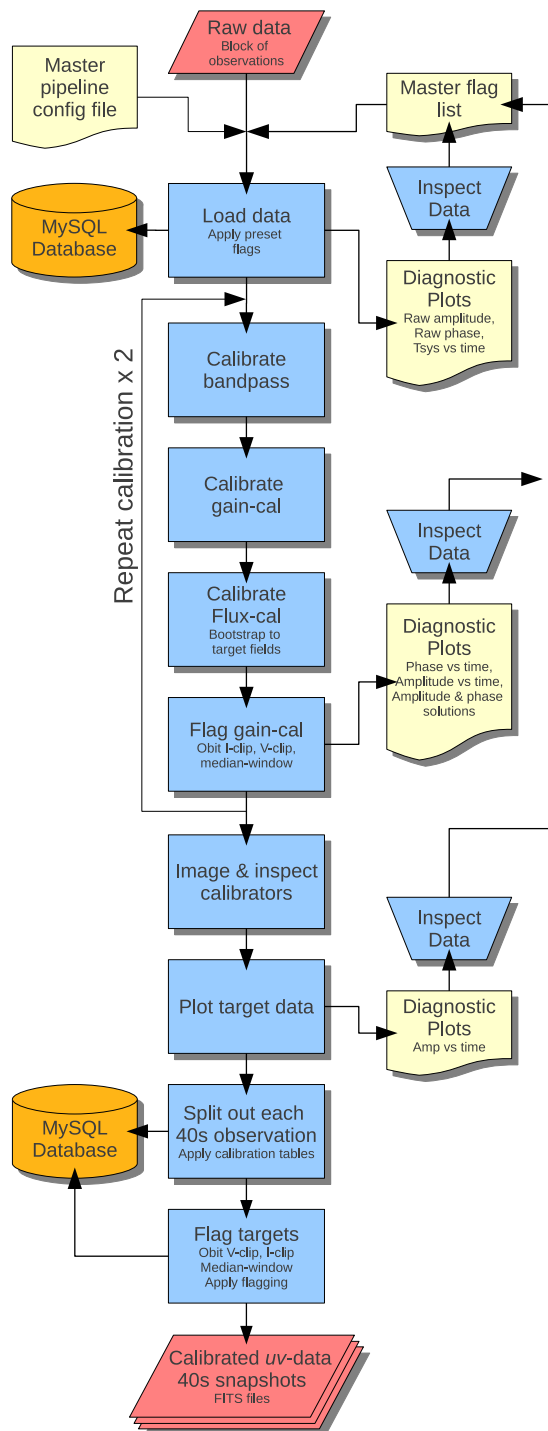
<sup>24</sup> The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

<sup>25</sup> <http://www.aips.nrao.edu/>

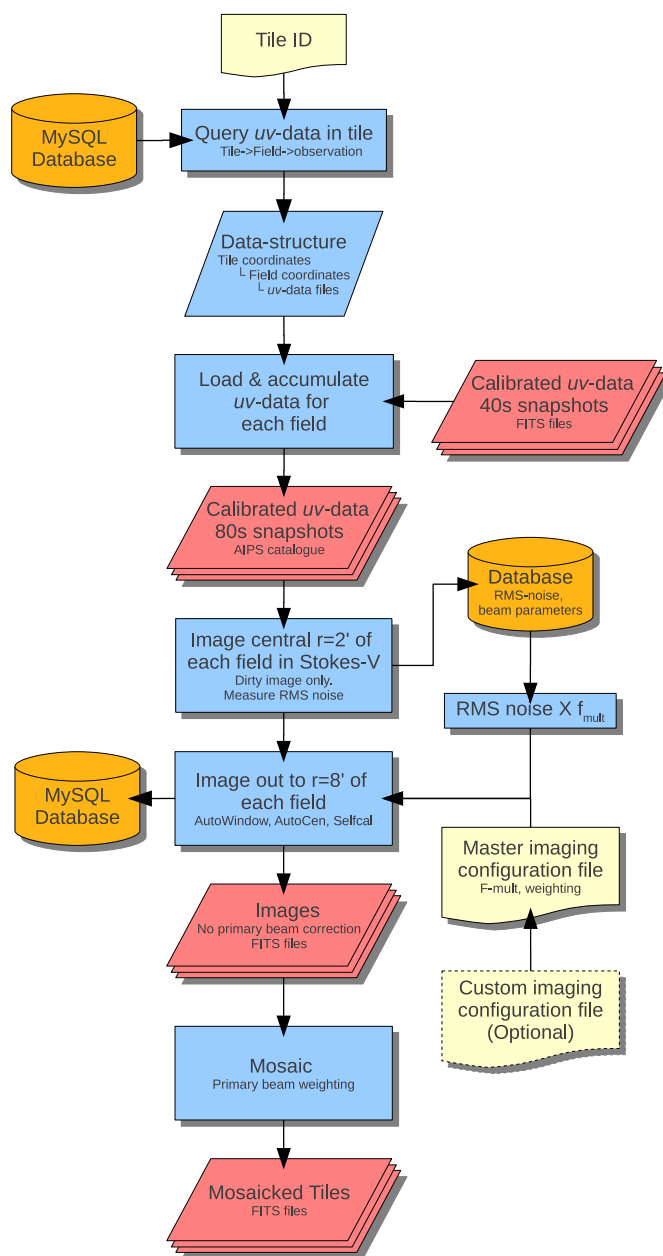
<sup>26</sup> <http://www.cv.nrao.edu/~bcotton/Obit.html>

<sup>27</sup> See <http://www.vla.nrao.edu/astro/guides/evlaretturn/> for details.

Calibration:



Imaging:



**Figure 1.** Graphical illustration of the CORNISH calibration and imaging pipeline. See the text of Section 3 for more detail. (A color version of this figure is available in the online journal.)

procedure subsequently queried this database when building a final mosaicked image.

3.2. Imaging

Fields were imaged using the OBIT Imager task, which performs imaging and deconvolution in a similar manner to the AIPS IMAGR routine. Imager automatically switches between the standard Cotton–Schwabb and SDI deconvolution algorithms (Schwab 1984; Steer et al. 1984), referred to sim-

ply as “CLEAN” in the following discussion. In addition, the task can be instructed to perform both phase and amplitude self-calibration. Imager is a complex task with many important input parameters that need careful tuning to result in a scientifically useful image. Key among these are the maximum residual flux, the threshold at which self-calibration begins, whether to perform both phase and amplitude self-calibration, and the weighting function used. The “average” CORNISH field was imaged using a Briggs robustness parameter of zero, which is

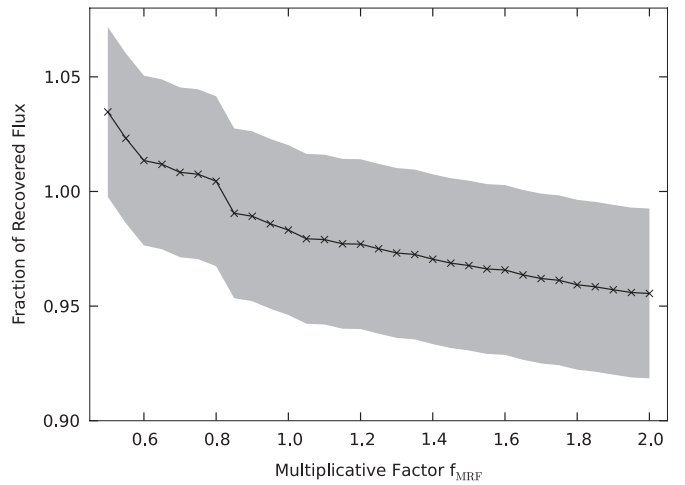
a compromise between natural (high-sensitivity) and uniform (high-resolution) weighting. A minority of complex fields were treated as a special case and imaged with a custom weighting scheme; see Section 3.2.2 below. Self-calibration was performed on sources with peak fluxes greater than  $30 \text{ mJy beam}^{-1}$ . During the deconvolution process the restoring Gaussian beam was forced to be circular and have an FWHM of  $1''.5$ . The cell size was set at  $0''.3$ , over-sampling the synthesized beam. We justify these choices in the following sections.

Imager implements two algorithms that improve the dynamic range and quality of the final image. First, the AutoWindow<sup>28</sup> function dynamically places small ( $<20$  pixels in radius) clean windows over regions of emission in the intermediate dirty map. The validity of each window is assessed periodically during the clean cycle and additional windows are created as necessary. The net effect is to CLEAN only real emission and avoid CLEANing noise, resulting in a smaller clean bias. The AutoCen<sup>29</sup> algorithm regrids  $uv$  data containing bright point sources, so their peaks fall on a pixel center. A point source at the center of a pixel can be represented as a single delta function, i.e., a single clean component, leading to a significant improvement in dynamic range. In contrast, a bright point-source offset from a pixel center requires multiple clean components, both positive and negative, to model its emission. This more complex model will inevitably suffer from rounding errors on finite-precision computers, leading to flux being scattered into the surrounding sky. The JVLA memos numbers 116 and 114 by Bill Cotton contain detailed descriptions of the AutoWindow and AutoCen algorithms.

### 3.2.1. Controlling the Deconvolution Algorithm

One of the most critical control parameters for the imaging task is the target maximum residual flux (MRF). The ideal value varies from field to field depending on the weather conditions, individual antenna system temperatures, and the structure and strength of emission in the field of view. To estimate the intrinsic sensitivity attainable we imaged each field in Stokes  $V$  and measured the rms noise. Sources with significant circularly polarized emission at 5 GHz are rare (Roberts et al. 1975; Homan & Lister 2006), so the rms noise measured from an *uncleaned* Stokes  $V$  image is expected to be comparable to the final CLEANed noise level. Because the two polarization beams of the VLA are not co-aligned on the sky they give rise to a strong instrumental polarization away from the pointing center. To compensate for this “beam squint”<sup>30</sup> only the central  $2'$  diameter portion of each field was imaged and measured. Despite this we expect noise in the Stokes  $V$  images to be higher than the ideal in their Stokes  $I$  counterparts. In addition, the  $1''.5$  restoring beam applied to the  $I$  images is larger, on average, than the unconstrained synthesized beam of the  $V$  images. This effectively smooths the noise in the Stokes  $I$  images compared to the  $V$ .

The target MRF was assumed to be equal to  $\text{rms}_V \times f_{\text{MRF}}$ , where  $f_{\text{MRF}}$  is a constant multiplicative factor. The canonical value for  $f_{\text{MRF}}$  was determined in two ways. First, artificial point sources were injected into the  $uv$  data for an emission-free field and the OBIT Imager task was applied using a range of values for  $f_{\text{MRF}}$ . After each iteration the recovered flux and rms noise of the final image was measured. Figure 2 shows



**Figure 2.** Fraction of recovered flux vs. multiplicative factor  $f_{\text{MRF}}$  for an artificial source injected into the empty field 18173–18247. The target maximum residual flux (MRF) driving the deconvolution procedure was set to  $\text{rms}_V \times f_{\text{MRF}}$ , where  $\text{rms}_V$  is the root mean square noise measured in Stokes  $V$  dirty map. The amplitude of the rms noise in the image expressed as a fraction of the recovered flux is illustrated by the gray shaded area. For factors below  $\sim 0.8$  the field is over-cleaned, leading to significant artifacts in the image. Values less than 2.0 recover greater than 96% of the flux.

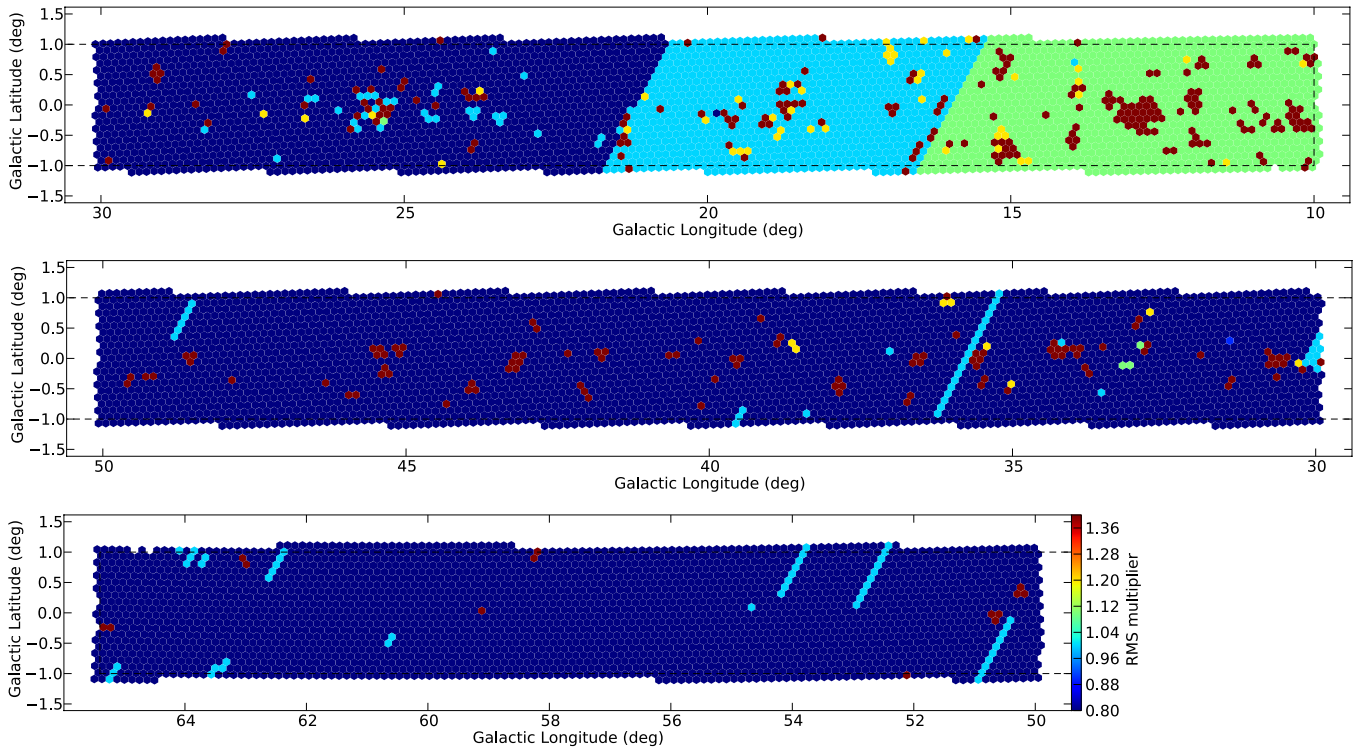
the results of this experiment. We found in practice that values of  $f_{\text{MRF}} = 0.8\text{--}2.0$  recovered greater than 96% of the flux, within the errors. Second, we chose a representative sample of compact sources with simple morphologies and inspected the deconvolved images for residual sidelobe structure. Values of  $f_{\text{MRF}} = 0.8\text{--}1.0$  were required to fully remove sidelobe structure from the images. We reimaged fields using a higher threshold where obvious clean artifacts were present. Figure 3 shows the values of  $f_{\text{MRF}}$  used across the survey. Most of the CORNISH area was imaged using an  $f_{\text{MRF}} = 0.8$ . Fields at lower declinations (including the BnA observations) required higher  $f_{\text{MRF}}$  values to avoid producing increased numbers of low-level artifacts in otherwise empty regions. Unresolved weak detections are dominated by the extragalactic population and hence have a flat distribution on the sky (Anglada et al. 1998; see also Sections 5.4 and 6.2.2). Based on an initial pass at the data reduction we used two further levels of  $f_{\text{MRF}} = 1.0$  and 1.1, chosen to keep the number of  $5\sigma\text{--}6\sigma$  point sources roughly constant away from the Galactic mid-plane. However, multiplier values between 0.8 and 1.1 lead to imaging artifacts in a minority of fields affected by poor calibration, containing very bright point sources ( $>1 \text{ Jy}$ ) or extended emission. Such fields were inspected and cleaned manually. A small proportion of fields ( $\ll 1\%$ ) were found to contain significant circularly polarized flux and were also cleaned by hand. It is likely that this emission is due to the instrumental beam squint, rather than real emission on the sky. The OBIT Imager task does not correct for beam squint, however, tests on selected CORNISH data did not find any believable Stokes  $V$  after a correction had been applied (B. Cotton 2011, private communication). The manually cleaned fields are mostly confined to high-mass star formation regions. They appear on Figure 3 as patches of red hexagons clustered around the mid-plane of the Galaxy. Values used in these cases ranged over  $1.2 \leq f_{\text{MRF}} \leq 10$ , with a mean of 3.2.

Applying the deconvolution algorithm close to the noise can result in an increase in the so-called “clean bias,” an effect that results in a systematic reduction in object fluxes. It is believed to be caused by inadvertently CLEANing bright sidelobes, leading

<sup>28</sup> <http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo116.pdf>

<sup>29</sup> <http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo114.pdf>

<sup>30</sup> <http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo113.pdf>



**Figure 3.** Map of the multiplier  $f_{\text{MRF}}$  used to control the deconvolution process on each field. Fields at lower declinations required higher values of  $f_{\text{MRF}}$  to compensate for the decrease in sensitivity which accompanied the lower observing elevations. Individual fields containing bright or complex structured emission also required higher  $f_{\text{MRF}}$  values. Partial rows of fields (in R.A.) with different multipliers compared to their neighbors are due to repeat observations from later epochs. The row at  $\delta = -16^{\circ}55'03''.13$  ( $13^{\circ}:3 < l < 14^{\circ}:5$ ) was observed only once and the  $f_{\text{MRF}}$  was increased to avoid over-cleaning. (A color version of this figure is available in the online journal.)

to a subtraction of real flux from astronomical objects (White et al. 1997; Condon et al. 1998). We measure the clean bias for CORNISH images in Section 5.6.

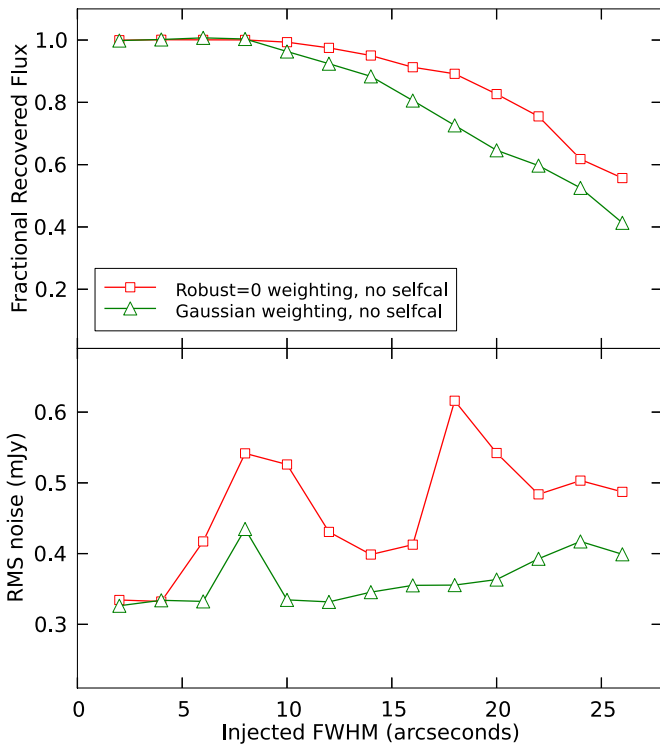
### 3.2.2. Imaging Extended Emission

Baseline lengths on the VLA B and BnA arrays range from approximately  $300 \text{ k}\lambda$  to  $2 \text{ k}\lambda$ , equivalent to spatial scales of  $1''.5$  to  $2''$ , respectively. These array configurations sample the  $uv$  plane less well at shorter spacings and the deconvolution algorithm has difficulty reconstructing image structure on scales greater than  $\sim 14''$ . The imaging procedure tends to produce “waves” or “ruffles” in the background of fields containing significant extended emission. Flux is also scattered over the image as the standard CLEAN algorithm attempts to model the emission as a series of delta functions. This can lead to high rms noise levels and multiple imaging artifacts, especially if self-calibration is allowed to run unchecked. A total of 193 fields ( $\sim 2\%$ ) were found to have poorly imaged extended emission. To combat this problem we imaged these fields using a custom weighting scheme. By default the CORNISH pipeline is configured to use robust weighting (Briggs 1995), which is a compromise between the low thermal noise of natural weighting and the high resolution of uniform weighting. In the case of uniform weighting the OBIT Imager weights each visibility by the sum of the weights present in each cell in the  $uv$  plane. For fields with extended emission we have instead weighted by the inverse of the number of visibilities within a radius of 10 cells, attenuated by a Gaussian function. For the B and BnA arrays this has the effect of weighting down the poorly sampled short spacings, similar to the effect of applying an inverse taper. We found that the rms noise and number of artifacts in the imaged

fields are reduced at the expense of additional “missing” flux. The reduction in flux compared to a uniform, or robust, weighted image is highly dependent on the structure of the emission.

In order to quantify the effect of the two weighting schemes (robust = 0 and Gaussian) we imaged artificial Gaussian sources of increasing size inserted into the  $uv$  data for a blank field. The peak flux was fixed at  $1.0 \text{ Jy beam}^{-1}$  while the FWHM was increased from  $2''.0$  to  $26''$  in steps of  $2''$ . Figure 4 (top panel) plots the recovered flux as a function of the source FWHM for both weighting schemes. When imaging using robust weighting, the fraction of recovered flux drops off above FWHMs greater than  $\sim 8''$ . When using the Gaussian-smoothed weighting scheme this drop-off also occurs at  $\text{FWHM} \geq 8''$ , however, the fraction of recovered flux falls more rapidly. The bottom panel of Figure 4 shows the rms noise as a function of injected source size. In this case the flux density was fixed at  $1.0 \text{ Jy}$  to avoid being dynamic-range limited at higher fluxes. It is clear that the rms noise in the Gaussian weighted images is significantly lower and more stable as a function of FWHM. In practice dynamic ranges of several thousand are achieved on isolated point sources, falling to several hundred for slightly resolved sources ( $> 1.8 \text{ arcsec}$ ).

Figure 5 shows the effect of the two weighting schemes on the fidelity of the images. In the top panel is plotted the absolute fractional value of the residual flux remaining after the model image is subtracted from the pipeline imaged data ( $|S_{\text{resid}}|/S_{\text{model}}$ ). Lower numbers mean that the image is similar to the model, while higher numbers mean that there are significant structural differences. It can be seen from the plot that the Gaussian weighting scheme is the most consistent at representing the source morphology, while the robust weighted images break down between  $\text{FWHM} = 8''$  and  $10''$ . To further



**Figure 4.** Top panel: fractional recovered flux as a function of FWHM for an artificial  $1 \text{ Jy beam}^{-1}$  peak Gaussian source injected into an empty field. Data imaged with a Gaussian smoothed weights is missing greater fractional flux compared to data imaged with robust = 0 weighting. Bottom panel: rms noise as a function of FWHM for  $1 \text{ Jy flux density}$  Gaussian sources. The images made using the Gaussian smoothed weights tend to have lower and more stable noise properties.

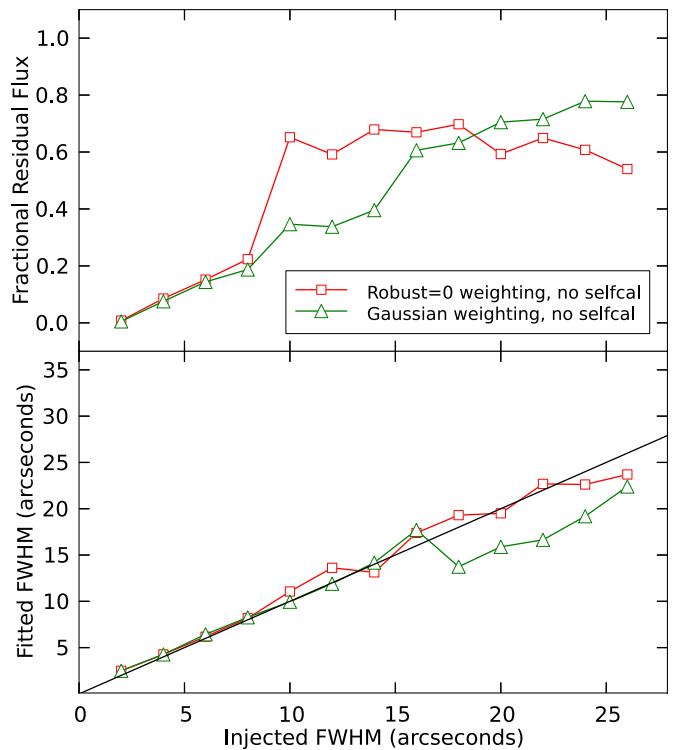
(A color version of this figure is available in the online journal.)

quantify the effect we fit the pipeline imaged data with a 2D-Gaussian using the MIRIAD task imfit. The fitted versus injected FWHM is plotted in the bottom panel of Figure 5. It is again clear that the robust weighted images begin to differ from the model at  $\text{FWHM} \approx 10''$ , while the Gaussian weighting scheme preserves structures out to  $\sim 14''$ . Note that real-world emission with complex morphology will react differently to the Gaussian weighting scheme, depending on its visibility function.

### 3.3. Mosaicking

All fields were imaged out to a radius of 8 arcmin ( $\sim 10\%$  power pattern) before being linearly mosaicked in the image plane onto  $20' \times 20'$  tiles oriented in equatorial (J2000) coordinates. In total 1408 tiles cover the survey area, each of which overlap by  $1'$ . Pointing centers sit on a close-packed hexagonal grid adapted from the 1.4 GHz NVSS survey and scaled to 5 GHz. Condon et al. (1998) justify this layout in detail and Hoare et al. (2012) describe the implementation in CORNISH. Here we provide a summary for convenience. Adjacent CORNISH pointing centers are separated by  $7.4$ , compared to the  $8.9$  FWHM of the primary beam at 4.86 GHz. The separation is optimized to maximize the uniformity of the noise pattern without appreciably degrading observing efficiency. At any point in the mosaic the sky brightness  $B$  is given by a weighted sum of the individual brightness values  $b_i$  contributed by the overlapping snapshots

$$B = \frac{\sum_{i=1}^m W_i b_i}{\sum_{i=1}^m W_i^2}. \quad (1)$$



**Figure 5.** Top panel: fractional absolute residual flux remaining after subtracting a model Gaussian source from the test image. Low values mean that the image is similar to the model while high values mean that there are significant differences in morphology. Bottom panel: fitted vs. injected FWHM. Sources with  $\text{FWHM} > 14''$  are poorly imaged by the VLA B array.

(A color version of this figure is available in the online journal.)

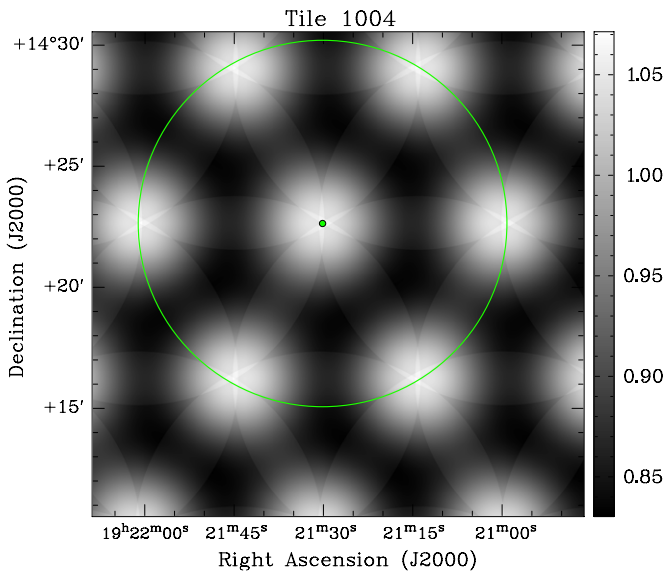
To maximize sensitivity the weighting factor  $W_i$  was set to be proportional to  $P(\rho)$ , the primary beam pattern as a function of offset  $\rho$  from the pointing center. This correction is necessary as the noise is constant across a raw snapshot image and must be weighted by the square of the signal-to-noise ratio. The weighting method is implemented in the CORNISH pipeline in two steps. Individual fields are first multiplied by  $P(\rho)$  and summed onto a blank tile. This image is then divided by a “weight image” created from the sum of  $P^2(\rho)$  functions (modeled by Gaussians for the VLA). The resultant data product is a mosaicked image which has been primary beam corrected (i.e., divided by  $P(\rho)$ ). Figure 6 illustrates an example of a weight image. The minimum weight is  $P^2(\rho) = 0.83$ , hence the worst-case relative-sensitivity is  $\sqrt{P^2(\rho)} = 0.91$ .

## 4. DATA QUALITY

Data were reduced and imaged for quality control purposes immediately after the observations were completed. Bad data were quickly identified allowing the affected fields to be rescheduled in the observing queue. The rapid turnaround time meant that we were able to reobserve most fields affected by poor weather in 2006 and system power glitches in 2007/2008.

### 4.1. Calibration

Three quasars, spaced equally along the plane of the Galaxy, were used as secondary (phase) calibrators for the whole survey. Although initially assumed to be point-like, we found that each exhibited structure at the 0.1%–2.0% level, in the form of radio jets and nearby confusing sources. Using the full complement



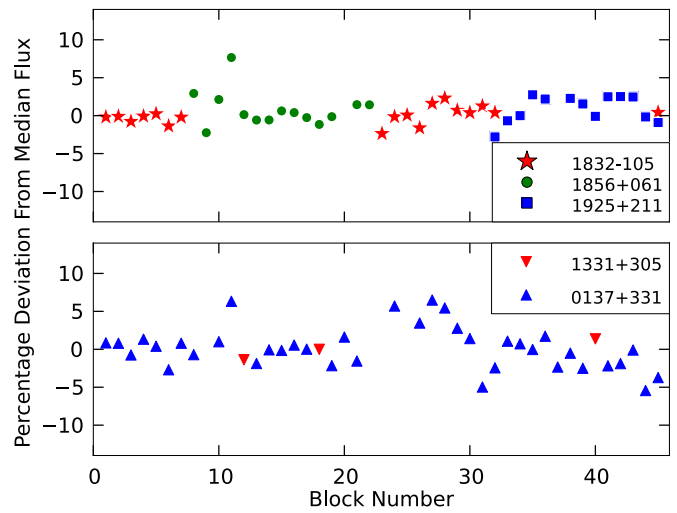
**Figure 6.** Example of a weight image used in the mosaicking process. The gray scale depicts the squared beam patterns  $P^2(\rho)$  accumulated onto a  $20' \times 20'$  tile. The green circle shows the extent of a single field and the dot its pointing center.

(A color version of this figure is available in the online journal.)

of data available we imaged and self-calibrated each secondary calibrator field out to the full-width half-power radius. The resulting clean-component models were used as inputs to the calibration procedure. Images of the secondary calibrators are presented in Figure 7. Quasars 1832–105 and 1856+061 deviate from point sources, exhibiting jets with flux densities peaking at 2% of the main peak. The source 1925+211 shows significant structure within 1 arcmin of the central source, including an elongated jet and two point sources of 1 mJy and 6 mJy (0.1% and 0.4% of the main peak, respectively).

Two primary flux density calibrators were observed, providing a redundant means of flux-calibrating the data. 1331+305 (3C286) was observed at the beginning of an observing block and 0137+331 (3C48) at the end. With a 5 GHz flux density of 7.47 Jy 1331+305 was the preferred calibrator. However, 0137+331 ( $S_{5\text{GHz}} = 5.48$  Jy) was used if technical or weather-related problems affected the initial data from a block.

The small number of calibrators observed allowed us to check the consistency of our calibration with time. Figure 8 shows the percentage deviation from the median flux densities



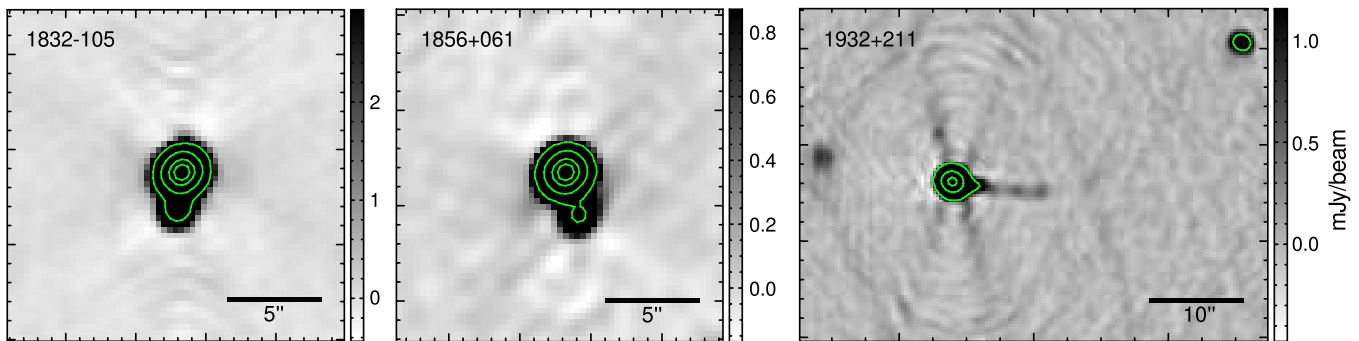
**Figure 8.** Top panel: plot of the percentage deviation from the median flux density vs. block number for the three secondary calibrators. Each point represents the integrated flux density measured from an 8 hr observing block. The scatter is mostly within 5%. Bottom panel: the same plot as above except for the backup primary calibrator.

(A color version of this figure is available in the online journal.)

of the three secondary calibrators and the backup primary calibrator. Each point on the plot represents an 8 hr block of observations. Calibrator flux densities were measured directly from the image data by manually drawing a polygon around each quasar and summing the flux within the polygon. For the secondary calibrators the standard deviation in flux is 2.7%, and for the backup calibrators 8.9%, consistent with the accuracy of previous VLA surveys (e.g., Condon et al. 1998, who quote 3% at 1.4 GHz). No variation with time is seen, implying that the calibration is stable over the two observing seasons. The scatter in the backup calibrator is a more appropriate error to quote for snapshot imaging and is adopted as the formal amplitude calibration error for CORNISH data.

All CORNISH observations are phase-referenced to one of the three secondary calibrators and hence adopt their positional uncertainties. The formal positional uncertainties may be found in the VLA calibrator manual<sup>31</sup> and are <150 mas for 1832–105, <10 mas for 1856+061 and <2 mas for 1925+211.

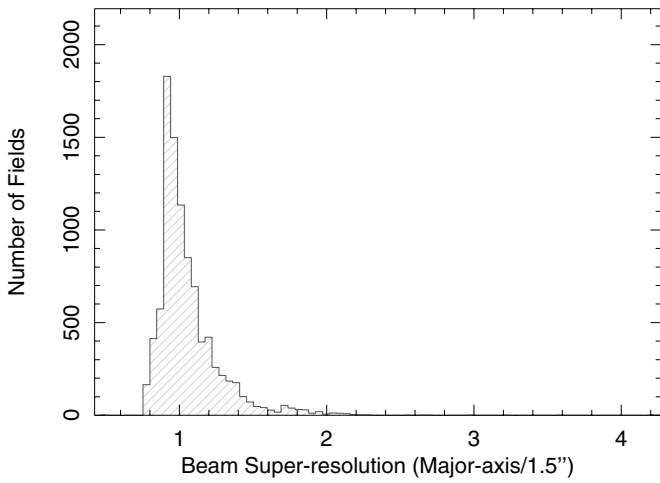
<sup>31</sup> <http://www.vla.nrao.edu/astro/calib/manual/>



**Figure 7.** Images of the secondary calibrators used to perform phase tracking. All three quasars exhibit jets, while 1925+211 has an extended jet and there are several sources brighter than 1 mJy within 1 arcmin. The image scales have been stretched to show all real emission, but also highlight very low level imaging artifacts. Only clean components from real emission were used when calibrating the data.

(A color version of this figure is available in the online journal.)





**Figure 9.** Histogram showing the distribution of super-resolution caused by forcing a 1.5'' circular restoring beam for all pointing positions.

#### 4.2. Synthesized Beam Shape

The dual-snapshot observing scheme was designed to deliver the most circular synthesized beam possible, while allowing both snapshots to be taken within a single 8 hr observing block. To minimize the total range of synthesized beam shapes in the survey each field should ideally be observed at an equal  $\pm 3$  hr hour-angle before and after its zenith position. Scheduling constraints meant that this was not achieved in practice and a compromise of 4 hr between snapshot images was implemented. Hoare et al. (2012) presents the parameters of the synthesized beams attained in the final images, which we briefly summarize here.

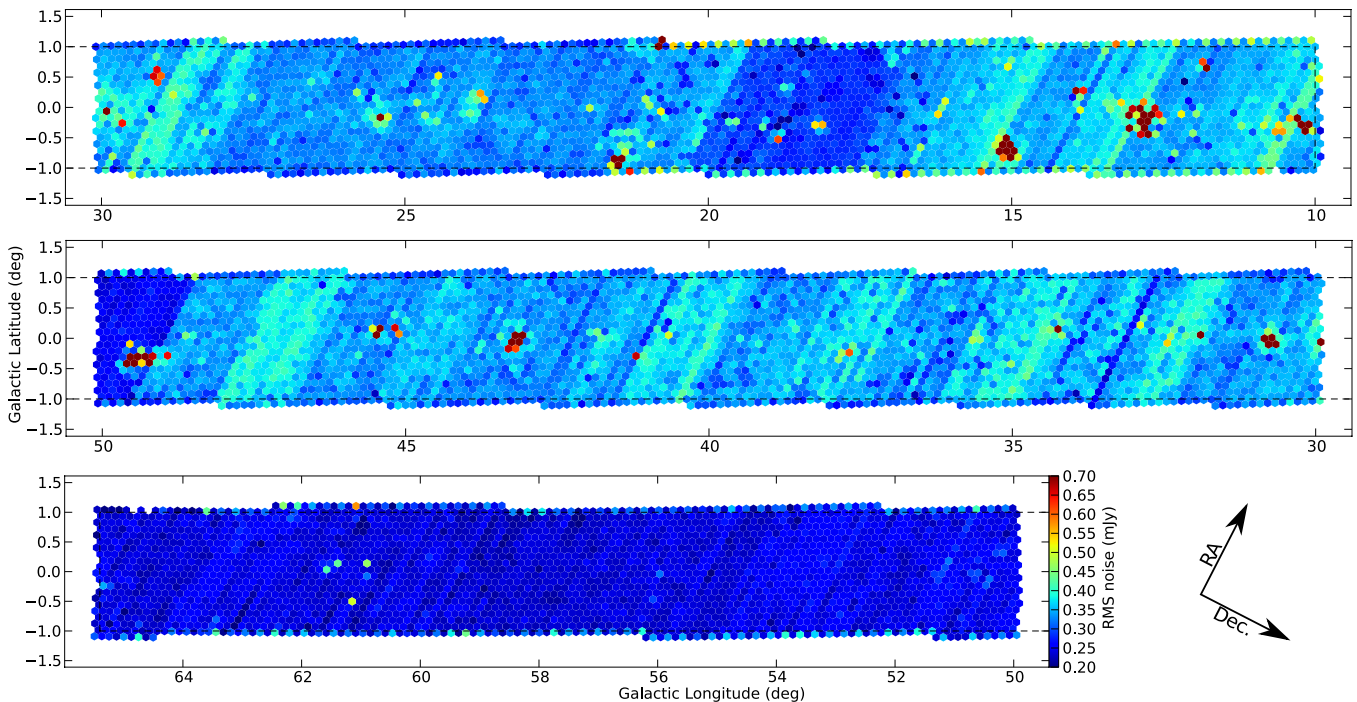
Within each observing block the beam elongation increases toward lower declinations, while the position angle varies by

$\sim 60$  deg. The distribution of beam minor-axes in the survey area separates into two distinct populations, with a small peak at  $0''.77$  and a large peak at  $1''.2$ . The smaller peak stems from the low-declination fields observed using the BnA array configuration, while the larger one contains the majority of fields observed using the B array. In contrast, the distribution of major axes values is monolithic, with a median at  $1''.5$  and a standard deviation of  $0''.32$ . Ninety-eight percent of fields have elongations less than 2 and 74% less than 1.5.

Based on these values, we chose to force a circular restoring beam of FWHM  $1''.5$  because this greatly simplified the mosaicking operation and meant that the restoring beam shape was constant across every mosaicked image. The value  $1''.5$  was chosen as the median value of the measured major-axes from all CORNISH fields. The degree of super-resolution is presented in Figure 9 and is less than 1.5 in 96% of fields. The restoring beam *area* is larger than the synthesized beam area for 8154 fields (87.2%) and is less than 1.5 times greater in 9343 fields (99.9%).

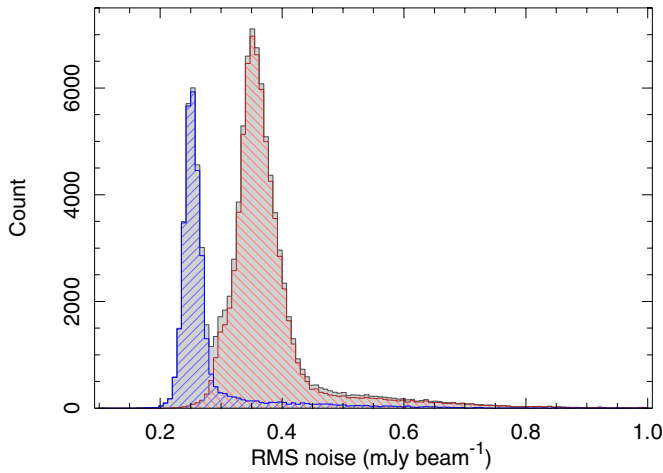
#### 4.3. Sensitivity and Uniformity

Figure 10 presents an image of the rms noise over the full survey area, with each color-coded hexagon representing a field. The locations of H II region complexes are prominent as clumps of high-noise fields located close to the mid-plane of the Galaxy. Away from such regions the noise level within individual scan-rows (scanning in R.A.) is relatively constant compared to the variation between rows, which is largely weather related. The observation area can be divided into two regions with noticeably different noise properties. At declinations greater than  $\delta = 14^\circ 2'$ , the median rms noise is significantly lower ( $\text{rms}_{\text{outer}} = 0.25 \text{ mJy beam}^{-1}$ ) than the remainder of the survey area ( $\text{rms}_{\text{inner}} = 0.35 \text{ mJy beam}^{-1}$ ). This outer CORNISH region corresponds directly to the epoch IIIb observations detailed in



**Figure 10.** Map of the rms noise in each field of the CORNISH survey. Striping in R.A. is due to changes in observing conditions between scan-rows. Clusters of high-noise fields (red) occur at the locations of star-forming complexes, which contain bright and extended emission.

(A color version of this figure is available in the online journal.)



**Figure 11.** Distribution of the total noise over the survey region made by sampling on  $2'$  scales (solid histogram). The right-slanted histogram (blue) contains only data from the epoch IIIb observations ( $\delta > 14^{\circ}2$ ) while the left-slanted histogram (red) contains the remainder (epochs I, II, and IIIa).

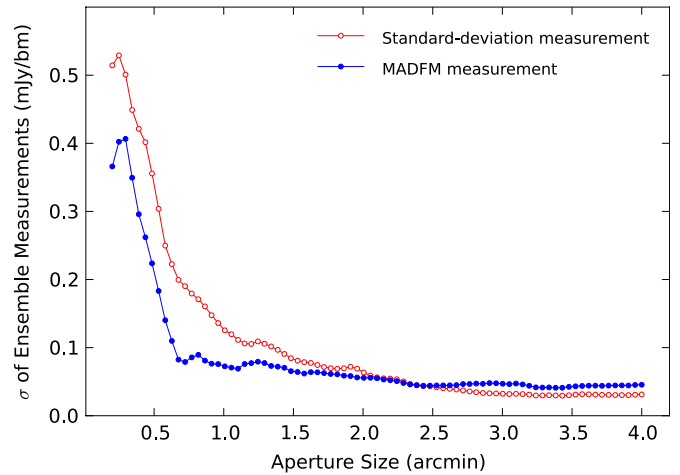
(A color version of this figure is available in the online journal.)

Table 1. From the 2007 season onward the VLA made extensive use of the upgraded JVLA antennas, which have more sensitive receivers. In addition, the weather conditions were better in the second season than in 2006, when observations were affected by electrical storms. Observations of approximately the inner 20 deg of the CORNISH area ( $\delta < -10^{\circ}5$ ) also took place during the second season, corresponding to epochs II and IIIa. However, the rms noise level is similar to the 2006 season for a number of reasons. In particular, the inner CORNISH region is seen at relatively low elevations from the VLA site, requiring the telescope to peer through a greater path-length of atmosphere. Emission from the atmosphere causes an increase in system temperature decreasing the signal-to-noise ratio in the data. The epoch II observations utilized fewer JVLA antennas and, because the telescope was at the beginning of the VLA/JVLA transition, required extensive flagging to render the data usable. Figure 11 presents a histogram of the distribution of noise measurements, sampled on  $2'$  scales, across the whole survey. The division between the inner and outer CORNISH regions is obvious. Both regions exhibit high-noise tails, corresponding to fields containing bright and extended emission.

#### 4.3.1. Spatial Scale of Noise

Interferometry data often exhibit non-Gaussian noise statistics, largely due to the nonlinear deconvolution process and poorly sampled  $uv$  coverage at large spatial scales. In regions with complex structures on scales greater than  $\sim 14''$  the emission is poorly constrained by  $uv$  coverage of the VLA B arrays. If only a few short baselines contain most of the flux a simple fringe pattern is produced on the sky. The flux is not evenly distributed but accumulates at specific spatial scales, depending on the sampling in  $uv$  space. The deconvolution algorithms used here also struggle to model this emission, resulting in some of the flux being scattered onto the surrounding sky (see Section 3.2). It is important to characterize this “ripple” noise pattern before attempting to search for real emission in the CORNISH data.

We have measured the noise characteristics of representative CORNISH data affected by a ripple. The region chosen was centered on  $\alpha = 18^{\text{h}}09^{\text{m}}21^{\text{s}}.96$ ,  $\delta = -20^{\circ}19'34''.9$  and the rms noise was measured using both the standard deviation (STDEV) and *median absolute deviation from the median* (MADFM)



**Figure 12.** Results of repeated noise measurements performed using a range of square apertures on CORNISH data containing a ripple. The y-axis plots the scatter in the ensemble set of measurements as a function of aperture size. It is clear the MADFM values are robust for apertures with scales greater than  $40''$ .

(A color version of this figure is available in the online journal.)

statistics. For a data set  $X = x_1, x_2 \dots x_i \dots x_n$  MADFM is given by

$$\sigma = K \text{median}(|x_i - \text{median}(X)|), \quad (2)$$

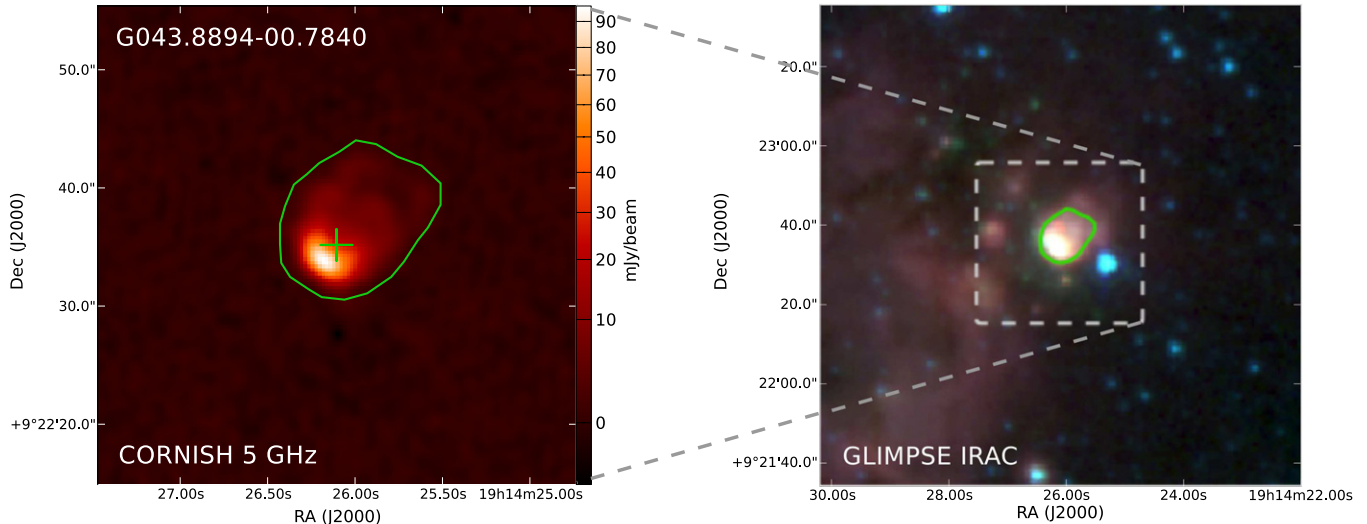
i.e., the median of the deviations from the median value. For a normal distribution MADFM is equivalent to the standard deviation using a scale factor  $K = 1.4826$ . The advantage of MADFM is that it is insensitive to the presence of outliers in the distribution and delivers a robust estimate of the true noise. Measurements were conducted using a range of aperture sizes, varying between  $12''$  and  $240''$  in steps of  $2''/82$ . In total twenty one positions were measured, offset in declination by  $6''$  along a line centered on the noise peak. The scatter in the results (expressed in standard deviations  $\sigma$ ) for each aperture size is plotted in Figure 12. From the plot we see that the scatter in the ensemble set of measurements increases as the aperture size decreases. The MADFM statistic remains stable at smaller spatial scales than the STDEV. At scales less than  $2'$  the scatter in the STDEV measurement slowly rises, compared to MADFM, whose scatter remains less than  $0.1 \text{ mJy beam}^{-1}$  until scales of  $40''$ . Measurements of the global noise-properties of the CORNISH data are therefore best performed using apertures spanning  $40''$  or larger using the MADFM statistic.

## 5. THE CORNISH SOURCE FINDER

We have developed an automated source finding procedure with the aim of producing a well-characterized catalog of 5 GHz emission in the northern Galactic plane. In the following subsections we describe the source-finding and measurement procedures and investigate the limits of the catalog.

### 5.1. Source Detection and Photometry

Tiles were automatically searched for emission using a custom procedure based on the OBIT FndSou task. FndSou identifies contiguous islands of emission above a global intensity threshold and attempts to fit one or more 2D Gaussians to each. This approach works well in the simplest case of an image with homogeneous noise properties, however, in the worst-case scenario the rms noise can change by a factor of a few over a  $20' \times 20'$  tile. This is especially true of tiles covering



**Figure 13.** Left panel: example of a polygonal photometry aperture drawn around a spatially extended CORNISH detection, in this case the UCH II region G043.8894–00.7840. The integrated flux density is calculated from the enclosed pixels minus an average background flux measured from the sky in the vicinity of the source. A cross marks the intensity-weighted position. Right panel: the GLIMPSE three-color image (red =  $8.3 \mu\text{m}$ , green =  $4.5 \mu\text{m}$ , blue =  $3.6 \mu\text{m}$ ) exhibits an overall morphology similar to that of the radio emission.

(A color version of this figure is available in the online journal.)

the Galactic mid-plane, where massive star-forming complexes are common. Using a single intensity threshold often results in spurious detections or omissions of real sources. To compensate for variable noise levels we ran FndSou on a  $9 \times 9$  grid of “patches” within the tile area. Each patch is 800 pixels ( $4'$ ) on a side and overlaps adjacent patches by 400 pixels in R.A. and decl. The local rms noise in each patch was determined using a histogram analysis clipped at  $3\sigma$  from the median value. With this patch layout a radio source within a  $2'$  band around the tile edge may be detected in two patches, except at the tile corners. A source in the interior may be detected in up to four overlapping patches. The maximum fitted Gaussian FWHM was constrained to be  $\leq 30''$  in keeping with the  $uv$  coverage. Fits within  $14''$  of the patch edge were deemed invalid, except where a patch abutted a tile edge. Running this patch-based emission finding procedure results in a degenerate list of sources with coincident positions derived from overlapping patches. A list of unique Gaussian fits to each tile was produced by searching for duplicates at similar positions (separation  $< 1''$ ) and with similar peak amplitudes ( $A_{\min}/A_{\max} > 0.7$ ). The Gaussian fit closest to the center of a patch was retained.

Initially, the search was conducted using a  $4\sigma$  local noise threshold and aperture photometry was performed to weed out detections with a signal-to-noise ratio  $\sigma < 5.0$  ( $\sigma = \text{maximum pixel/rms noise}$ ). An elliptical aperture was used to measure the source properties, which extended to the  $3\sigma$  Gaussian major and minor axes ( $2.548 \times \text{FWHM}$ ). If the emission was indeed Gaussian in shape this aperture would encompass 99.7% of the emitting flux. The rms noise and median background level of the sky were measured from a  $20''$  wide annulus centered on the source and offset from the measurement aperture by  $5''$ . The annulus width was chosen to sample the local noise pattern without being influenced by ripples or negative bowls (see Section 4.3.1). In crowded regions the sky annulus is likely to contain bright and real sources so the noise was measured using the robust MADFM statistic. The parameters of the valid Gaussian fits and photometric measurements were both recorded to the MySQL database, although the Gaussian fits are preferentially used in the default CORNISH catalog.

## 5.2. Resolved Emission

The source finder determines accurate fluxes for isolated and unresolved sources but decomposes complex structures into multiple overlapping Gaussians fits. It is highly desirable to merge these into a single measurement to avoid over-interpreting the number counts and properties of sources in the final catalog. Clusters of Gaussians were identified in the catalog using a friends-of-friends search: a Gaussian was associated with a cluster if it was within  $12''$  of any other member. In total, 741 clusters were found and these were all inspected manually. To distinguish between adjacent but unrelated sources and over-resolved emission the morphology at 5 GHz was compared to that in the *Spitzer* GLIMPSE mid-infrared images. The most common extended sources in the images are UCH II regions and planetary nebulae, each of which have distinctive mid-infrared signatures. For these types of object the morphology of the  $8 \mu\text{m}$  emission often echos that of the radio continuum (Hoare et al. 2007). If a cluster of Gaussian fits was found to trace an over-resolved source then the fitted parameters were replaced with a single measurement under a polygonal aperture manually drawn around the emission. Figure 13 shows an example of a polygon carefully drawn around the border of a cometary H II region. The flux density is calculated from the sum of the pixels within the source aperture minus the median background level in the vicinity of the source. In addition to the coordinates of the peak emission (for which the source is named in  $l$  and  $b$ ) we also record the geometric and intensity-weighted positions.

## 5.3. Measurements and Uncertainties

Below we explain how the properties of the sources were measured and the uncertainties calculated. The final values are presented in the CORNISH catalog, including the measurement-error and the absolute uncertainty on each parameter, incorporating the calibration error of 8.9%.

### 5.3.1. Gaussian Fits

Uncertainties on the Gaussian fits are calculated using the equations derived by Condon (1997), summarized here for

convenience. Noise in interferometric data is correlated on the scale of the synthesized beam FWHM, in this case  $\theta_{\text{bm}} = 1''.5$ . The effective signal-to-noise level  $\rho$  of a source with measured peak amplitude  $A_{\text{peak}}$  seen against a background of correlated Gaussian noise is given by

$$\rho^2 = \frac{\theta_M \theta_m}{4\theta_{\text{bm}}^2} \left[ 1 + \left( \frac{\theta_{\text{bm}}}{\theta_M} \right)^2 \right]^{\alpha_M} \left[ 1 + \left( \frac{\theta_{\text{bm}}}{\theta_m} \right)^2 \right]^{\alpha_m} \frac{A_{\text{peak}}^2}{\sigma_{\text{sky}}^2}, \quad (3)$$

where  $\theta_M$  and  $\theta_m$  are the respective major and minor fitted axes and  $\sigma_{\text{sky}}$  is the rms noise measured directly from the image. The exponents  $\alpha_M$  and  $\alpha_m$  have been estimated by Condon (1997) via Monte Carlo simulations and are  $\alpha_M = \alpha_m = 3/2$  for the amplitude and flux density errors,  $\alpha_M = 5/2$ ,  $\alpha_m = 1/2$  for the error on the major axis, and  $\alpha_M = 1/2$ ,  $\alpha_m = 5/2$  for the minor axes, position angle and absolute coordinate errors. On average, the signal-to-noise ratio is *increased* by a factor of 1.4. The positional uncertainties parallel to the major ( $\sigma_M$ ) and minor ( $\sigma_m$ ) fitted axes are given by

$$\frac{\sigma_M^2}{\theta_M^2} = \frac{\sigma_m^2}{\theta_m^2} \approx \frac{1}{(2 \ln 2)\rho^2}, \quad (4)$$

using values for  $\rho$  calculated from Equation (3). When the fit is projected onto equatorial axes the absolute position errors in right ascension ( $\sigma_\alpha$ ) and declination ( $\sigma_\delta$ ) become

$$\begin{aligned} \sigma_\alpha^2 &\approx \epsilon_\alpha^2 + \sigma_M^2 \sin^2(\text{P.A.}) + \sigma_m^2 \cos^2(\text{P.A.}), \\ \sigma_\delta^2 &\approx \epsilon_\delta^2 + \sigma_M^2 \cos^2(\text{P.A.}) + \sigma_m^2 \sin^2(\text{P.A.}), \end{aligned} \quad (5)$$

where P.A. is the position angle of the fitted major axis east of north and  $\epsilon_\alpha = \epsilon_\delta \approx 0.1$  arcsec is the systematic positional uncertainty. This value was determined via a comparison between CORNISH and catalogs of quasars whose positions are determined to milliarcsecond accuracy. The 15 matching quasars were drawn from the Goddard VLBI astrometric catalogs,<sup>32</sup> Very Long Baseline Array Galactic Plane Survey (VGaPS; Petrov et al. 2011) and the VLA calibrator manual<sup>33</sup> and their median offset of 0.1 arcsec was adopted as the systematic positional uncertainty for CORNISH.

Errors in the P.A. may be calculated from

$$\sigma_{\text{PA}}^2 = \frac{4}{\rho^2} \left( \frac{\theta_M \theta_m}{\theta_M^2 - \theta_m^2} \right)^2, \quad (6)$$

although we note that position angle values are only relevant when one or more axes is significantly resolved. Uncertainties associated with the fitted major and minor Gaussian FWHM are given by

$$\sigma^2(\theta_M) = \frac{2\theta_M^2}{\rho^2} + \epsilon_\theta^2 \theta_M^2 \quad \text{and} \quad \sigma^2(\theta_m) = \frac{2\theta_m^2}{\rho^2} + \epsilon_\theta^2 \theta_m^2. \quad (7)$$

The fractional calibration uncertainty  $\epsilon_\theta = 0.02$  is adopted from the VLA NVSS survey (Condon et al. 1998), which was observed using a similar snapshot mode. A single characteristic measured angular size  $\theta_f$  may be obtained from the geometric mean of the major and minor axes

$$\theta_f = \sqrt{\theta_M \theta_m}, \quad (8)$$

and its associated measurement uncertainty given by

$$\sigma(\theta_f) = \frac{\theta_f}{2} \sqrt{\frac{\sigma^2(\theta_M)}{\theta_M^2} + \frac{\sigma^2(\theta_m)}{\theta_m^2}}. \quad (9)$$

The restoring beam was forced to be a circular Gaussian of FWHM  $\theta_{\text{bm}} = 1''.5$  over the whole survey area so the deconvolved source size  $\theta_s$  in arcseconds may be found from

$$\theta_s = \sqrt{\theta_f^2 - 1.5^2}, \quad (10)$$

although we note that detections with  $\theta_f < 1''.8$  are considered unresolved in CORNISH.

The fitted amplitude  $A_{\text{peak}}$  must be corrected for the clean bias  $\Delta A_{\text{cb}} = -0.94 \sigma_{\text{sky}}$ , which we measured in Section 5.6 below, so

$$A = A_{\text{peak}} - \Delta A_{\text{cb}}. \quad (11)$$

The uncertainty on the fitted amplitude may be calculated from

$$\sigma_A^2 = \frac{2A^2}{\rho^2} + \epsilon_A^2 A^2, \quad (12)$$

where  $\epsilon_A = 0.089$  is the fractional amplitude calibration error. Finally, the integrated flux density  $S$  under a 2D Gaussian is given by

$$S = \frac{A \pi}{4 \ln(2)} \frac{\theta_M \theta_m}{\theta_{\text{bm}}^2} \quad (13)$$

and the corresponding uncertainty is

$$\frac{\sigma_S^2}{S^2} \approx \frac{\sigma_A^2}{A^2} + \frac{\theta_{\text{bm}}^2}{\theta_M \theta_m} \left[ \frac{\sigma^2(\theta_M)}{\theta_M^2} + \frac{\sigma^2(\theta_m)}{\theta_m^2} \right]. \quad (14)$$

### 5.3.2. Aperture Photometry

The peak amplitude reported for sources measured using aperture photometry is simply the intensity of the brightest pixel within the source aperture, corrected for the clean bias.

$$A = A_{\text{max}} - \Delta A_{\text{cb}}. \quad (15)$$

The effective signal-to-noise of the source in the presence of correlated Gaussian noise may be determined from a modified version of Equation (3) in which  $\theta_M$  and  $\theta_m$  are both replaced with the intensity-weighted diameter

$$\theta_d = 2 \frac{\sum_{i=1}^{N_{\text{src}}} r_i A_i}{\sum_{i=1}^{N_{\text{src}}} A_i}. \quad (16)$$

For a perfectly circular Gaussian source  $\theta_d \equiv \theta_f \equiv \text{FWHM}$ . In Equation (16)  $r_i$  is the angular distance from the  $i$ th pixel to the brightness-weighted center and  $A_i$  is its intensity. The sky noise corrected for Gaussian correlation is then

$$\frac{1}{\sigma_g^2} = \frac{\theta_d^2}{4\theta_{\text{bm}}^2} \left[ 1 + \left( \frac{\theta_{\text{bm}}}{\theta_d} \right)^2 \right]^\alpha \frac{1}{\sigma_{\text{sky}}^2}, \quad (17)$$

where  $\alpha = 3$  for all errors. The uncertainty on the peak amplitude is

$$\sigma_A^2 = \sigma_g^2 + \epsilon_A^2 A_{\text{max}}^2, \quad (18)$$

where  $\epsilon_A = 0.089$  is the calibration error.

<sup>32</sup> <http://gemini.gsfc.nasa.gov/solutions/>

<sup>33</sup> <http://www.vla.nrao.edu/astro/calib/manual/>

The equation for measuring the integrated flux density  $S_{\text{phot}}$  of a source using aperture photometry can be written as

$$S_{\text{phot}} = \left( \sum_{i=1}^{N_{\text{src}}} A_i - N_{\text{src}} \bar{B} \right) / a_{\text{bm}}, \quad (19)$$

where  $\sum_{i=1}^{N_{\text{src}}} A_i$  is the total flux in the source aperture summed over  $N_{\text{src}}$  pixel elements (in units of Jy pixel<sup>-1</sup>),  $\bar{B}$  is the background flux density estimated from the median level in the sky-annulus and  $a_{\text{bm}} = \pi \theta_{\text{bm}}^2 / (4 \ln(2) \Delta_{\text{pix}}^2)$  is the beam-area in pixels (28.33 pixels for CORNISH data).  $S_{\text{phot}}$  must also be corrected for clean-bias. If the source is unresolved the missing flux  $\Delta S_{\text{cb}}$  is given by

$$\Delta S_{\text{cb}} = -0.94 \sigma_{\text{sky}} \frac{\pi}{4 \ln(2)} \frac{\theta_{\text{bm}}^2}{a_{\text{bm}}}, \quad (20)$$

however, because the clean-bias reduces the flux in all clean-components (CCs) by a constant factor the effect on extended emission is difficult to gauge. The minimum number of CCs required to model an extended source can be estimated from the number of beam-areas  $n_{\text{beams}}$  subtended by the emission. The integrated flux density is then

$$S = S_{\text{phot}} - \Delta S_{\text{cb}} * n_{\text{beams}}. \quad (21)$$

The error on the integrated flux density may be found from

$$\sigma_S^2 = \left( \sigma(\sum A_i)^2 + \frac{\pi N_{\text{src}}^2 \sigma_g^2}{2 N_{\text{sky}}} \right) / a_{\text{bm}}^2, \quad (22)$$

where  $\sigma_g^2$  is the corrected variance,  $N_{\text{src}}$  and  $N_{\text{sky}}$  are the number of pixels in the source- and sky-apertures, respectively. The term  $\sigma(\sum A_i)$  is the uncertainty on the sum over the pixels in the source aperture given by

$$\sigma(\sum A_i) = \sum_{i=1}^{N_{\text{src}}} \left( \sqrt{\sigma_g^2 + \epsilon_A^2 A_i^2} \right). \quad (23)$$

The uncertainty on the intensity weighted diameter maybe found from

$$\frac{\sigma^2(\theta_d)}{\theta_d^2} = \frac{\sigma_g^2 \sum r_i^2}{(\sum r_i A_i)^2} + \frac{N_{\text{src}} \sigma_g^2}{(\sum A_i)^2}, \quad (24)$$

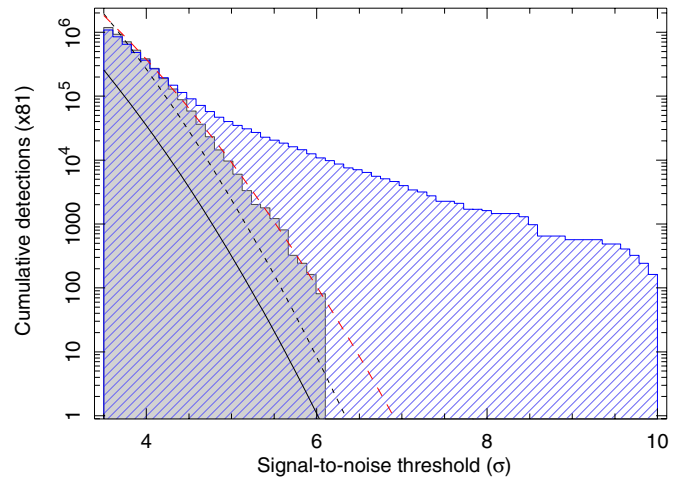
where the sums are taken over the pixels in the source-aperture. For emission measured using a polygonal aperture the intensity-weighted position is given by

$$\bar{x} = \sum_{i=1}^{N_{\text{src}}} x_i A_i / \sum_{i=1}^{N_{\text{src}}} A_i, \quad (25)$$

where  $x_i$  is the right ascension ( $\alpha$ ) or declination ( $\delta$ ) in the  $i$ th pixel. The corresponding error in  $\bar{x}$  is given by

$$\sigma_x^2 = \left[ \frac{\sigma_g^2 \sum x_i^2}{(\sum x_i A_i)^2} + \frac{N_{\text{src}} \sigma_g^2}{(\sum A_i)^2} \right] \bar{x}^2 + \epsilon_x^2, \quad (26)$$

with  $\epsilon_x = \epsilon_\alpha = \epsilon_\delta$ , the absolute positional error of the associated phase calibrator.



**Figure 14.** Solid histogram illustrates the cumulative distribution of spurious detections in the CORNISH catalog as a function of the signal-to-noise ratio measured from fourteen inverted tiles. The plot has been scaled to the total CORNISH area ( $\sim 1300$  tile areas) by multiplying by 81. In comparison, the hatched histogram shows the distribution of detections from the same tiles before inversion. See Section 5.4 for further details.

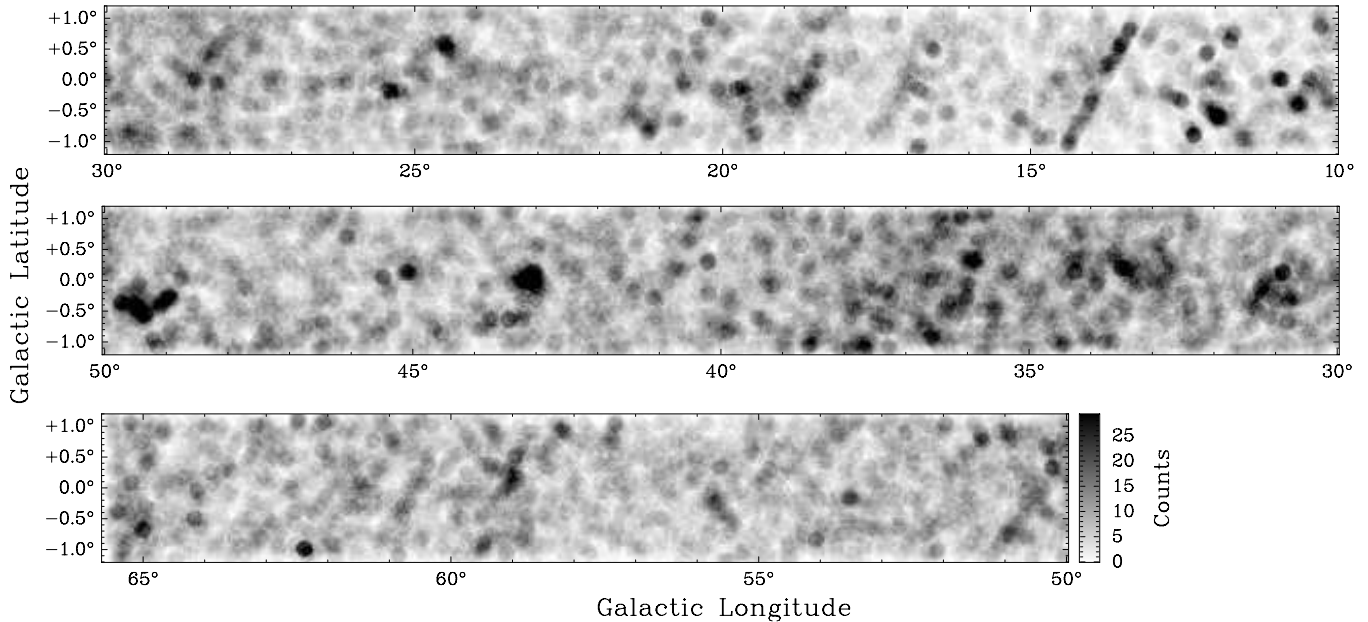
(A color version of this figure is available in the online journal.)

#### 5.4. Spurious Sources

We have attempted to estimate the number of spurious sources detected in *well calibrated and well behaved data* by running the source finder on inverted tiles, i.e., tiles where the pixel values have been multiplied by  $-1$ . Any negative detections will be false and allow us to estimate the number of spurious sources as a function of the signal-to-noise ratio. Fourteen tiles were selected to be representative of the emission properties across the survey region. They contained variously: no strong emission, one or more point sources with  $S_{5\text{GHz}} > 50$  mJy, weak extended emission, and bright extended sources causing moderately elevated noise levels ( $0.5 \text{ mJy} < \text{rms} < 0.8 \text{ mJy}$ ). For comparison, we ran the source finder using a  $3.5\sigma$  cutoff on both the inverted and the regular tiles. Figure 14 plots the cumulative counts of detections as above a signal-to-noise ratio, expressed as  $\sigma$ . The gray shaded histogram illustrates the number of spurious detections in the inverted tiles, while the hatched histogram illustrates the detections in the normal tiles, some of which will be real. Note that below  $4.5\sigma$  the detections are dominated by spurious sources. The 14 tiles represent 1.23% of the survey area, so by scaling the plot by 81 we can estimate the total distribution of spurious sources in CORNISH. In Figure 14 the number of spurious sources found decreases to 81 at  $6.1\sigma$ , above which our scaling is too crude to sample. For populations governed by Gaussian statistics the fraction  $f(\sigma)$  of the populations lying within a  $\sigma$ -threshold is given by

$$f(\sigma) = 1 - \text{erf}(\sigma/\sqrt{2}), \quad (27)$$

where  $\text{erf}(\sigma)$  is the Gaussian error function. The solid black curve in Figure 14 plots  $f(\sigma)$  assuming the total number of possible detections is equal to the number of synthesized beam areas in CORNISH ( $\approx 5.6 \times 10^8$  beams). It is clear that this assumption underestimates the number of spurious sources found, so we fit  $f(\sigma)$  with the total number of sources as a free parameter. The fit is shown by the short-dashed line and is dominated by the large number of sources in the bins with  $\sigma < 4$ . Above  $4\sigma$  the distribution has a shallower fall-off than expected for Gaussian statistics. An alternative is shown by the



**Figure 15.** Density map of all CORNISH sources detected above  $5\sigma$ . The gray-scale level illustrates the number of sources found within an 8 arcmin radius of any position. Away from complexes of H II regions, which are mostly near the Galactic mid-plane, variations in the source counts are due to spurious sources, e.g., the bad scan-row at  $l \approx 14^\circ$ .

long-dashed line, which is fitted only to the bins with  $\sigma > 4$  and uses the error-function of a distribution with a narrower width than purely Gaussian ( $\sigma = 0.9\sigma_{\text{gauss}}$ ). It is a significantly better match to the high signal-to-noise end of the distribution and predicts less than one spurious source above  $7\sigma$ . Based on this reasoning we have chosen a signal-to-noise cutoff of  $7\sigma$  for a high-reliability CORNISH catalog. We caution that data with greater complexity or poor calibration may introduce significant numbers of false sources above this level, so this does not mitigate the need to manually inspect the data for artifacts. Sources detected below  $7\sigma$  are not offered as an official data-product, but this low-reliability catalog will be made available on the CORNISH web page

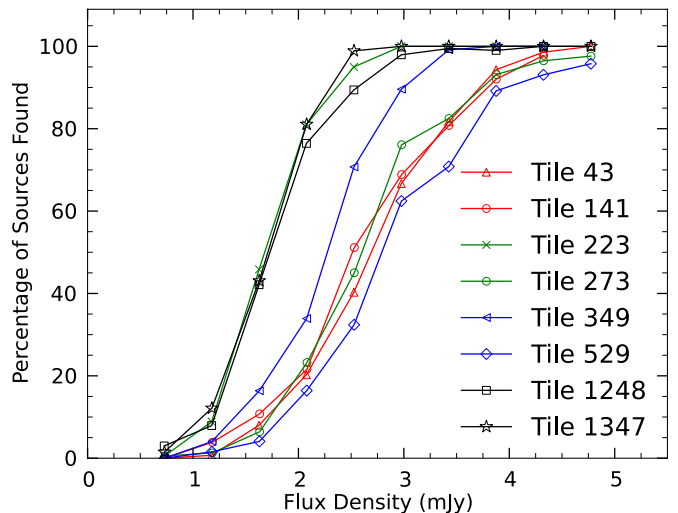
#### 5.4.1. Density of Low Signal-to-noise Sources

A density map of the CORNISH detections serves to highlight regions containing excessive numbers of weak sources, some of which may be spurious. Figure 15 plots the number density of CORNISH sources above  $5\sigma$  summed within an 8 arcmin radius. The most prominent feature is a line of elevated pixels corresponding to the scan row at  $\delta = -16^\circ 55' 03''.13$  ( $13^\circ 3' < l < 14^\circ 5'$ ). This row is unique in the survey as each field has only a single 40 s snapshot observation. Data in the second pass were found to be corrupted and no repeat observations were scheduled due to an array configuration change.

Isolated clumps of pixels with high source counts (e.g., at  $l = 12$ ,  $l = 31$ ,  $l = 43$  and  $l = 49$ ) correspond to molecular cloud complexes forming massive stars. The over-density of weak sources in these regions could be due to either a real increase in source counts or to an increase spurious in sources generated by the deconvolution process. Both scenarios warrant careful inspection of the data.

#### 5.5. Completeness

To quantify the formal sensitivity limits of the pipeline reduced data we conducted completeness tests on tiles from “empty” parts of the sky. The tiles were chosen to have few



**Figure 16.** Percentage completeness as a function of flux density for point sources within eight representative tiles. See the text in Section 5.5 for further details.

(A color version of this figure is available in the online journal.)

detections above  $5\sigma$ , homogeneous noise properties and be free of obvious imaging artifacts. One-hundred artificial point sources were injected into the calibrated  $uv$  data for each tile before creating a mosaicked image. The flux densities of the injected sources were varied randomly between 0.5 and 5.0 mJy, so as to bracket the expected sensitivity limit. Positions were also chosen randomly, but avoided known emission, the tile edges or regions where noise spikes were common. FndSou was then used to find and fit the emission with Gaussians. After 20 iterations of the injection-imaging-fitting routine the aggregate results were compared to the injected source parameters.

Figure 16 plots the percentage of sources recovered as a function of the flux density for tiles covering a range of rms values and epochs. The image parameters are presented in

**Table 2**  
Completeness Limits of Tiles Spanning a  
Representative Range of Noise Levels

Tile	rms (mJy beam <sup>-1</sup> )	$f_{\text{MRF}}$	Epoch	50% (mJy)	90% (mJy)
43	0.384	1.1	II	2.7	3.8
141	0.384	1.1	II	2.5	3.8
223	0.299	1.0	IIIa	1.7	2.3
273	0.358	1.0	IIIa	2.6	3.8
349	0.344	0.8	I	2.2	3.0
529	0.367	0.8	I	2.8	3.9
1248	0.239	0.8	IIIb	1.7	2.5
1347	0.257	0.8	IIIb	1.7	2.3

Table 2 alongside the 50% and 90% completeness limits. As expected, tiles with lower measured rms noise levels tend to have lower completeness limits. There is, however, significant variation between tiles as the local completeness limit ultimately depends on the uniformity of the noise pattern within the tile. At worst (tile 529, epoch I) the CORNISH survey is 90% complete to point sources at the 3.9 mJy level.

### 5.6. Clean Bias

When deconvolving the synthesized beam from the images, the flux level at which the CLEAN algorithm halts must be chosen carefully. If the cutoff is set too far above the noise then the residual images will be dominated by sidelobe patterns. If it is too low CLEAN will inadvertently identify noise spikes and sidelobes as real emission. Both negative and positive clean components on sidelobes will result in flux being subtracted from the positions of real sources and can artificially lower the rms noise. The AutoWindow function described in Section 3.2 has been shown to reduce the clean bias. However, we chose a relatively low CLEAN cutoff when imaging CORNISH data and need to measure the bias level in order to evaluate the correct flux-densities and uncertainties for the CORNISH catalog. In a similar test to the one presented in Section 5.5 we inserted 20 point sources into the  $uv$  data for each of tiles 43, 273, 529, and 1248. The flux densities were set randomly between 2 and 20 mJy. The data were imaged and mosaicked using the CORNISH pipeline, and aperture photometry was used to recover the artificial source fluxes. We found that the clean bias was consistent across all four tiles, despite being representative of different epochs. We adopt a mean clean bias of  $\Delta A_{\text{peak}} = -0.94 \sigma_{\text{sky}}$  (typically 0.33 mJy), indicating a moderate level of over-cleaning compared to NVSS, which quotes  $\Delta A_{\text{peak}} = -0.67 \sigma$ . We judge that this will not affect the utility of the catalog.

### 5.7. Manual Quality Control

According to the results of Section 5.4 we do not expect to find significant numbers of spurious sources above  $7\sigma$  in well-behaved CORNISH data. This statement is not necessarily true of high-noise fields containing bright and extended emission associated with massive star-forming complexes. Occasionally, peaks in the rippled noise pattern may be mistaken for real emission, or calibration errors may conspire to create false sources. To alleviate this problem we visually inspected all high-reliability CORNISH detections (i.e., those peaking above  $7\sigma$ ) to assess them as potential artifacts.

The CORNISH team visually inspected all mosaic tiles and individual sources in the  $7\sigma$  catalog. All sources were classified

as being either “unlikely,” “possibly” or “likely” an artifact based on the criteria above, i.e., located on a peak in a high noise ripple region, near a very bright source, or in a region where there appears to be an excessive number of potentially spurious  $5\sigma$ – $7\sigma$  sources (see Figure 15). If a source suspected of being possibly or likely an artifact was found to have a radio or infrared counterpart then, of course, the flag was left as “unlikely” in the CORNISH database. Smaller UCH II regions lying within the noise radius of a much brighter emission often have counterparts in the GLIMPSE IRAC data, while planetary nebulae (PNs) are often seen in the UKIDSS data, confirming them as real detections. Both source types appear in the far-infrared MIPS GAL bands ( $24 \mu\text{m}$  and  $70 \mu\text{m}$ ). Real extragalactic sources are not likely have counterparts in the infrared data sets and so retain their possible-artifact flag in suspect regions.

## 6. RESULTS

We found 3062 sources in the CORNISH data above a  $7\sigma$  detection threshold. Of these, 2591 were well fit by model Gaussians and the remaining 471 sources required measurement using a hand-drawn polygonal aperture. A total of 286 and 138 sources were classified as “possible” or “likely” artifacts, respectively, and a flag set in the final high-reliability catalog. They remain available in the online catalog and users will be able to include possible and likely artifact sources in their searches. Below we present the new, high-reliability catalog of 5 GHz radio emission containing 2638 sources.

### 6.1. Catalog Format

Isolated and unresolved sources identified by the source finder have two recorded entries taken from fitted Gaussian parameters and aperture photometry measurements. Sources exhibiting structured and extended emission have a single entry, based on aperture photometry performed using a manually drawn polygon. When assembling an aggregate catalog we favored the Gaussian fitted values. The photometric measurements are useful for diagnostic purposes.

An excerpt from the final CORNISH catalog is presented in Table 3. The columns are as follows. Column 1 contains the CORNISH source name, constructed from the Galactic longitude and latitude of the source. The equivalent right ascension ( $\alpha$ ) and declination ( $\delta$ ) are displayed in Columns 2 and 3, respectively. For sources well fitted by Gaussians the adopted coordinates are simply the peak positions of the fits. The intensity-weighted position is quoted for extended sources measured using a polygonal aperture. The associated positional uncertainties are given in Columns 4 and 5. Two uncertainty values are quoted for catalog entries. The first value is the absolute uncertainty, incorporating both measurement and calibration errors. The second value (in brackets) is the error associated with the photometry or Gaussian fit alone. Column 6 presents the peak flux density in units of mJy beam<sup>-1</sup>. The 5 GHz integrated flux density ( $S_{5\text{GHz}}$ ) is presented in Column 7. Column 8 contains the measured angular scale of the emission  $\theta_f$ , which has been determined from the geometric average of the major and minor Gaussian fit axes, or the intensity-weighted diameter in the case of extended emission. Sources with  $\theta_s > 1.8$  are considered to be resolved in the CORNISH images and their deconvolved sizes are presented in Column 9. The local rms noise measured from the photometric sky-annulus is recorded in Column 10. Column 11 notes how the flux

**Table 3**  
5 GHz Sources in the CORNISH Catalog

(1) Name ( <i>l</i> & <i>b</i> ) (deg)	(2) $\alpha$ (J2000) ( <sup>h</sup> <sup>m</sup> <sup>s</sup> )	(3) $\delta$ (J2000) ( <sup>o</sup> ' ")	(4) $\sigma_\alpha$ (")	(5) $\sigma_\delta$ (")	(6) <i>A</i> (mJy $\text{bm}^{-1}$ )	(7) $S_{5\text{GHz}}$ (mJy)	(8) $\theta_f$ (")	(9) $\theta_s^a$ (")	(10) rms (mJy $\text{bm}^{-1}$ )	(11) Measure Type <sup>b</sup>	(12) Flags <sup>c</sup>			
G010.1640–00.3655	18 09 27.854	–20 19 25.99	0.54 (0.53)	0.29 (0.27)	14.25 $\pm$ 2.46 (2.11)	86.99 $\pm$ 17.32 (14.89)	3.706 $\pm$ 0.454	3.4	2.03	G		N	W	A
G010.4168+00.9356	18 05 09.172	–19 28 11.15	0.11 (0.03)	0.10 (0.03)	13.00 $\pm$ 1.23 (0.42)	14.38 $\pm$ 1.51 (0.78)	1.578 $\pm$ 0.043	–	0.40	G				
G011.2436+01.0526	18 06 25.797	–18 41 28.90	0.36 (0.35)	0.42 (0.41)	2.50 $\pm$ 0.46 (0.40)	11.90 $\pm$ 2.62 (2.26)	3.272 $\pm$ 0.441	2.9	0.42	G				
G011.3441–00.0381	18 10 40.146	–19 07 57.14	0.10 (0.01)	0.10 (0.01)	25.34 $\pm$ 2.28 (0.36)	25.34 $\pm$ 2.34 (0.64)	1.500 $\pm$ 0.027	–	0.35	G	C			7
G011.7210–00.4916	18 13 07.156	–19 01 12.34	0.17 (0.13)	0.17 (0.13)	3.76 $\pm$ 0.52 (0.40)	5.93 $\pm$ 1.13 (0.94)	1.883 $\pm$ 0.157	1.1	0.37	G				5
G011.7434–00.6502	18 13 45.167	–19 04 34.68	0.16 (0.13)	0.16 (0.13)	16.97 $\pm$ 1.52 (0.28)	156.88 $\pm$ 14.98 (4.16)	4.324 $\pm$ 0.008	4.1	0.28	P			W	
G011.9368–00.6158	18 14 01.100	–18 53 23.53	0.10 (0.03)	0.10 (0.03)	163.93 $\pm$ 14.59 (0.35)	1155.90 $\pm$ 105.38 (11.79)	5.886 $\pm$ 0.003	5.7	0.35	P			SW	
G013.7481+00.2673	18 14 23.833	–16 52 37.43	0.26 (0.24)	0.24 (0.22)	4.25 $\pm$ 0.56 (0.41)	22.93 $\pm$ 3.36 (2.59)	3.483 $\pm$ 0.270	3.1	0.48	G			NS	A
G014.2365+00.2117	18 15 34.470	–16 28 28.00	0.17 (0.14)	0.17 (0.13)	5.74 $\pm$ 0.63 (0.37)	24.34 $\pm$ 2.92 (1.91)	3.088 $\pm$ 0.157	2.7	0.38	G				
G014.2460–00.0728	18 16 38.295	–16 36 06.15	0.50 (0.49)	0.50 (0.49)	11.40 $\pm$ 1.04 (0.39)	51.26 $\pm$ 6.18 (3.96)	3.935 $\pm$ 0.026	3.6	0.39	P				
G015.1577–00.0401	18 18 19.203	–15 47 01.43	1.68 (1.67)	1.67 (1.67)	7.08 $\pm$ 0.70 (0.42)	12.77 $\pm$ 2.32 (1.98)	2.837 $\pm$ 0.064	2.4	0.42	P				
G017.7794–00.0082	18 23 18.623	–13 27 21.64	0.12 (0.07)	0.12 (0.06)	4.66 $\pm$ 0.52 (0.32)	4.66 $\pm$ 0.70 (0.56)	1.500 $\pm$ 0.081	–	0.30	G				
G019.6087–00.2351	18 27 38.266	–11 56 36.07	0.10 (0.02)	0.10 (0.02)	153.16 $\pm$ 13.63 (0.56)	2900.88 $\pm$ 260.93 (26.36)	13.108 $\pm$ 0.003	13.0	0.56	P	C		NS	7
G020.3969–00.7921	18 31 08.707	–11 30 16.21	0.11 (0.04)	0.11 (0.04)	10.68 $\pm$ 1.00 (0.32)	18.18 $\pm$ 1.82 (0.82)	1.957 $\pm$ 0.052	1.3	0.31	G	C			7
G023.0883+00.2242	18 32 32.290	–08 38 53.38	0.10 (0.01)	0.10 (0.01)	28.75 $\pm$ 2.58 (0.34)	34.06 $\pm$ 3.10 (0.67)	1.633 $\pm$ 0.027	–	0.33	G			S	
G024.3973+00.7938	18 32 56.271	–07 13 26.84	0.12 (0.07)	0.13 (0.08)	5.50 $\pm$ 0.63 (0.40)	6.83 $\pm$ 1.03 (0.80)	1.671 $\pm$ 0.094	–	0.38	G				
G025.0079+01.0336	18 33 12.925	–06 34 18.37	0.16 (0.12)	0.19 (0.16)	2.94 $\pm$ 0.40 (0.30)	5.60 $\pm$ 0.99 (0.81)	2.071 $\pm$ 0.166	1.4	0.28	G				
G028.1985–00.0503	18 42 58.139	–04 14 04.87	0.19 (0.17)	0.18 (0.16)	33.62 $\pm$ 3.00 (0.36)	136.26 $\pm$ 12.94 (3.35)	3.048 $\pm$ 0.007	2.7	0.36	P	C		S	
G029.0447–00.5989	18 46 28.464	–03 43 58.01	1.10 (1.09)	1.08 (1.08)	8.07 $\pm$ 0.79 (0.46)	20.21 $\pm$ 2.84 (2.16)	2.752 $\pm$ 0.039	2.3	0.46	P			N	
G030.8000–01.0444	18 51 16.124	–02 22 24.81	0.15 (0.12)	0.23 (0.21)	2.50 $\pm$ 0.38 (0.31)	3.66 $\pm$ 0.82 (0.70)	1.814 $\pm$ 0.183	1.0	0.30	G				
G032.9686–00.4681	18 53 10.271	–00 10 50.78	0.10 (0.01)	0.10 (0.01)	87.52 $\pm$ 7.80 (0.31)	87.52 $\pm$ 7.81 (0.53)	1.500 $\pm$ 0.022	–	0.32	G				
G033.4543–00.6149	18 54 34.789	00 11 04.31	0.10 (0.01)	0.10 (0.01)	55.24 $\pm$ 4.92 (0.28)	75.82 $\pm$ 6.78 (0.61)	1.757 $\pm$ 0.026	–	0.27	G				
G035.3304+00.6133	18 53 37.776	02 24 51.19	0.10 (0.03)	0.10 (0.03)	11.69 $\pm$ 1.09 (0.32)	11.69 $\pm$ 1.19 (0.56)	1.500 $\pm$ 0.037	–	0.31	G	C			
G036.6579–00.0417	18 58 23.408	03 17 47.40	0.11 (0.05)	0.11 (0.05)	6.12 $\pm$ 0.63 (0.32)	6.12 $\pm$ 0.78 (0.55)	1.500 $\pm$ 0.061	–	0.32	G	C			
G040.4633+00.9671	19 01 46.445	07 08 31.37	0.11 (0.05)	0.11 (0.05)	7.36 $\pm$ 0.75 (0.37)	7.36 $\pm$ 0.93 (0.63)	1.500 $\pm$ 0.059	–	0.35	G			S	
G043.1684+00.0087	19 10 14.145	09 06 15.01	0.53 (0.52)	0.52 (0.51)	78.74 $\pm$ 7.13 (1.69)	185.25 $\pm$ 21.98 (11.55)	2.305 $\pm$ 0.018	1.8	1.69	P	C		NSW7	
G043.1701+00.0078	19 10 14.716	09 06 19.18	0.14 (0.10)	0.15 (0.11)	90.46 $\pm$ 8.08 (1.70)	1108.08 $\pm$ 103.11 (32.49)	7.533 $\pm$ 0.010	7.4	1.70	P	C		NSW75	
G045.2343–01.1294	19 18 13.261	10 24 18.43	0.10 (0.03)	0.10 (0.02)	13.37 $\pm$ 1.23 (0.32)	16.14 $\pm$ 1.57 (0.63)	1.648 $\pm$ 0.037	–	0.31	G				
G048.8229+00.5618	19 18 58.546	14 22 03.89	0.15 (0.11)	0.15 (0.11)	2.04 $\pm$ 0.29 (0.23)	2.04 $\pm$ 0.46 (0.39)	1.500 $\pm$ 0.134	–	0.23	G				
G048.8900–00.7569	19 23 54.299	13 48 23.21	0.14 (0.10)	0.19 (0.17)	4.24 $\pm$ 0.55 (0.40)	7.30 $\pm$ 1.25 (1.02)	1.969 $\pm$ 0.149	1.3	0.37	G				
G050.5722–01.1229	19 28 32.225	15 06 44.73	0.16 (0.13)	0.16 (0.13)	2.07 $\pm$ 0.32 (0.26)	2.07 $\pm$ 0.53 (0.46)	1.500 $\pm$ 0.156	–	0.27	G				A
G051.9859+00.8436	19 24 07.840	17 17 27.81	0.11 (0.05)	0.11 (0.05)	5.35 $\pm$ 0.54 (0.26)	5.35 $\pm$ 0.67 (0.45)	1.500 $\pm$ 0.058	–	0.25	G				
G052.0011–00.5305	19 29 12.901	16 39 01.94	0.14 (0.09)	0.13 (0.08)	3.06 $\pm$ 0.38 (0.27)	3.06 $\pm$ 0.55 (0.48)	1.500 $\pm$ 0.106	–	0.27	G			S	
G055.8169–00.8698	19 38 14.930	19 49 27.92	0.11 (0.05)	0.11 (0.05)	4.00 $\pm$ 0.42 (0.23)	4.00 $\pm$ 0.54 (0.39)	1.500 $\pm$ 0.067	–	0.23	G				
G058.3595–00.5207	19 42 19.210	22 12 30.37	1.61 (1.61)	1.60 (1.59)	2.59 $\pm$ 0.30 (0.26)	8.17 $\pm$ 1.38 (1.20)	3.106 $\pm$ 0.060	2.7	0.26	P				
G061.4758+00.0913	19 46 48.020	25 12 46.18	0.10 (0.01)	0.10 (0.01)	49.80 $\pm$ 4.43 (0.46)	3466.32 $\pm$ 305.90 (29.02)	15.978 $\pm$ 0.003	15.9	0.46	P	C		NS	75
G061.4763+00.0892	19 46 49.067	25 12 44.00	0.11 (0.04)	0.11 (0.04)	90.52 $\pm$ 8.06 (0.46)	718.71 $\pm$ 64.37 (9.23)	6.536 $\pm$ 0.004	6.4	0.46	P	C		NS	7
G062.3292–01.0738	19 53 10.033	25 21 21.15	0.10 (0.02)	0.10 (0.02)	11.64 $\pm$ 1.07 (0.28)	11.64 $\pm$ 1.15 (0.49)	1.500 $\pm$ 0.034	–	0.30	G	CE			
G065.3071–00.2139	19 56 46.041	28 20 57.99	0.10 (0.01)	0.10 (0.01)	988.70 $\pm$ 87.99 (0.37)	988.70 $\pm$ 88.00 (0.64)	1.500 $\pm$ 0.021	–	0.35	G	C			7

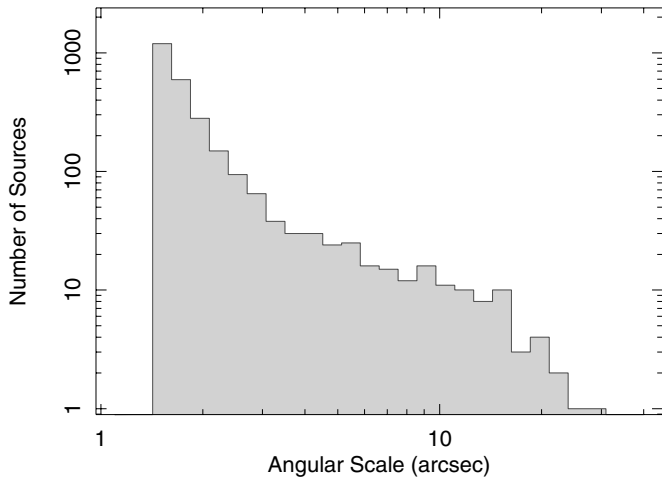
**Notes.** Two values are quoted for the uncertainty on all parameters. The first value is the absolute uncertainty, including measurement and calibration errors. The second value, in parentheses, is the uncertainty on the measurement alone.

<sup>a</sup> Sources with  $\theta_s < 1''8$  are considered unresolved in the catalog. We note that the  $1''8$  limit is only  $\sim 2\sigma$  from  $1''5$  for the weakest sources, which may cause some weak and unresolved sources to be labeled as resolved.

<sup>b</sup> The flux density of sources marked with an “P” in Column 11 was measured using polygonal apertures drawn by hand on the images, while a “G” means the flux density and peak-flux measurements were taken directly from the Gaussian fit.

<sup>c</sup> The flag codes in Column 12 have the following meanings: C = the source is part of a cluster, i.e., within  $12''$  of another source; E = the source is within 2 arcmin of a survey edge; N = the source lies within a high-noise region (rms  $> 0.45$  mJy); B = the source lies within  $3'$  of a bright (0.5 Jy) source; W =  $\nu\nu$  data for one or more fields contributing to a source was imaged using a smoothed weighting scheme; 7 = the source overlaps with another  $7\sigma$  catalog source; 5 = the source overlaps with a  $5\sigma$ – $7\sigma$  source; S = the source is located in a region with a high concentration of  $5\sigma$ – $7\sigma$  sources.





**Figure 17.** Distribution of source angular sizes for the high-reliability catalog.

density was measured, either with an polygonal aperture, or using the Gaussian fit. Finally, Column 12 contains a range of flags notifying the reader if the source

1. is within  $12''$  of another source,
2. is lying on an unusually high noise region ( $\text{rms} > 0.45 \text{ mJy beam}^{-1}$ ),
3. is imaged using the smoothed weighting scheme described in Section 3.2.2 (within  $4/45$  of a field center, i.e., half a primary beam FWHM),
4. is within  $3'$  of a bright ( $>0.5 \text{ Jy}$ ) source,
5. is within  $2'$  of the edge of the survey,
6. is within an area containing numerous low signal-to-noise detections likely to be spurious,
7. overlaps with another source,
8. is flagged as a suspected artifact during manual inspection.

The flags are described in more detail in the footnotes to Table 3.

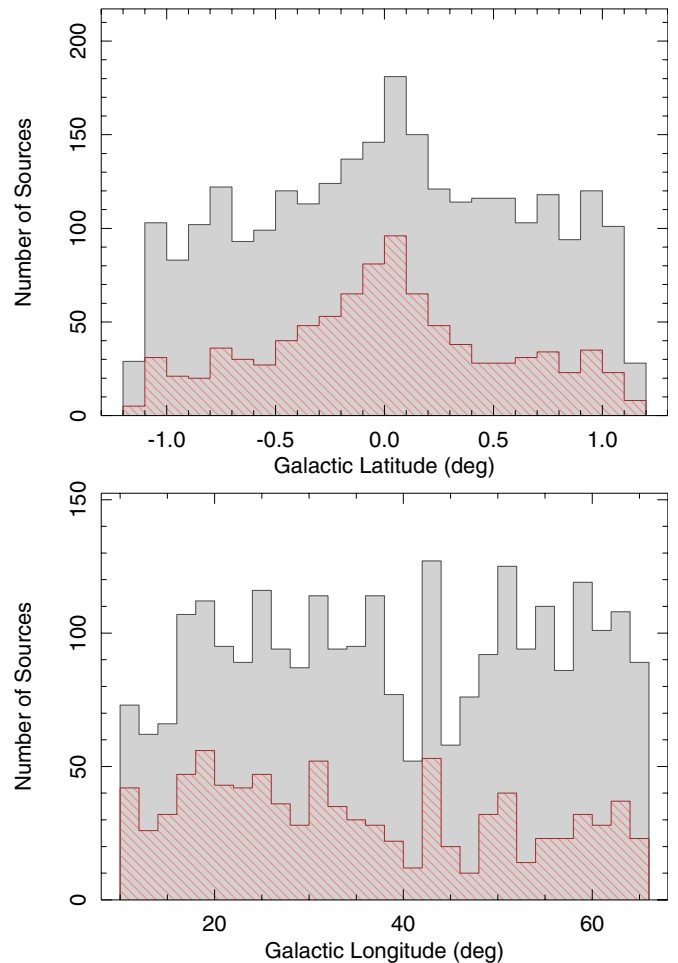
The full CORNISH catalog is offered to the astronomical community as a plain text file or VO table (XML-based table format defined by the Virtual Observatory) on the project Web site (<http://cornish.leeds.ac.uk>). A query-based Web interface is also available which allows the user to retrieve specific catalog subsets and drill down to the underlying data.

## 6.2. Ensemble Source Properties

### 6.2.1. Angular Size

The measured angular size  $\theta_f$  quoted in the catalog is given by the geometric average of the major and minor fitted Gaussian FWHM axes ( $\sqrt{\theta_M \theta_m}$ ), or by the intensity-weighted diameter ( $\theta_d$ ) in the case of emission measured using a polygonal aperture. For a bright source with a Gaussian morphology these measurements are equivalent within the errors. Figure 17 plots the distribution of measured angular sizes for the high-reliability catalog. The distribution begins to flatten at size scales greater than 5 arcsec, before tapering out at  $30''$ , where the upper limit for the source-fitter is set. Above sizes of  $14''$  the deconvolution algorithm struggles to model the poorly sampled longer  $uv$ -spacings (see Section 3.2.2), hence the slight increase in counts at that scale—broad sources are artificially truncated at sizes of  $14''$ .

The uncertainty on the angular size of CORNISH sources is better than  $0'.3$  for 96% of compact ( $\theta_f < 5''$ ) catalog entries. We consider sources with  $\theta_f < 1'.8$  (i.e., the restoring beam size plus



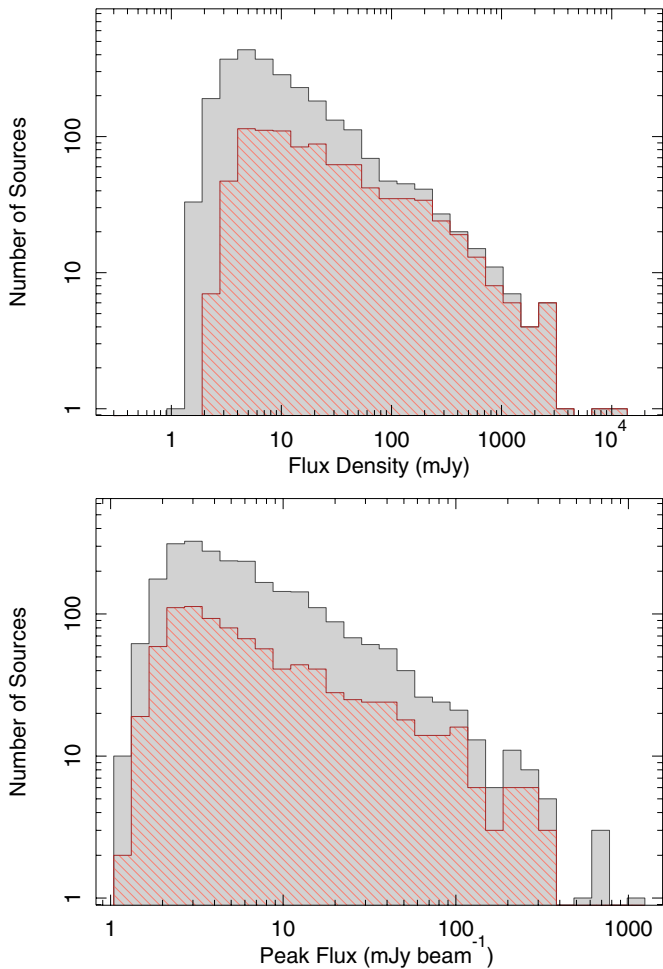
**Figure 18.** Distribution of CORNISH sources as a function of Galactic latitude (upper panel) and longitude (lower panel). In both panels the high-reliability catalog ( $\sigma \geq 7$ ) is plotted using solid shading, while the hatched histogram contains a subsample of resolved sources ( $\theta_f > 1'.8$ ).

(A color version of this figure is available in the online journal.)

$0'.3$ ) to be unresolved. Sixty-one percent of the  $7\sigma$  catalog fall into this category. Below we examine the differences between the resolved and unresolved populations.

### 6.2.2. Galactic Distribution

Figure 18 illustrates the distribution of CORNISH sources as a function of Galactic latitude (upper panel) and longitude (lower panel). The solid-shaded histogram contains all sources in the high-reliability catalog, while the hatched histogram contains only the subset of resolved detections (859 sources). Resolved sources account entirely for the broad peak seen in the latitude distribution, with the remaining unresolved detections exhibiting a flat profile. The scale-height of the resolved latitude distribution is  $0'.47$ , consistent with that of UCH II regions, 6.7 GHz methanol masers and other tracers of high-mass star formation (e.g., Green et al. 2009; Urquhart et al. 2007, 2009, 2011). The supposition that a large fraction of the resolved sources arise in high-mass star-forming regions is lent weight by their Galactic longitude distribution. The number of sources per  $2^\circ$  bin increases gradually toward longitude zero, while the two spikes at  $l \approx 43^\circ$  and  $l \approx 50^\circ$  correspond to the W49 and W51 complexes, respectively. Conversely, unresolved detections (1719 sources) show a flat distribution with Galactic longitude and are likely to contain significant numbers of



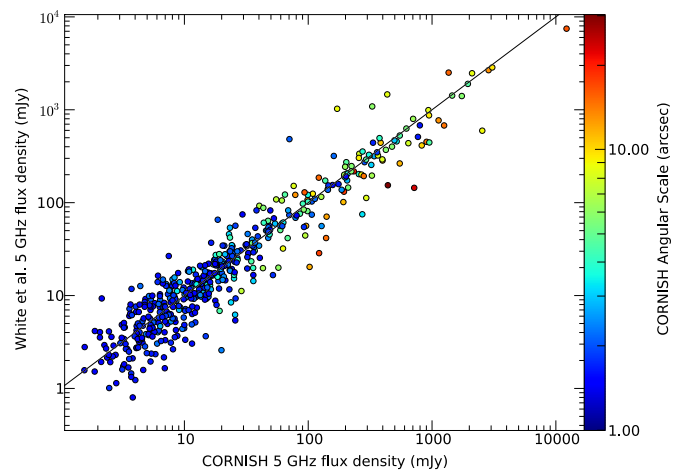
**Figure 19.** Distribution of integrated flux density (upper panel) and peak flux (lower panel). The high-reliability catalog is illustrated by the solid shaded histogram, while the hatched histogram contains only the subset of resolved sources.

(A color version of this figure is available in the online journal.)

AGNs, and other extragalactic sources. We note that this is partly by design as course adjustments were made to the deconvolution algorithm in order to keep the number of low-level sources roughly constant (Figure 3). The expected density of extragalactic sources in CORNISH may be calculated from Equation (A2) of Anglada et al. (1998), using the 5 GHz source counts of Condon (1984). Assuming median values of rms noise for the two regions presented in Figure 11, we expect to find  $\sim 2400$  extragalactic sources in our  $7\sigma$  catalog, consistent with the number of unresolved detections. Most extragalactic sources can be classified as they are not detected in any of the infrared wavebands. Specific catalogs of CORNISH sources identified as UCH II regions, PNs and AGNs will be presented in a forthcoming paper.

### 6.2.3. Flux Density and Peak Flux

Figure 19 shows the distributions of flux densities and peak fluxes for the CORNISH sources. At flux density levels of  $\sim 5$  mJy or greater the distribution is well fitted with a power law of index  $-0.81$ . Below 3 mJy the number of sources begins to decrease as the  $7\sigma$  detection limit is encountered. The distribution of flux densities for resolved sources turns over at approximately 6 mJy due to the constraints imposed by their selection and the  $7\sigma$  signal-to-noise cutoff.



**Figure 20.** Comparison of the 5 GHz flux density measurements for sources common to the CORNISH and White et al. (2005) catalogs. No systematic differences are apparent in the plot, however, the absolute differences between the two catalogs increase with flux density. The points are color-coded to show the angular size of the CORNISH detections and it is clear that the most extended sources are responsible for the outliers seen above  $\sim 100$  mJy.

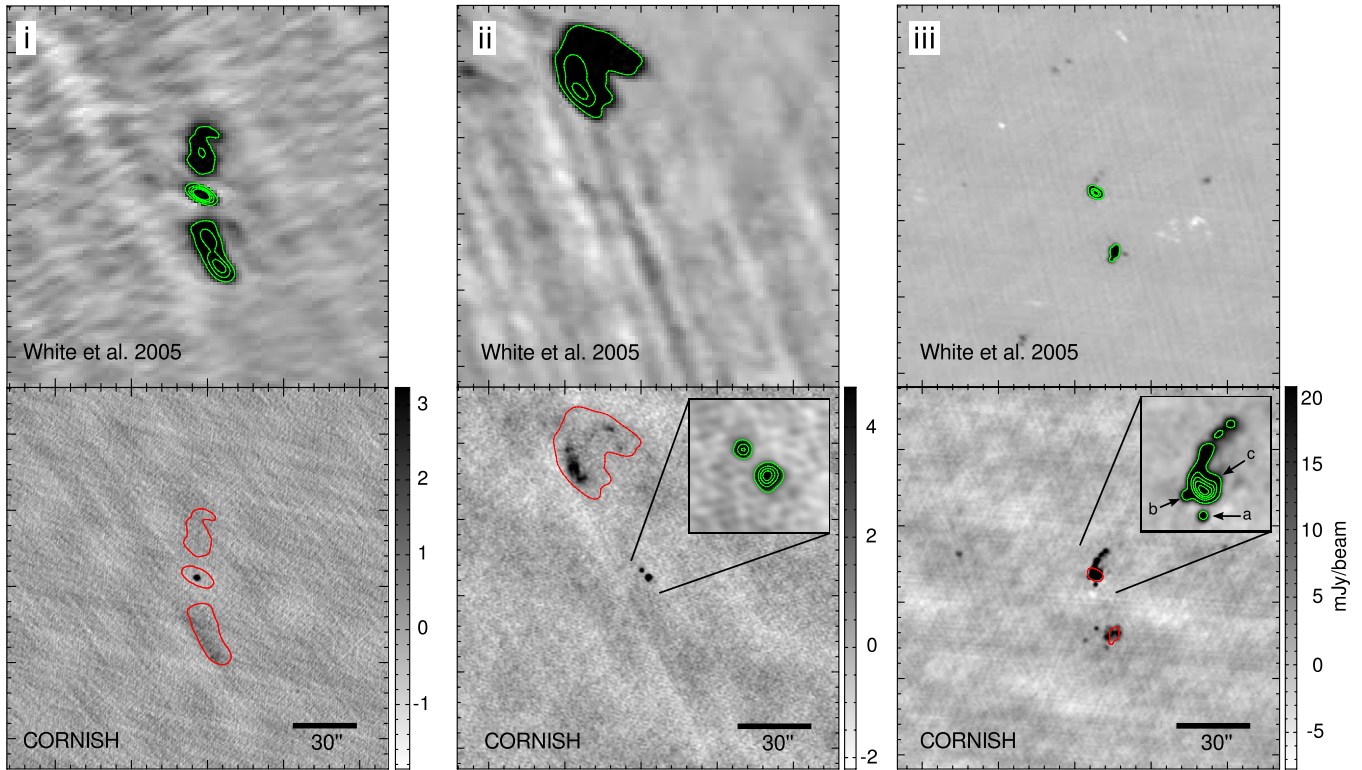
(A color version of this figure is available in the online journal.)

The peak flux distributions are identical for the high-reliability and the resolved source catalogs. Above the sensitivity cutoff ( $\sim 2.5$  mJy beam $^{-1}$ ), both are fit by a power law of  $n_{\text{src}} \propto S_{5\text{GHz}}^{-0.91}$ .

### 6.3. Comparison to Other Catalogs

Of all prior observations the White et al. (2005) VLA survey of the Galactic plane has the most similar observing setup and sky-coverage. A comparison to that work serves as a useful sanity-check on the ensemble properties of the CORNISH catalog and images. The 5 GHz component of the White et al. survey was observed using the D, DnC, and C arrays. These more compact VLA configurations (compared to B and BnA) yield better sensitivity to extended emission than CORNISH, but at a lower resolution ( $\sim 6''$ ). White et al. (2005) imaged the Galactic plane between  $-10^\circ < l < 42^\circ$  and  $|b| < 0.4$ , of which  $25.6$  deg $^2$  overlap with the CORNISH target area. The measured noise properties of their images are lower, with a median rms of  $\sim 0.27$  mJy beam $^{-1}$  compared to  $\sim 0.35$  mJy beam $^{-1}$  for CORNISH data. The cutoff limit for the White et al. source catalog was chosen to be  $5.5\sigma$  ( $\sim 1.4$  mJy beam $^{-1}$ ) compared to  $7\sigma$  ( $\sim 2.5$  mJy beam $^{-1}$ ) in this work. We would expect similar flux densities for compact sources ( $\leq 6''$ , see Figure 4) common to both catalogs, despite differences in  $uv$  coverage. Systematic errors present in either catalog should be obvious in a flux-flux comparison plot.

The White et al. 5 GHz catalog contains 1822 entries in the overlapping area and we match 558 of these with 521 CORNISH sources using a  $5''$  search radius. The number of matches diminishes significantly at matching radii greater than  $2''$ , however, a  $5''$  matching radius was chosen to allow for offsets in the positions assigned to resolved sources in both catalogs. Figure 20 presents a comparison of the measured flux densities for sources successfully cross-matched between the two surveys. The measurements agree, on average, to within 39%, with no evidence of systematic differences. A greater fraction of sources in the high flux density bins are resolved, hence the outliers in the plot above  $\sim 100$  mJy may be attributed to imaging and measurement differences between the two surveys. Points



**Figure 21.** Comparison between 5 GHz continuum images from the White et al. (2005) VLA survey (top) and CORNISH (bottom). Panel (i) shows a radio galaxy detected in both surveys. The unresolved central driving source is detected in CORNISH while the extended emission from the radio lobes is resolved out. Similarly, emission on scales of  $\sim 30$  arcsec is filtered from the cometary H II region G24.799+0.097, shown in panel (ii). However, CORNISH does an excellent job of imaging structures on scales less than  $\sim 14$  arcsec. Panel (iii) shows the star-forming region G34.26+0.15. The detailed structure is well-imaged by CORNISH. In comparison the White et al. survey struggles to resolve the individual knots of radio emission and significant imaging artifacts are present in the image. The two compact sources at the center of panel (ii) are not detected in the White et al. survey.

(A color version of this figure is available in the online journal.)

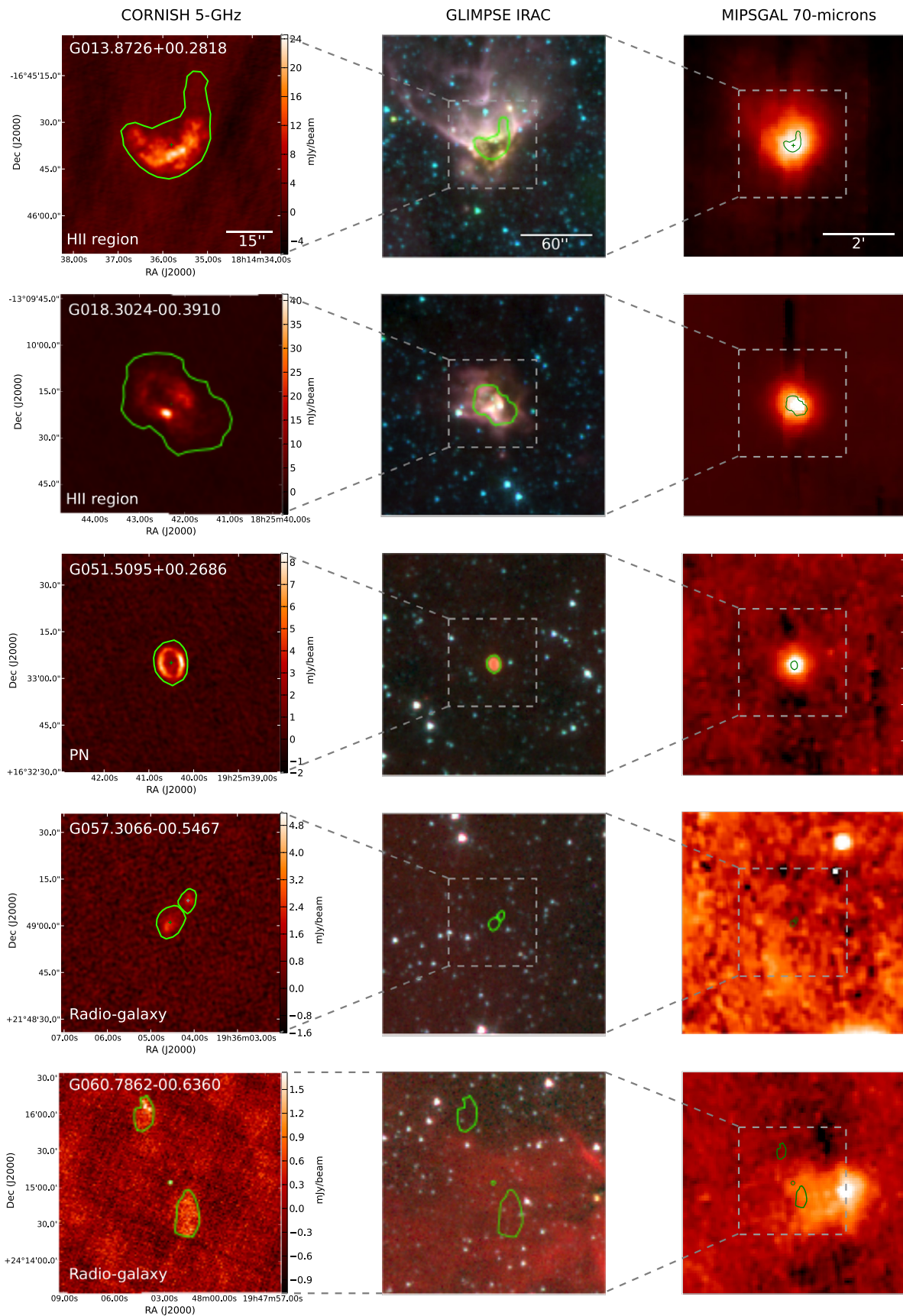
representing individual sources in Figure 20 are color-coded to indicate angular size in the CORNISH catalog. Unresolved sources (blue) cluster around the equality line while the outliers are almost all extended (red).

In total, 1264 sources in the White et al. catalog remain unmatched using a simple cone search within 5 arcsec. Of these, 50% lie above our  $7\sigma$  sensitivity threshold and are sufficiently bright to have been detected in CORNISH. The reasons for the disparity become apparent upon comparing the White et al. and CORNISH images. A significant fraction of the bright, unmatched sources have angular scales greater than  $\sim 20''$  in the White et al. data and are simply resolved out by the VLA B configurations used by CORNISH. Figure 21 (panels (i) and (ii)) presents examples of such objects. The central source powering the radio galaxy shown in panel (i) is detected in both surveys as a 25 mJy point source. The radio lobes have angular scales of  $\sim 30''$  in the White et al. image and the brighter southern lobe has a peak flux density of  $14 \text{ mJy beam}^{-1}$ . In the corresponding CORNISH image the northern lobe is completely resolved out, while the southern lobe is detected at a  $5.5\sigma$  level and is therefore not included in the high-reliability catalog. Similarly, the cometary H II region G24.799+0.097, shown in panel (ii), is resolved into multiple components by CORNISH. When assembling the CORNISH catalog we took great care to identify such over-resolved emission as a single source (see Section 5.2). In the White et al. catalog individual Gaussian fits to complex emission are left separate. G24.799+0.097, for example, has four catalog entries and a  $15''$  matching radius is required to correctly match these to their CORNISH counterpart.

Panel (iii) of Figure 21 presents an image of the G34.26+0.15 star formation region. G34.26+0.15 is divided into three components (*a*, *b*, and *c*), of which *c* is the prototype cometary UCH II region (van Buren et al. 1990). The CORNISH image clearly resolves all three components with excellent image fidelity. By contrast the White et al. image barely resolves the *c* component and contains significant numbers of imaging artifacts.

The remaining unmatched White et al. sources derive from intrinsic differences in the image quality between the two surveys. In general, the CORNISH images have more homogeneous noise properties and are of higher quality. The White et al. images contain numerous compact sites of emission not present in the equivalent CORNISH mosaics. These often occur in regions adjacent to very bright sources or poorly imaged extended emission, and are themselves generally unresolved and weak ( $90\% < 10 \text{ mJy beam}^{-1}$ ). Their morphology and location makes them likely to be artifacts of the imaging process. Examples of such artifacts are visible in panel (iii) and, to a lesser extent in panel (ii) of Figure 21. Artifacts in the CORNISH image are limited to a moderate level of ripple and few noise spikes. By comparison the White et al. images often contain significant sidelobe structure and a number of spurious emission sources. The advantages of utilizing a semi-automatic pipeline are apparent in the high-quality of the CORNISH images. In general, such fields also illustrate the importance of manually inspecting the results of automatic source-finders when dealing with under-sampled interferometric data.

There are 262 CORNISH sources which have no counterpart in the White et al. catalog, despite peaking well above the



**Figure 22.** Sample CORNISH images (left column) alongside *Spitzer* GLIMPSE (middle column) and MIPS GAL (right column) infrared data. The green polygon illustrates the aperture used to measure the properties of the 5 GHz emission. In the three-color GLIMPSE images the 8.0  $\mu\text{m}$  band is coded red, the 4.5  $\mu\text{m}$  green and 3.6  $\mu\text{m}$  blue.

(A color version of this figure is available in the online journal.)

nominal  $5.5\sigma$  detection limit. A small fraction of these are pathological cases, like the two compact sources at the center of the CORNISH image in panel (ii) of Figure 21. The sources have flux densities of 12 and 20 mJy and should be visible in the White et al. image. A linear discontinuity cuts through the image at this position, so we speculate that the omission derives from a problem with the imaging process used by White et al. (2005). The majority of the remaining unmatched CORNISH sources are detected below the  $10\sigma$  level and when present in the White et al. images are washed out by ripples in the noise or other imaging artifacts.

## 7. EXAMPLE CORNISH DATA

The CORNISH data set contains objects of many types, including H II and UCH II regions, PNs, evolved stars, active binaries, radio lobes from external galaxies, and many AGNs and quasars. Figure 22 presents sample CORNISH images for known objects opposite their counterpart data from the *Spitzer* GLIMPSE and MIPS GAL surveys. The first column of images shows the CORNISH data, the second a three-color image made from the mid-infrared GLIMPSE IRAC bands and the third column the  $70\ \mu\text{m}$  MIPS GAL image. The first two rows present examples of resolved H II regions with different morphologies. G013.8726+00.2818 is a classical cometary H II region, while G018.3024–00.3910 is irregularly shaped. For most resolved H II regions in CORNISH the shape of the 5 GHz continuum emission is echoed and extended in the GLIMPSE three-color image. The  $3.6\ \mu\text{m}$  and  $8.0\ \mu\text{m}$  bands (coded blue and red, respectively, in Figure 22) contain broad lines from polycyclic aromatic hydrocarbons (PAHs), which are excited by the strong ultraviolet radiation field (Peeters et al. 2002). Galactic massive star-forming regions are readily identifiable via the appearance of the filamentary PAH emission surrounding them. The extended flocculent emission (appearing purple in the GLIMPSE color coding in Figure 22) traces the clumpy photodissociation region (PDR) at the interface between the ionized gas and the enveloping molecular cloud. The spectral-energy distribution of H II regions peaks in the far-infrared and they are easily detected as a bright source in the MIPS GAL  $70\ \mu\text{m}$  images.

The third row shows a particularly good example of a resolved PNs. The GLIMPSE colors of PNs are similar to those of the H II regions, but the SED falls off more steeply in the far-infrared, hence the MIPS GAL  $70\ \mu\text{m}$  band is noticeably less bright. PNs tend to be isolated objects in the GLIMPSE images, having long since dispersed their natal molecular clouds. The expanding shell of gas surrounding the PN also contains PAHs, which are excited by the ultraviolet photons generated by the central stellar remnant (see Smith & McLean 2008 and references therein). Unlike the H II regions, the PAH emission is confined to the ejected envelope leading to simple mid-infrared morphology. G051.5095+00.2686 in Figure 22 exhibits a similar ring-shape in both CORNISH and GLIMPSE images.

The final two rows present examples of radio galaxies in which the radio lobes have been resolved. In G057.3066+005467 the central driving source is not detected and the lobes are barely resolved as two tear-drop shaped sources extended toward each other. The central driving source of G060.7862–00.6360 is detected as a point source near the center of the image and both radio lobes are well-resolved, if weak. Neither radio galaxy has a counterpart in any of the associated mid- or far-infrared images.

## 8. SUMMARY AND FUTURE WORK

The CORNISH project has delivered the best ever complementary radio view of the northern GLIMPSE region at 5 GHz (6 cm wavelength). With a resolution of  $\sim 1''.5$  and an rms noise level of  $<0.4\ \text{mJy beam}^{-1}$ , the survey is tailored to search for UCH II regions across the Galaxy, but has also detected a wide range of radio-bright objects that are also identified in other categories.

We present here a catalog of 3062 compact radio sources detected in CORNISH data above a  $7\sigma$  signal-to-noise threshold. A high-reliability subset (2638 sources) contains has been flagged to exclude potential spurious detections in poorly imaged  $uv$  data. Fields containing emission extended on scales greater than  $14''$  are poorly sampled by the  $uv$  coverage of the VLA B configuration, giving rise to a small number of spurious sources. Such fields represent only two% of the survey area and a rigorous program of manual inspection has flagged suspected artifacts, hence, we estimate the catalog reliability to be better than 99%. To date, the CORNISH catalog is the most uniformly sensitive, homogeneous and complete list of compact radio-emission sources at 5 GHz toward the northern Galactic plane.

Mosaicked images and calibrated  $uv$  data in FITS format are available to download from the CORNISH Web site (<http://cornish.leeds.ac.uk>). We have created a data server, which is operated by submitting a list of positions and serves either postage-stamp images or calibrated  $uv$  data. The full CORNISH catalog is also available online via a query-based interface as a plain-text format file or a VO table. General access is also available through the Vizier service.

Much work remains to be done in order to fully exploit the CORNISH data set. A future paper in the survey (C. R. Purcell et al., in preparation) will cross-match the 5 GHz radio emission to the complementary *Spitzer* GLIMPSE and UKIDSS data sets, allowing the identification of specific source types via their SEDs.

The authors thank the referee, Jim Condon, for his thorough comments which very much improved the quality of the paper. We also thank the Director and staff of the VLA for their assistance during the preparation of these observations. Thanks also to James Alison for many helpful discussions. C.R.P. was supported by a STFC postdoctoral grant while at the Universities of Manchester and Leeds.

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

*Facilities:* VLA, Spitzer (IRAC, MIPS)

## REFERENCES

- Altenhoff, W. J., Downes, D., Pauls, T., & Schraml, J. 1979, *A&AS*, **35**, 23  
 Anglada, G., Villuendas, E., Estalella, R., et al. 1998, *AJ*, **116**, 2953  
 Arvidsson, K., Kerton, C. R., Alexander, M. J., Koblunick, H. A., & Uzpen, B. 2010, *AJ*, **140**, 462  
 Becker, R. H., White, R. L., Helfand, D. J., & Zoonematkermani, S. 1994, *ApJS*, **91**, 347  
 Briggs, D. S. 1995, *BAAS*, **27**, 1444  
 Carey, S. J., Noriega-Crespo, A., Mizuno, D. R., et al. 2009, *PASP*, **121**, 76  
 Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, **121**, 213  
 Condon, J. J. 1984, *ApJ*, **287**, 461  
 Condon, J. J. 1997, *PASP*, **109**, 166  
 Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, **115**, 1693  
 Drew, J. E., Greimel, R., Irwin, M. J., et al. 2005, *MNRAS*, **362**, 753  
 Froebrich, D., Davis, C. J., Ioannidis, G., et al. 2011, *MNRAS*, **413**, 480

- Giveon, U., Becker, R. H., Helfand, D. J., & White, R. L. 2005, *AJ*, **129**, 348
- Green, J. A., McClure-Griffiths, N. M., Caswell, J. L., et al. 2009, *ApJL*, **696**, L156
- Hoare, M. G., Kurtz, S. E., Lizano, S., Keto, E., & Hofner, P. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 181
- Hoare, M. G., Purcell, C. R., Churchwell, E. B., et al. 2012, *PASP*, **124**, 939
- Homan, D. C., & Lister, M. L. 2006, *AJ*, **131**, 1262
- Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, *ApJS*, **163**, 145
- Kurtz, S., Churchwell, E., & Wood, D. O. S. 1994, *ApJS*, **91**, 659
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, **379**, 1599
- Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, *MNRAS*, **391**, 136
- Molinari, S., Swinyard, B., Bally, J., et al. 2010, *PASP*, **122**, 314
- Mottram, J. C., Hoare, M. G., Davies, B., et al. 2011, *ApJL*, **730**, L33
- Peeters, E., Hony, S., Van Kerckhoven, C., et al. 2002, *A&A*, **390**, 1089
- Petrov, L., Kovalev, Y. Y., Fomalont, E. B., & Gordon, D. 2011, *AJ*, **142**, 35
- Roberts, J. A., Cooke, D. J., Murray, J. D., et al. 1975, *AuJPh*, **28**, 325
- Robitaille, T. P., Cohen, M., Whitney, B. A., et al. 2007, *AJ*, **134**, 2099
- Schwab, F. R. 1984, *AJ*, **89**, 1076
- Sewifo, M., Churchwell, E., Kurtz, S., Goss, W. M., & Hofner, P. 2011, *ApJS*, **194**, 44
- Smith, E. C. D., & McLean, I. S. 2008, *ApJ*, **676**, 408
- Smith, N., Povich, M. S., Whitney, B. A., et al. 2010, *MNRAS*, **406**, 952
- Steer, D. G., Dewdney, P. E., & Ito, M. R. 1984, *A&A*, **137**, 159
- Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, *AJ*, **132**, 1158
- Urquhart, J. S., Busfield, A. L., Hoare, M. G., et al. 2007, *A&A*, **461**, 11
- Urquhart, J. S., Hoare, M. G., Purcell, C. R., et al. 2009, *A&A*, **501**, 539
- Urquhart, J. S., Moore, T. J. T., Hoare, M. G., et al. 2011, *MNRAS*, **410**, 1237
- van Buren, D., Mac Low, M.-M., Wood, D. O. S., & Churchwell, E. 1990, *ApJ*, **353**, 570
- White, R. L., Becker, R. H., & Helfand, D. J. 2005, *AJ*, **130**, 586
- White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1997, *ApJ*, **475**, 479
- Wood, D. O. S., & Churchwell, E. 1989, *ApJS*, **69**, 831
- Wright, N. J., Drake, J. J., Drew, J. E., & Vink, J. S. 2010, *ApJ*, **713**, 871