

A global view on star formation: The GLOSTAR Galactic plane survey

XI. Radio source catalog IV: $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$

S.-N. X. Medina^{1,2,*}, S. A. Dzib², J. S. Urquhart³, A. Y. Yang^{4,5}, A. Brunthaler², K. M. Menten²,
F. Wyrowski², W. D. Cotton⁶, A. Cheema², R. Dokara², Y. Gong², S. Khan², H. Nguyen²,
G. N. Ortiz-León^{7,2}, M. R. Ruge^{2,6,8}, V. S. Veena^{2,9}, H. Beuther¹⁰, T. Csengeri¹¹,
J. D. Pandian¹², and N. Roy¹³

¹ German Aerospace Center, Scientific Information, 51147 Cologne, Germany

² Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
e-mail: sdzib@mpi-fr-bonn.mpg.de

³ Centre for Astrophysics and Planetary Science, University of Kent, Canterbury, CT2 7NH, UK

⁴ National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, PR China

⁵ Key Laboratory of Radio Astronomy and Technology, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, PR China

⁶ National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

⁷ Instituto Nacional de Astrofísica, Óptica y Electrónica, Apartado Postal 51 y 216, 72000 Puebla, Mexico

⁸ Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA

⁹ I. Physikalisches Institut, Universität zu Köln, Zùlpicher Str. 77, 50937 Köln, Germany

¹⁰ Max Planck Institute for Astronomy, Koenigstuhl 17, 69117 Heidelberg, Germany

¹¹ Laboratoire d'astrophysique de Bordeaux, Univ. Bordeaux, CNRS, B18N, allée Geoffroy Saint-Hilaire, 33615 Pessac, France

¹² Department of Earth & Space Sciences, Indian Institute of Space Science and Technology, Trivandrum 695547, India

¹³ Department of Physics, Indian Institute of Science, Bangalore 560012, India

Received 6 March 2024 / Accepted 13 July 2024

ABSTRACT

Context. The GLOSTAR view on STAR formation in the Milky Way (GLOSTAR) survey studies star formation with the Very Large Array (VLA) and the Effelsberg 100 meter radio telescope in the Galactic plane between $-2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$, and the Cygnus X region ($76^\circ < \ell < 83^\circ$ and $-1^\circ < b < 2^\circ$), with unprecedented sensitivity in both flux density ($1\sigma \sim 50 \mu\text{Jy beam}^{-1}$) and the capability of detecting emission with angular scales in the range from $1''$ to the largest radio structures in the Galaxy on the order of a few degrees in size.

Aims. Here, we provide a complete GLOSTAR-VLA D-configuration radio source catalog for the part of the Galactic disk covered. A catalog for the “pilot region” between $28^\circ < \ell < 36^\circ$ has been published in a previous paper and here we present the complementary catalog for the area within $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$.

Methods. Observations were taken with the VLA in a 4–8 GHz band to image 100 square degrees of the inner Galactic disk at a reference frequency of 5.8 GHz, using a total of 260 h of telescope time. We determined spectral indices (α ; $S_\nu \propto \nu^\alpha$) inside the observed band and in the frequency range of 1.4–5.8 GHz by complementing our results with those from The HI/OH/Recombination line survey of the inner Milky Way (THOR), which covers 1–2 GHz.

Results. The final images have an angular resolution of $18''$ and an average sensitivity of $123 \mu\text{Jy beam}^{-1}$. The sensitivity is better ($\sim 60 \mu\text{Jy beam}^{-1}$) in areas free of extended emission. The complementary Galactic disk catalog presented in this work consists of 11 211 radio sources. Of these, 1965 are known large-scale structure sources such as star-forming region complexes, well-known supernova remnants (SNRs), SNR candidates, or parts thereof. The remaining 9227 are discrete individual sources. Source parameters—namely flux densities, sizes, spectral indices, and classifications—are reported. We identify 769 H II region candidates, 359 of which have been newly classified as such. The mean value of spectral indices of 225 H II regions is $+0.14 \pm 0.02$, consistent with most of them emitting optically thin thermal radio emission. Combining our results with the previously published catalog of the pilot region, the final GLOSTAR-VLA D-configuration catalog contains 12 981 radio sources.

Key words. techniques: interferometric – catalogs – surveys – stars: formation – stars: massive – H II regions

1. Introduction

The wide-field radio continuum surveys have been fundamental in significantly increasing the number of known radio objects,

particularly sources representing high-mass young stellar objects still embedded in or close to dense gas. These objects, when far away in the Galactic plane, are typically obscured by large volumes of dust, making observations of stars in the earliest evolutionary stages difficult (see a review in high-mass star

*Corresponding author; smedina@mpi-fr-bonn.mpg.de

formation by Motte et al. 2018). Therefore, the GLOSTAR view on STAR formation in the Milky Way (GLOSTAR)¹ survey (Medina et al. 2019; Brunthaler et al. 2021) focuses on finding and characterizing star-forming regions in the Galactic plane at radio frequencies with unprecedented sensitivity using the Karl G. Jansky Very Large Array (VLA) and the Effelsberg 100 meter telescope. The GLOSTAR survey detects tell-tale tracers of star formation: compact, ultra- and hyper-compact H II regions and class II methanol masers that trace different stages of early stellar evolution. The capabilities of the GLOSTAR survey help to locate the center of early stages of star-forming activity and complement previous radio surveys like MAGPIS (the Multi-Array Galactic Plane Imaging Survey; Helfand et al. 2006), CORNISH (Co-Ordinated Radio ‘N’ Infrared Survey for High-mass star formation; Hoare et al. 2012), and THOR (The HI/OH/Recombination line survey of the inner Milky Way; Beuther et al. 2016) that have also largely contributed to star formation studies. Furthermore, the sensitivity achieved by GLOSTAR will highly increase and confirm the number of known sources within the Galactic plane.

To achieve its goal, the GLOSTAR survey covers a large area of the Galactic plane within 145 deg^2 , covering $-2^\circ < \ell < 60^\circ$; $|b| < 1^\circ$, which includes the Galactic center, and, in addition, the Cygnus X region ($76^\circ < \ell < 83^\circ$; $-1^\circ < b < 2^\circ$). It uses the results of a complementary set of observations with the VLA of the National Radio Astronomy Observatory (NRAO)² and the Effelsberg 100 m telescope. With this combination, the GLOSTAR survey can detect compact radio sources and recover extended emission. The GLOSTAR-VLA observations used the wideband C-band (4–8 GHz) receivers. They were performed using both the compact D-configuration (as well as complementary DnC- and C-configurations) and the extended B-configuration (as well as complementary BnA- and A-configurations) to obtain good surface brightness sensitivity and a range of resolutions to map the radio emission over a large range of angular scales. The C-band also comprises the frequency of the prominent class II CH₃OH (methanol) maser line (6.668 GHz), the 4.829 GHz transition line H₂CO (formaldehyde), and several radio recombination lines (RRLs). Spectrally resolved data have also been taken within the GLOSTAR survey.

The GLOSTAR survey results can be found in a series of publications, starting with the first catalog of radio continuum sources by Medina et al. (2019) based on the VLA D-configuration data of the GLOSTAR “pilot region” (defined within the following limits: $28^\circ < \ell < 36^\circ$ and $|b| < 1^\circ$). An overview of the GLOSTAR survey is presented by Brunthaler et al. (2021), who additionally provided examples of the combination of the VLA and the Effelsberg observations. At the same time, Dokara et al. (2021) presented a study of Galactic supernova remnants (SNRs), Nguyen et al. (2021) presented radio continuum detections of young stellar objects (YSOs) in the Galactic center, and Ortiz-León et al. (2021) reported 6.7 GHz methanol maser detections in the Cygnus X region, followed by Nguyen et al. (2022) who reported methanol maser detections within $-2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$. The GLOSTAR VLA B-configuration data were used by Dzib et al. (2023) to provide the first radio source catalog of compact sources in the pilot region and by Yang et al. (2023) for the region within $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 40^\circ$, $56^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$.

Moreover, GLOSTAR VLA and Effelsberg data were recently used by Dokara et al. (2023) to analyze nonthermal synchrotron emission from SNR candidates in the GLOSTAR pilot region, and by Gong et al. (2023) for a deep analysis of formaldehyde absorption in the Cygnus X region, while Khan et al. (2024) reports RRLs related to bright H II regions of the full survey.

The focus of this paper is the characterization of continuum radio emission from GLOSTAR VLA D-configuration data (angular resolution $\sim 18''$ and mean sensitivity $\sim 128 \mu\text{Jy beam}^{-1}$) to provide a reliable catalog of radio sources with angular sizes of $\sim 18''$ to a few arcminutes. The work presented in this paper extends the analysis already conducted on the radio continuum map of the pilot region (Medina et al. 2019, in which we reported 1575 radio sources) to the rest of the inner Galactic disk covered by the survey. We follow a similar set of steps to create a complete catalog of the Galactic disk covered by the GLOSTAR survey ($2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$).

This paper is organized as follows. Section 2 gives a description of the observations and the obtained data. Section 3 describes the resulting image of the radio continuum emission and the method used for source extraction. Section 4 discusses the catalog construction, and Section 5 summarizes and discusses the catalog properties. In Section 6, we give a summary of the work and outline our main findings.

2. Observations and data reduction

The observations were performed with the VLA in the D-, DnC-, and C-configurations (see Table 1). The hybrid DnC configuration was preferred for targets with low elevation (that is, $\ell < 12^\circ$) to recover a synthesized beam similar in shape to that obtained with D-configuration observations at higher elevations. However, the use of the hybrid configurations was discontinued in 2016. To compensate for this, the regions not observed in the DnC-configuration were observed in both D- and C-configurations, as was suggested by NRAO³. Though diverse array configurations were used, for the simplicity of convention, we refer to all these low-resolution VLA observations as the GLOSTAR VLA D-configuration observations. A detailed overview of the observing strategy, data reduction, and imaging of the observed data is given in Brunthaler et al. (2021), and more specific details on the calibration and imaging of the D-configuration data are described in Medina et al. (2019). While we refer the reader to these works for details, a summary is provided below.

The observed band corresponds to the so-called C-band, covering a frequency range from 4.0 to 8.0 GHz. In particular, the correlator was tuned to record two 1.0 GHz wide sub-bands centered at 4.7 and 6.9 GHz in a semi-continuum mode⁴. The portion of the Galactic disk for which results are reported in this paper has a size of 100 square degrees and is delimited by $|b| \leq 1^\circ$ and $2^\circ \leq \ell \leq 28^\circ$ and $36^\circ \leq \ell \leq 60^\circ$. The results for the areas corresponding to the Galactic center region ($-2^\circ < \ell < +2^\circ$) and the Cygnus X region are not included in this work and are left for future GLOSTAR catalog releases. A total of 52 observing sessions were performed to cover smaller areas using the hexagonal

³ https://science.nrao.edu/facilities/vla/docs/manuals/propvla/array_configs

⁴ The observations also comprised additional high spectral resolution correlation setups covering spectral lines (see Brunthaler et al. 2021). In this paper, however, we focus on the continuum data.

¹ <https://glostar.mpifr-bonn.mpg.de/glostar/>

² The NRAO is operated by Associated Universities Inc. under cooperative agreement with the National Science Foundation.

mosaicing scheme of pointed observations (Medina et al. 2019; Brunthaler et al. 2021)⁵.

The sessions observed while the VLA was in its D- and DnC-configurations covered rectangular areas with sizes of two degrees in the Galactic latitude direction and one degree in the Galactic longitude direction, using ~ 650 target pointings (spending ~ 22 s per target pointing). C-configuration observations, on the other hand, covered 1.5 degrees in the Galactic longitude direction and observed ~ 950 target pointings (spending ~ 11 s per target pointing).

Quasars were observed for calibration purposes; in particular, 3C 286 and 3C 48 were observed for amplitude and bandpass calibrations. A gain calibrator was observed for each session and chosen to be at an angular distance of $\lesssim 10.0^\circ$ from the region covered in that session. The observations are part of the VLA projects with IDs 11B-168, 13A-334, 15B-175, and 17A-197. The total telescope time used for these observations was 260 h (that is, five hours per session). The observations occurred from December 2011 to June 2017. Individual observation dates, corresponding VLA project ID, VLA configuration, and used gain calibrators are listed in Table 1.

The data sets were edited, calibrated, and imaged using the *Obit* software package (Cotton 2008), designed for handling radio astronomical data⁶. Prior to calibration, the data sets were edited to remove the first 10 seconds of the scans on calibrators, which are affected by the antennas slewing. Then, a standard calibration is applied. This includes amplitude corrections based on the switched power signal, group delay, bandpass, amplitude and phase, and polarization calibration. The standard calibration was alternated with the editing of data affected by instrumental problems or radio frequency interference. Then, the data were reset and calibrated again, excluding the identified problematic data.

The images were produced using the *Obit* task MFImage. First, the data for every target pointing are imaged individually, covering the FWHM primary beam corresponding to the lowest continuum band, 4.7 GHz, which is $8''.4$. In cases for which the target pointing covered strong sources located outside the primary beam that are bright enough to affect the imaging, we add outlier fields at the location of these sources to account for them in the cleaning process. The observed bandwidth is divided into frequency bins that are narrow enough for the effects of variable spectral index and antenna pattern variations to be minor, enabling a joint spectral deconvolution. MFImage then generates an image for each individual frequency bin, and these images are then weighted averaged to produce the final map of the target field. The images are produced using a Briggs' weighting with robustness parameter 0.0. Finally, the frequency bins and the weighted averaged map are combined to produce the mosaiced images with a restoring beam of $18''$ and a pixel size of $2''.5$ as described by Medina et al. (2019) and Brunthaler et al. (2021). The shortest baseline of the VLA in the D-configuration is 35 m. Thus, sources with angular sizes larger than $\sim 4'$ are filtered out from the images. In order to conserve computational resources, the total observed area was divided into six sub-mosaics that

Table 1. All observed epochs of the GLOSTAR-VLA survey presented in this work.

ℓ ($^\circ$) range	VLA Project ID	Date of Obs. dd/mm/yyyy	VLA Conf.	Phase calibrator
02–03	15B-175	17/01/2016	DnC	J1820–2528
03–04	15B-175	21/01/2016	DnC	J1820–2528
04–05	15B-175	22/01/2016	DnC	J1820–2528
05–06	15B-175	16/01/2016	DnC	J1820–2528
06–07	17A-197	03/04/2017	D	J1820–2528
06–7.5	17A-197	31/05/2017	C	J1820–2528
07–08	17A-197	31/03/2017	D	J1820–2528
08–09	17A-197	21/02/2017	D	J1820–2528
7.5–09	17A-197	05/06/2017	C	J1820–2528
09–10	15B-175	24/01/2016	DnC	J1820–2528
10–11	13A-334	16/05/2013	DnC	J1820–2528
11–12	13A-334	17/05/2013	DnC	J1820–2528
12–13	17A-197	03/03/2017	D	J1825–0737
13–14	17A-197	04/04/2017	D	J1825–0737
14–15	17A-197	19/02/2017	D	J1825–0737
15–16	14A-420	14/07/2014	D	J1825–0737
16–17	14A-420	24/07/2014	D	J1825–0737
17–18	14A-420	05/08/2014	D	J1825–0737
18–19	14A-420	14/08/2014	D	J1825–0737
19–20	14A-420	12/07/2014	D	J1825–0737
20–21	14A-420	23/07/2014	D	J1825–0737
21–22	14A-420	28/07/2014	D	J1825–0737
22–23	14A-420	27/07/2014	D	J1825–0737
23–24	14A-420	26/08/2014	D	J1825–0737
24–25	14A-420	16/07/2014	D	J1825–0737
25–26	14A-420	29/07/2014	D	J1825–0737
26–27	14A-420	13/08/2014	D	J1825–0737
27–28	14A-420	28/08/2014	D	J1825–0737
36–37	14A-420	07/07/2014	D	J1907+0127
37–38	14A-420	04/07/2014	D	J1907+0127
38–39	14A-420	01/08/2014	D	J1907+0127
39–40	14A-420	25/08/2014	D	J1907+0127
40–41	14A-420	07/08/2014	D	J1907+0127
41–42	14A-420	02/07/2014	D	J1907+0127
42–43	14A-420	09/07/2014	D	J1907+0127
43–44	14A-420	17/07/2014	D	J1907+0127
44–45	14A-420	03/08/2014	D	J1907+0127
45–46	14A-420	29/06/2014	D	J1907+0127
46–47	15B-175	25/11/2015	D	J1922+1530
47–48	15B-175	13/11/2015	D	J1922+1530
48–49	15B-175	21/11/2015	D	J1922+1530
49–50	15B-175	14/11/2015	D	J1922+1530
50–51	15B-175	22/11/2015	D	J1922+1530
51–52	15B-175	11/11/2015	D	J1922+1530
52–53	15B-175	20/11/2015	D	J1922+1530
53–54	15B-175	10/11/2015	D	J1922+1530
54–55	15B-175	27/11/2015	D	J1922+1530
55–56	15B-175	17/12/2015	D	J1922+1530
56–57	15B-175	28/11/2015	D	J1925+2106
57–58	15B-175	08/11/2015	D	J1925+2106
58–59	11B-168	15/12/2011	D	J1931+2243
59–60	11B-168	29/12/2011	D	J1931+2243

⁵ Note that the values given for the spacing and primary beam size in Sect. 2.1 of Medina et al. (2019) have incorrect units; they were given in arcseconds, whereas the correct units should be arcminutes. The hexagonal grid spacing is $\theta_{\text{hex}} = 3'.25$ and the FWHM primary beam size at the higher frequency band, 6.9 GHz, is $\theta_{\text{b}} = 45/(\nu [\text{GHz}]) = 6'.5$.

⁶ *Obit* applications can be accessed through a Python interface (*ObitTalk*). *Obit* also inter-operates with classic AIPS (The Astronomical Image Processing System; Greisen 2003) and has access to its tasks.

constitute the final radio continuum map. The full continuum image of the Galactic disk area covered by the GLOSTAR survey, $2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$ (i.e., including the pilot region), is shown in Figure 1.

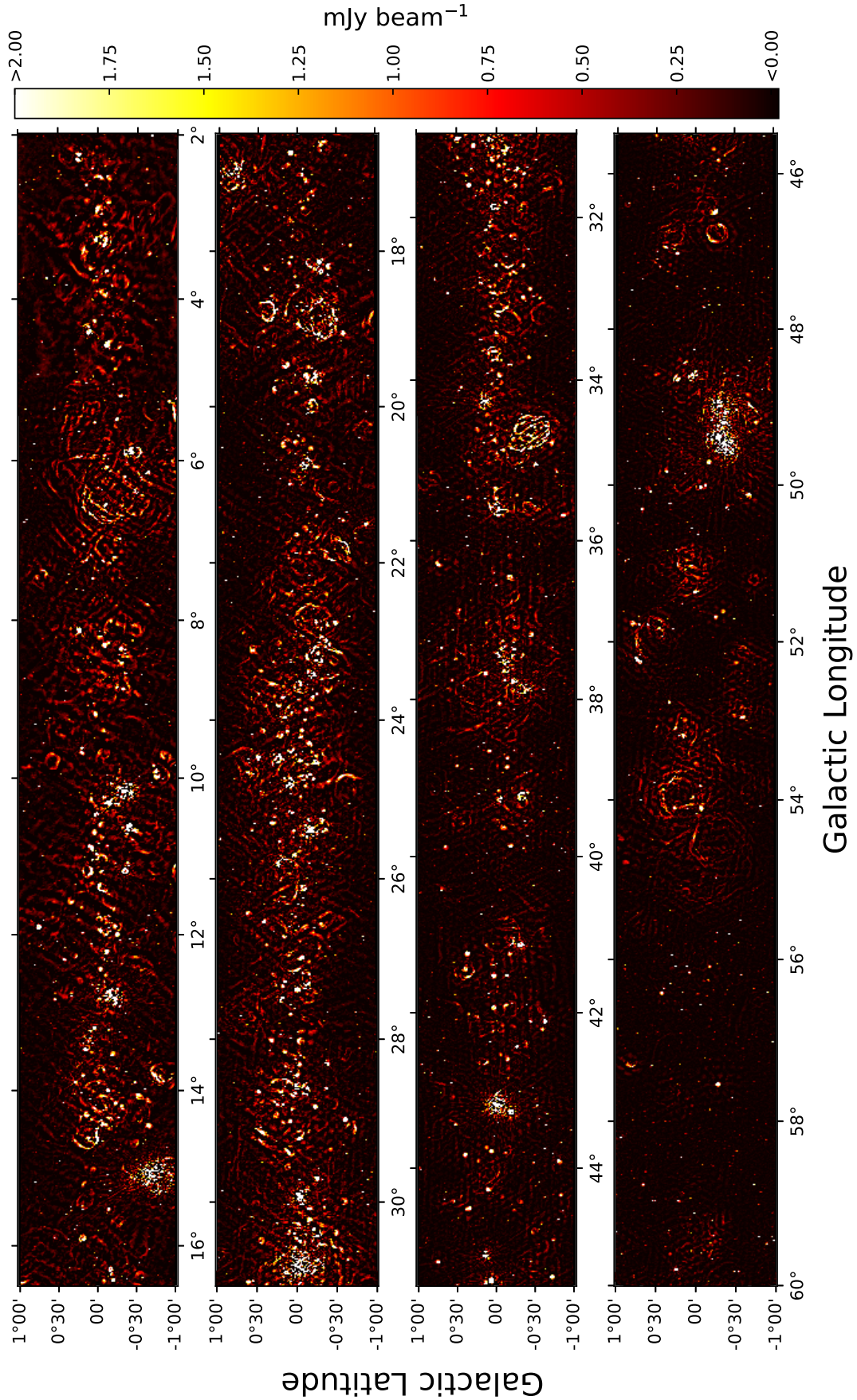


Fig. 1. Full GLOSTAR radio continuum map at 5.8 GHz of 116 square degrees of the Galactic disk plane ($2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$). Though the most significant peak emission has a brightness of $\sim 10 \text{ Jy beam}^{-1}$, we have constrained the contrast by choosing an upper limit of 2 mJy beam^{-1} , in order to make most of the weak extended and compact sources recognizable.

Table 2. Average noise values estimated for the seven mosaic maps produced of the survey area across the full GLOSTAR radio map.

Galactic longitude range	σ_{rms} [$\mu\text{Jy beam}^{-1}$]	Radio sources	LSS
$2^\circ\text{--}12^\circ$	171	2026	31
$12^\circ\text{--}20^\circ$	165	1655	19
$20^\circ\text{--}28^\circ$	157	1888	29
$28^\circ\text{--}36^\circ$ (*)	150	1770(**)	27
$36^\circ\text{--}44^\circ$	119	1756	8
$44^\circ\text{--}52^\circ$	98	1766	13
$52^\circ\text{--}60^\circ$	84	2120	5
$2^\circ\text{--}28^\circ$; $36^\circ\text{--}60^\circ$ (***)	123	11211	106
$2^\circ\text{--}60^\circ$	123	12981	132

Notes. (*)Corresponding to the pilot region, previously reported by Medina et al. (2019). (**)This number includes 195 sources related to the 27 LSS, as well as the 1575 individual sources. (***)This work.

3. Analysis of the continuum map

3.1. Radio continuum map

The observational data previously described were used to image 100 square degrees of the Galactic plane by producing six sub-mosaics (see Table 2 for each longitude range, including the pilot region). Together with the sub-mosaic of the pilot region, previously reported by Medina et al. (2019), we have a complete map of 116 square degrees of the Galactic plane (that is, $2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$). Figure 1 presents the radio continuum map of the complete region produced from the D, DnC, and C-configuration data. This radio map has an effective central frequency of 5.8 GHz and has been restored using a circular beam of $18''$. Although the observed region extends slightly beyond $|b| = 1.0^\circ$, the noise increases significantly in these outer regions because of the primary beam attenuation and because these areas do not overlap with other pointings. Hence, sources detected beyond this latitude range have been excluded from the analysis presented here, although they are listed in our final catalog.

Inspection of Fig. 1 reveals the presence of thousands of compact sources and many large-scale structures (LSSs, see Sect. 3.2 for details) associated, for example, with prominent star-forming complexes (e.g., W31, W33, W49, and W51) or SNRs (e.g., W28 and W44). The emission towards these more complex regions is not fully recovered due to the lack of short (u, v) baselines, which also results in strong negative bowls that affect the noise level around them. Therefore, these LSSs cause significantly higher noise levels in the Galactic mid-plane and the inner part of the Galactic disk, where they are more densely concentrated, and affect the quality of the imaging around them. This is demonstrated in Table 2 where the mean noise values determined from each of the mosaics can be seen to decrease with longitude from the Galactic center to the outer part of the disk where less star formation is occurring. In Fig. 2 we present a histogram of the pixel values from the whole survey region. The significant noise variations across the GLOSTAR VLA D-configuration map produce the non-Gaussian profile of the pixel noise values. We use a Gaussian fit to this pixel distribution to estimate the overall sensitivity of the final map. The standard deviation of the distribution of the pixel values, σ_{rms} , is $123 \mu\text{Jy beam}^{-1}$; however, the noise values can be significantly higher towards prominent complexes concentrated towards the mid-plane. The noise level in regions free of extended emission is approximately $60 \mu\text{Jy beam}^{-1}$.

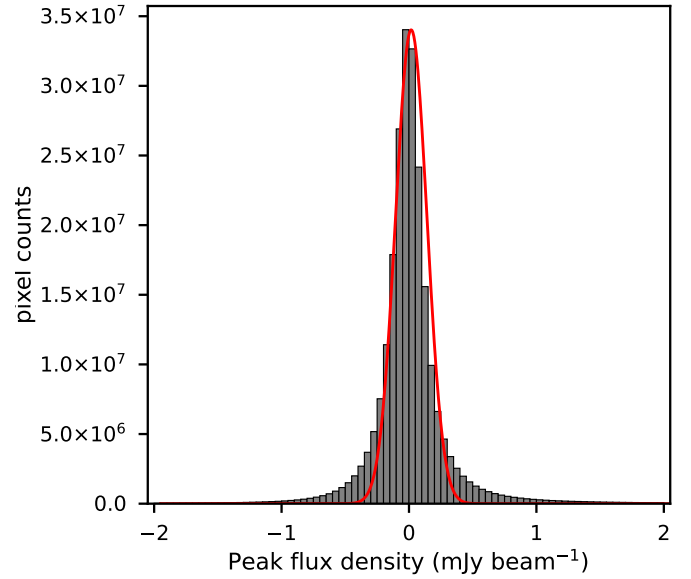


Fig. 2. Noise distribution of the full GLOSTAR VLA D-configuration mosaic. The range of pixel values ranges between -2 and 2 mJy . The red line shows the results of a Gaussian fit to the distribution, $\sigma_{\text{rms}} = 123 \mu\text{Jy beam}^{-1}$. The bin size is $50 \mu\text{Jy beam}^{-1}$.

3.2. Source extraction

The range of detected structures, level of complexity, variation, and survey coverage make source extraction and analysis difficult in the GLOSTAR low-resolution images presented here. In our previous analysis of the pilot region, we developed a number of steps to automatically identify sources and confirm them visually and by spatially correlating them with published source catalogs at different wavelengths (Medina et al. 2019). In this paper, we followed the same process to create a catalog of radio sources for the rest of the covered Galactic plane, which were then combined with the sources detected in the pilot field to create a complete catalog of radio sources in the GLOSTAR VLA D-configuration radio map (excluding the Galactic center and the Cygnus X regions); that is $2^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$. We have again used the BLOBCAT software package to perform the source extraction; see Hales et al. (2012) for a detailed description of this Python code. This code has also been previously used for the source extraction of THOR, a complementary 21 cm radio continuum survey (Beuther et al. 2016; Wang et al. 2018), which will facilitate further comparisons. Below we provide a brief outline of the method used and refer the reader to Medina et al. (2019) for more details. The methodology steps are also visualized in Fig. 3.

Creation of the noise maps. To perform the automatic source extraction, we first need to produce an independent noise map of the region due to the position-dependent nature of the noise in GLOSTAR VLA D-configuration radio continuum maps. We use the *rms* estimation algorithm within the SExtractor package (Bertin & Arnouts 1996; Holwerda 2005). This algorithm defines the *rms* value for each pixel in an image by determining the distribution of pixel values within a local mesh until all values are around a chosen sigma value. Following our previous work Medina et al. (2019), for the calculations we used a detection and analysis threshold of 5σ , a minimum size of 5 pixels, and a mesh size of 80×80 pixels². As a result, most real emission is removed from the noise image, and the determined noise map represents the correct noise level.

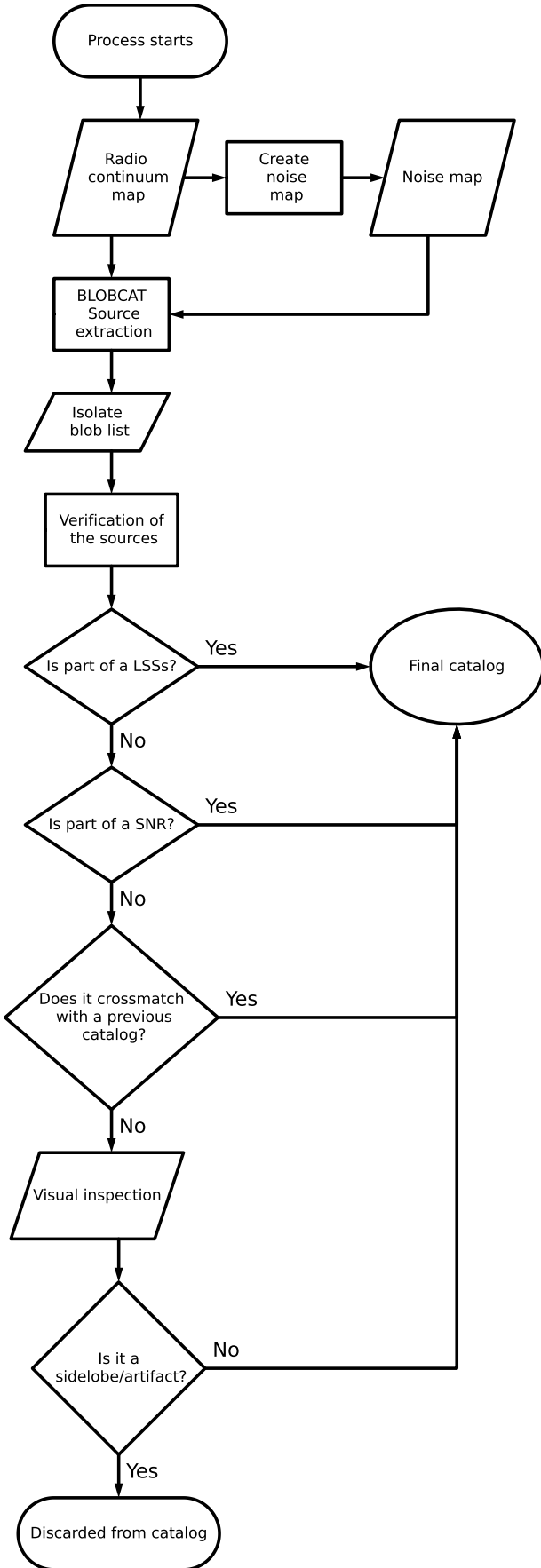


Fig. 3. Methodology of the source extraction and verification followed for the GLOSTAR-VLA survey data to create the final catalog.

Automatic source extraction with BLOBCAT. We employ the BLOBCAT software package that uses an algorithm to detect and identify blobs, or islands of pixels representing sources, in two-dimensional astronomical radio-wavelength images (see Hales et al. 2012, for details). The algorithm makes a pixel-by-pixel comparison between the main image and the background noise map, locates pixels above a given threshold, and defines “blobs” around these pixels. Then, a first source parameter set is obtained by fitting a 2D elliptical Gaussian. Later, the peak and integrated flux densities are corrected by considering biases from the Gaussian fits, such as source morphology (see Hales et al. 2012, for details). BLOBCAT returns several parameters from which we use the following (and their corresponding errors) for our catalog construction: the 2D source position (RA_p and Dec_p; notice these are software internal names and not the image coordinate system), peak flux density (S_p_CBBWS), integrated flux density (S_int_CB), and the number of pixels covered by the source (npix).

Following our previous work, we applied a detection threshold (dSNR) of 4σ and a minimum source size of 5 pixels in diameter ($\sim 13''$). This resulted in the extraction of 13 001 blobs potentially associated with radio sources in the area delimited by $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$. All of them have flux above four times that in the local noise map and have a size comparable to or larger than the beam. However, this sample contains a significant number of artifacts that need to be removed, and also emission from LSSs; that is, sources fragmenting into multiple components that need to be grouped. Furthermore, at the edges of the map, the noise increases significantly; therefore, the sources located at $|b| > 1^\circ$ are excluded from further analysis and their identification names are marked with an *. The following steps will address these issues. Additionally, in our previous work, we also flagged (that is labeled) sources that appeared to consist of two or more distinct emission peaks; however, in this work, we have not attempted to flag them or to separate the emission because in our previous work, only around 3% of the sources are affected by this problem, and those are resolved in the GLOSTAR VLA B-configuration images (e.g., Dzib et al. 2023). Finally, as in our previous works (e.g., Medina et al. 2019), we considered the Y-factor ($Y_{\text{factor}} = S_{\nu, \text{Int}}/S_{\nu, \text{Peak}}$) and roughly divided the sources into extended ($Y_{\text{factor}} > 2.0$), compact ($1.2 < Y_{\text{factor}} \leq 2.0$) and unresolved ($Y_{\text{factor}} \leq 1.2$).

Identification of large-scale structures. In our previous work, we defined LSSs as over-resolved structures whose emission is broken up into multiple independent fragments of radio sources. To identify LSSs, we performed a visual inspection and compared the radio emission distribution with their corresponding $8.0 \mu\text{m}$ maps extracted from the Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE) archive. Most GLOSTAR sources with corresponding $8.0 \mu\text{m}$ have thermal radio emission; that is they are H II regions. SNRs are also generally faint or even not detected at mid-infrared (MIR) wavelengths (Dokara et al. 2021). In Fig. 4 we show some examples of these LSSs. We delimited a rectangular area around the coherent infrared structures morphologically correlated with the radio emission (see the left panel of Fig. 4). Also, we have delimited a rectangular area around radio emission sources with shell-like morphology that are clearly part of one source even if there is no clear correlation with the $8.0 \mu\text{m}$ emission, which is the case for many SNRs (see the lower right panel of Fig. 4). Finally, we also classified as LSSs the complexes with two or more independent sources that clearly belong to the same star-forming region as the one seen in the $8.0 \mu\text{m}$ maps (see the top right panel of Fig. 4).

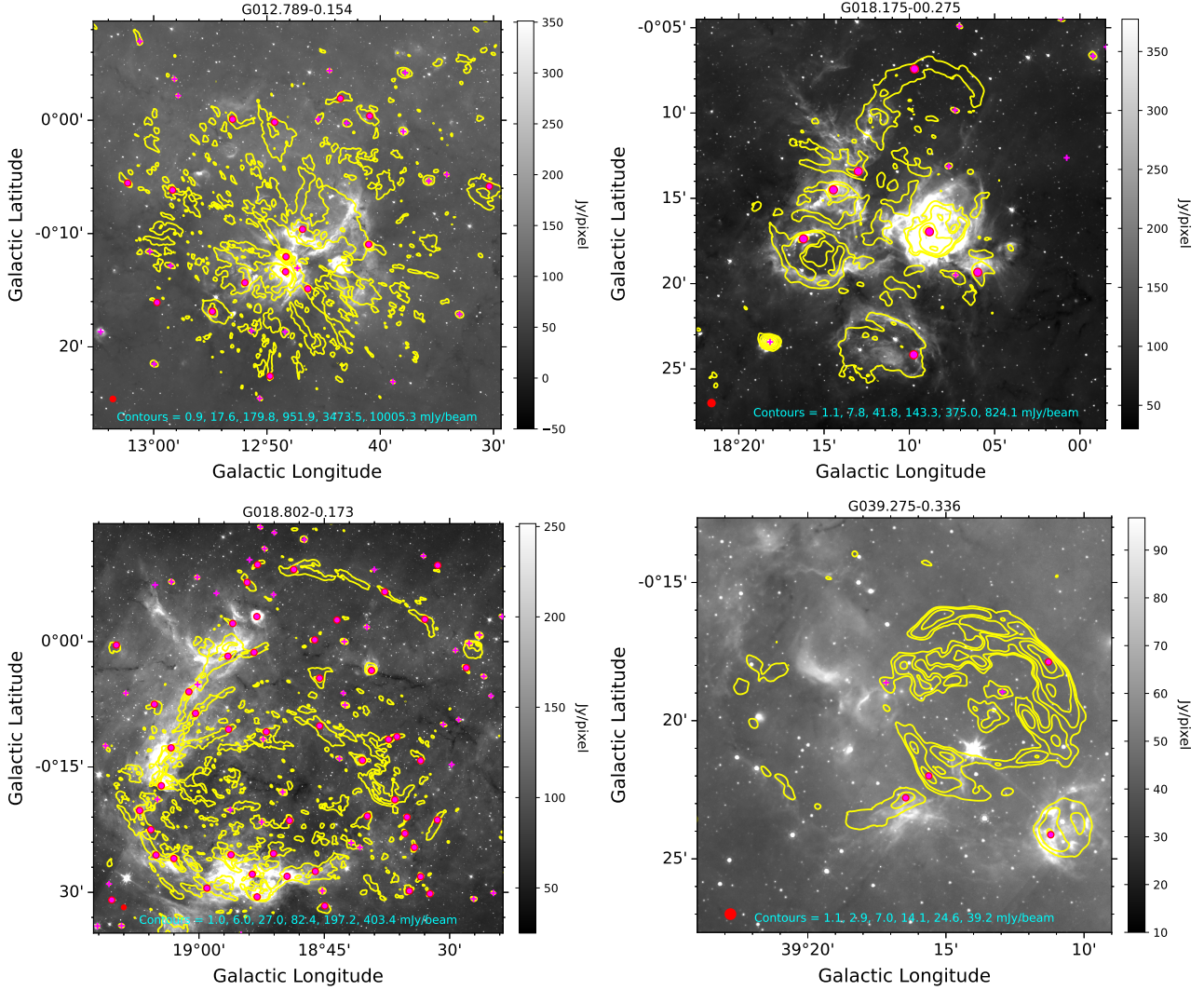


Fig. 4. Examples of radio emission associated with large-scale structures. The grayscale background represents GLIMPSE 8 μm emission, while the yellow contours are the GLOSTAR 5.8 GHz radio continuum emission. These regions often show coherent infrared structures that are morphologically correlated with most of the radio emission (see examples presented in the upper panels). Although this correlation is not always present, the different radio emission blobs are clearly related with each other (see examples presented in the lower panels). The contours are determined as described in the figures, and the red-filled circle shown in the lower left corners shows the GLOSTAR VLA D-configuration beam FWHM. The magenta-filled circles indicate the peak positions of radio sources associated with these regions, and magenta crosses indicate compact radio sources in the area that are treated as separate sources.

All extended radio sources in these boxed regions are considered to be part of one LSS. Individual fragment components and their properties are given in the catalog; however, they have been excluded from the statistical analysis presented here. Compact and unresolved radio sources that are considered unlikely to be associated with LSSs have been retained in the final catalog as distinct sources, since these are possible individual sources and not part of the extended emission. They are included in our further analysis.

In this work, we have visually identified a total of 106 large-scale radio emission regions containing 839 individual fragments within $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$, which includes the SNRs like W28, W30, and W49 (e.g., Green 2019) and star-forming region complexes such as W31 (Ghosh et al. 1989), W33 (Messineo et al. 2011), M16 (Oliveira 2008), M17 (Chini & Hoffmeister 2008), W39 (Kerton et al. 2013), and W51 (Ginsburg et al. 2017).

In Table 3, we give the list of LSSs, including the 27 from the pilot region. The name of each structure has been constructed from the central coordinates of the enclosing box. The table includes the extent of the boxed region in ℓ and b and the integrated flux of the total emission in the box minus the emission from any compact sources that are considered as distinct sources. The total integrated flux density of each LSS, is calculated by adding the flux densities of all components together. The measured parameters of LSSs should be used with caution as the flux densities and sizes are unlikely to correspond to discrete sources, but rather to source complexes. We also point the readers to Dokara et al. (2021) and Dokara et al. (2023) for a discussion on the calculation of flux densities of sources identified as SNRs.

Crossmatch with recently identified SNR candidates. SNRs are sources of nonthermal radio emission (Anderson et al.

2017) that can also be detected with the GLOSTAR observations (see Medina et al. 2019; Dokara et al. 2021, 2023). The radio emission from most of these radio sources is extended and, in many cases, has a weak surface brightness. Because of these properties, automatic source extraction is not optimal to fully recover these sources. Their identification and total flux determination require a more detailed visual inspection that goes beyond the scope of this work. SNRs and SNR candidates in the observed area of the GLOSTAR survey are the subject of the studies presented by Dokara et al. (2021) and Dokara et al. (2023). However, our procedure recovered sources located in the regions containing well-known SNRs and previously identified SNR candidates. Previously, well-known SNRs with large angular scales were identified as LSSs (see also Table 3). Additionally, we also identified 1126 extended sources ($Y_{\text{factor}} > 2.0$) located in areas of other SNR candidates, including the 80 new SNR candidates from the GLOSTAR survey (Dokara et al. 2021). These sources are labeled in the final catalog.

Cross match with previous catalogs. In order to verify the remaining sources, the next step was to assume that any radio source that has a counterpart in a published radio or MIR catalog is real. We used several complementary multiwavelength surveys to cross match with our detected radio sources to identify possible counterparts. The selected surveys are better described in the following sections. They either have comparable angular resolution: GLIMPSE (Fazio et al. 2004), WISE (Wide-Field Infrared Survey Explorer; Wright et al. 2010), HiGAL (the Herschel infrared Galactic Plane Survey; Molinari et al. 2010, 2016), ATLASGAL (the APEX Telescope Large Area Survey of the Galaxy; Schuller et al. 2009), GLOSTAR VLA B-configuration (Yang et al. 2023), CORNISH (Hoare et al. 2012), RMS (the Red MSX Source survey; Urquhart et al. 2009), and THOR (Beuther et al. 2016; Wang et al. 2020) all have angular resolutions between 6.0 and 20". The CORNISH and RMS VLA surveys, both at 5.0 GHz, comparable to the GLOSTAR survey's wavelength, have resolutions of 1.5" and 2", respectively. We used an initial cross match radius of 10", the same as chosen in our previous paper (Medina et al. 2019).

The offset distribution between GLOSTAR VLA D-configuration sources and the above catalogs was analyzed based on the empirical cumulative distribution function (ECDF), see Figure 5. The offsets between GLOSTAR and other radio centimeter wavelength surveys (THOR, CORNISH, and RMS) in Fig. 5 show a good position match. This is because they are tracing radio emission with similar morphology. For ATLASGAL, GLIMPSE, and WISE, the situation is different. ATLASGAL 870 μm submillimeter continuum sources trace compact high column density emission of cold dust (and associated molecular gas) in star-forming regions and are quite rare. Thus, a radio source the direction of an ATLASGAL source may be related to it even if the angular separation between the ATLASGAL and GLOSTAR positions is larger than 10". The cases of the WISE and GLIMPSE catalogs are more complex: due to the high density of foreground field stars detected in these survey there is a significant probability of false line of sight matches with the GLOSTAR detections. As a precaution, a smaller matching radius of 6" was used for WISE and GLIMPSE in our final counterpart search.

We have found 8235 radio sources with at least one counterpart at a 10" cross match radius. In Table 4, we summarize the counterparts of the GLOSTAR-VLA sources with the complementary multiwavelength surveys.

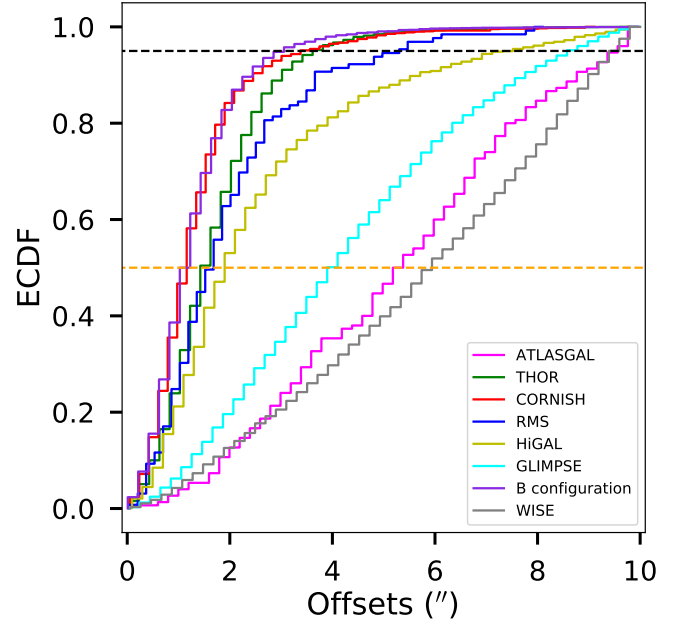


Fig. 5. Empirical cumulative density function of the position offsets among GLOSTAR compact sources ($Y\text{-factor} \leq 2.0$) with the other survey counterparts. The dashed lines indicate the 50% (orange) and 95% (black) of matches.

Visual inspection. After our cross match analysis, we find 2807 blobs to be unrelated to extended sources and to have no counterparts in any other survey. To understand the nature of these sources, we first compare the statistics with the expected number of false detections. As discussed in our previous works (e.g., Medina et al. 2019; Yang et al. 2023) the number of false detections can be calculated using the complementary cumulative distribution function $\Phi(x) = 1 - \phi(x)$, where $\phi(x)$ is the cumulative distribution function. Assuming that the noise in our radio maps follows a Gaussian distribution,

$$\phi(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (1)$$

where erf is the error function given by

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt. \quad (2)$$

$\Phi(x)$ is the probability that the value of a random variable x with a standard normal distribution will exceed the value x . Considering the 100 square degrees of the analyzed radio map, the expected false detection is calculated like $\theta^2 \times \phi(x)$, where θ is the synthesized beam size. Therefore, considering sources above 4σ , we expect only 127 false detections, which is much smaller than the 2807 sources without counterparts.

Next, we carried out a visual inspection to identify and remove side lobes (which appear as elongated blobs in the direction of a very bright source) and artifacts (weak sources, between $4\text{--}5\sigma$, and have a size much smaller than the beam size). This led to the removal of 1790 blobs. We also identified 19 sources as unclear; these sources have a signal-to-noise ratio (S/N) between 4.0 and 5.0; they have no counterpart and do not comply with the characteristics of a side-lobe or an artifact; however, they are located in a noisy region or close to a very bright source.

Table 4. Statistics of the matches between GLOSTAR VLA D-configuration and other published surveys.

Survey	Ref.	Wavelength	Resolution ($''$)	rms level mJy beam^{-1}	Num. of sources in GLOSTAR region	Num. of matches ($10''$)	Num. of matches ($6''$)
THOR	(1)	20 cm	25	0.3–1	6337	5179	5092
RMS	(2)	6 cm	1.0–2.0	0.22	737	150	145
CORNISH	(3)	6 cm	1.5	0.33	1754	1231	1221
GLOSTAR B configuration	(4)	6 cm	1.0	0.08	5437	2496	2473
ATLASGAL	(5)	870 μm	19.2	50–70	3096	198	109
HiGAL	(6)	70–500 μm	5.8–35	...	35 344	1096	918
WISE	(7)	3.4–22 μm	6–12	...	2 426 781	5473	3034
GLIMPSE	(8,9)	3.6–8.0 μm	1.4–2	...	18 100 849	7326	5525

Notes. A match radius of $10''$ and $6''$ has been used, centered on the peak of the GLOSTAR emission. References are: (1) Wang et al. (2018), (2) Urquhart et al. (2009), (3) Hoare et al. (2012), (4) Yang et al. (2023), (5) Schuller et al. (2009), (6) Molinari et al. (2016), (7) Wright et al. (2010), (8) Churchwell et al. (2009) and (9) (Fazio et al. 2004).

Table 5. Summary of detection categories.

Description	Number of sources
Total number of extracted sources	13 001
Number of sources assoc. with large structures	839
Number of sources assoc. with SNR	1126
Number of sources with a counterpart	8235
Number of sources with $Y_{\text{factor}} > 3$	297
Number of recovered real sources	701
Number of unclear sources	19
Number of artifacts or side lobes ^(a)	1790
Number of sources in final catalog	11 211

Notes. ^(a)Excluded from final catalog.

Finally, we identified 701 as real sources, and included all detections with a S/N ratio ≥ 5 that are not elongated (and thus a not part of sidelobes). Table 5 summarizes the numbers of detected sources ascribed to the different categories.

4. Final catalog

After removing all artifacts extracted by BLOBCAT, the final catalog consists of 11 211 entries. From these, 9 254 represent discrete sources; the remaining are part of LSS and SNR candidates. In this section, we use the discrete sources and compare them with previous surveys in order to quantify the reliability of their properties based on our source extraction process (see Sect. 3.2). Additionally, we use the information from the counterparts of previous radio and infrared (IR) surveys to classify the sources.

4.1. Astrometry

To check the quality of the GLOSTAR source positions, we have compared the positions of unresolved ($Y_{\text{factor}} < 1.2$) sources that have counterparts in the radio fundamental catalog of compact radio sources⁷. These radio sources are quasars that were observed with Very Long Baseline Interferometry (VLBI) and

⁷ This catalog is provided via the project webpage at <https://astrogeo.smce.nasa.gov>. Responsible NASA (National Aeronautics and Space Administration) official: L. Petrov.

their celestial positions are known with accuracies better than a few milli-arcseconds.

Using a position matching of $6''$ between both catalogs, we found 55 sources in common. In the upper panel of Fig. 6, we show the ℓ and b position offsets between GLOSTAR D-configuration and VLBI sources. The mean values of the position offsets are $-0''34 \pm 0''10^8$ and $+0''31 \pm 0''11$ in Galactic longitude and latitude, respectively. The standard deviations of the offsets are $0''8$ in both directions. Thus, the positions of the GLOSTAR catalog presented in this paper are accurate to within $1''$.

A second test can be done by comparing the GLOSTAR sources with the CORNISH catalog (Purcell et al. 2013). CORNISH used the VLA prior to its upgrade to observe at 5.0 GHz an area similar to that covered by us. Most of the sources in this catalog are expected to be background extragalactic objects, which is also the case observed for GLOSTAR. No detectable proper motions are expected from these sources. We have also used a matching radius of $6''$, and we have restricted our analysis to sources that are unresolved in both catalogs. In the lower panel of Fig. 6, the measured position offsets of 668 sources meeting these criteria are shown. The mean values of the position offsets are $-0''30 \pm 0''04$ and $+0''28 \pm 0''04$ in ℓ and b , respectively. The standard deviation of the offsets is $0''80$ in both Galactic coordinates. This result is in agreement with that found for the VLBI sources.

4.2. Flux density levels

In this section, we check the accuracy of flux density measurements. By using the 668 CORNISH unresolved sources detected in the GLOSTAR survey, we can check the quality of the measured flux densities in GLOSTAR (Medina et al. 2019; Dzib et al. 2023; Yang et al. 2023). This can be done since most of these sources are background extragalactic objects whose radio emission is expected to have a low degree of variability.

In the upper panel of Fig. 7, the flux densities of the CORNISH sources are plotted against the GLOSTAR flux densities. For a visual guide, the equality line is also shown. This panel shows that the radio emission from both catalogs is consistent. In the lower panel of Fig. 7, the distribution of the ratios of the flux

⁸ The errors reported on the mean values, here and through the manuscript, are estimated using the standard error of means (SEM = σ/\sqrt{N} , where σ is the standard deviation and N is the number of elements in the sample).

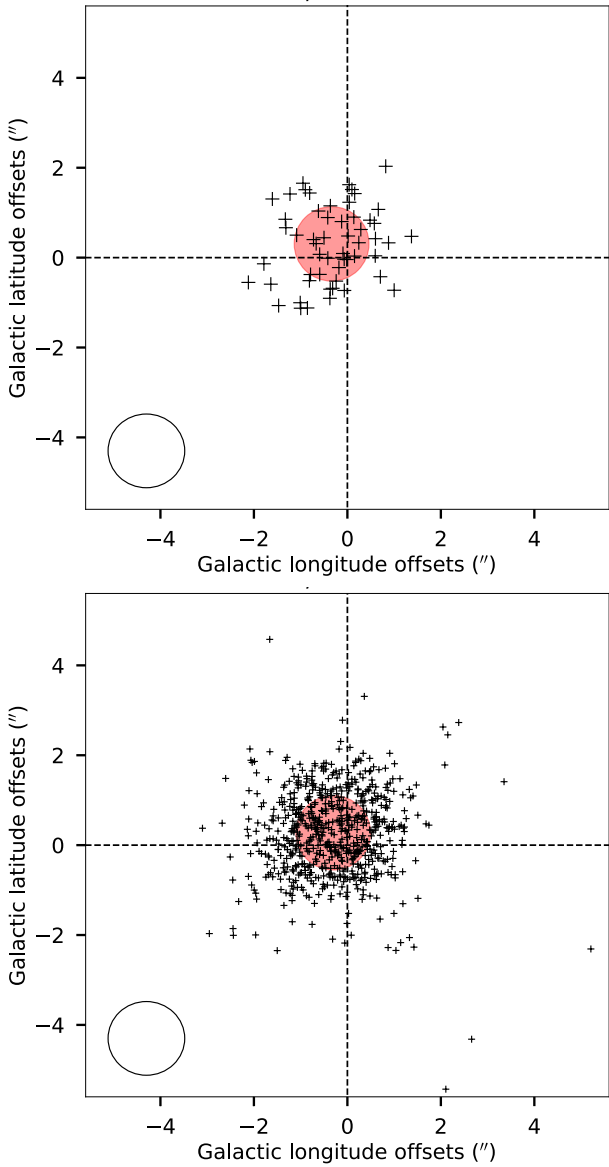


Fig. 6. Position offsets between GLOSTAR unresolved sources (Y factor ≤ 1.2) and VLBI extragalactic sources (upper panel), and CORNISH unresolved sources (lower panel). Red circles are centered in the mean values of the offsets, and the radii sizes are equal to the standard deviations. The circles in the bottom left corners of both panels represent the mean position error of GLOSTAR-VLA D-configuration radio sources reported in this work, which is $1''.6$.

densities of sources from both catalogs are shown, with mean and standard deviation values of 1.01 ± 0.02 and 0.42 , respectively. Taking a 3σ value, it is then concluded that the fluxes of the GLOSTAR survey have an accuracy better than 6%.

4.3. Source effective radius

The BLOCAT software returns the number of pixels within a source (output parameter `npix`). This number does not contain any information on the structure, elongation, or position angle of the source. While source size cannot be recovered from

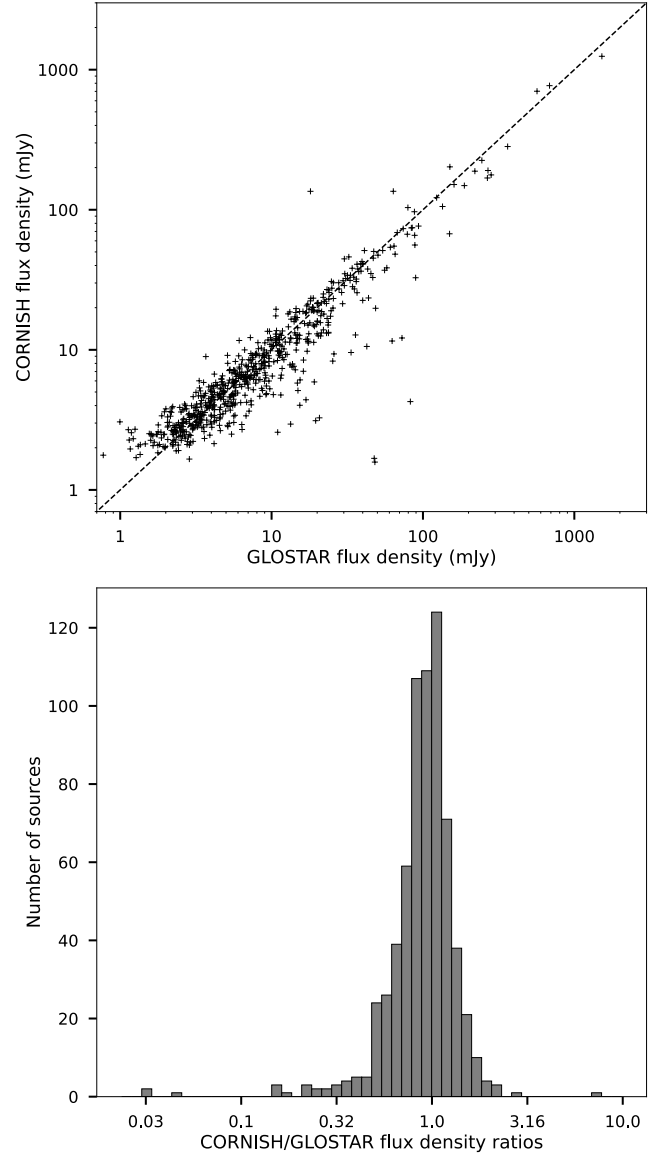


Fig. 7. Upper panel: comparison of the flux densities of CORNISH unresolved sources that are also detected as unresolved sources in the GLOSTAR survey (this work). The dashed black line is the line of equality. Lower panel: flux density ratio distribution of sources plotted in the upper panel. The mean value of the flux density ratios is 1.01 ± 0.02 and the standard deviation is 0.42 .

BLOCAT, following the strategy by Medina et al. (2019) we can determine the effective radius of a source.

We aim to estimate the radius of a circular source that covers the same number of pixels as the source extracted using BLOCAT. The area of a single pixel in the GLOSTAR D-configuration images is $2''.5 \times 2''.5$. The area, A , covered by a source is n_{pix} times the pixel size squared, and the effective source radius is given by

$$R_{\text{source}} = \sqrt{\frac{A}{\pi}}.$$

The effective radius, as obtained from this procedure, is listed for all sources in Table 6.

4.4. Spectral indices

The flux density of radio sources as a function of the observed frequency can be expressed in a power law form:

$$S_\nu \propto \nu^\alpha, \quad (3)$$

where S_ν is the flux density at the observed frequency, ν , and α is the spectral index. The sensitive and wideband GLOSTAR-VLA observations allow an in-band spectral index (α_{GLOSTAR}) to be determined. We followed the successful strategy used in the other GLOSTAR-VLA catalogs (Medina et al. 2019; Dzib et al. 2023; Yang et al. 2023). First, to avoid possible spectral index biases, we constrain our spectral index determinations to sources with $S/N \geq 10$ and $Y_{\text{factor}} \leq 2.0$; that is, compact sources. A lower limit for S/N was chosen to ensure that the source could be detected in most of the seven frequency bin images. For instance, as the noise level per frequency bin image is expected to be $\sqrt{7}$ times larger than that of the combined image, for sources with $S/N=10$ we expect detection of the order of $10/\sqrt{7} = 3.8\sigma_{\text{bin}}$ in each frequency bin image. This rough value ensures detections in most frequency bin images. For extended sources, as the (u, v) -coverage is different across the band, the fraction of total flux density recovered at the lower sideband can greatly exceed that at the upper sideband. In these cases, the spectral index derived will have apparently steeper (more negative) spectra to values that can be physically impossible. Second, source flux densities are measured on the imaged frequency bins (see Sect. 2). We use the logarithmic form of Equation (3), which is a linear equation, and the measured flux densities in the individual frequency bin images to perform a least-squares fitting and determine the spectral index. Following these steps, we have estimated the spectral indices of 3 968 sources.

In Dzib et al. (2023) and Yang et al. (2023), we have compared the spectral indices determined from GLOSTAR data with those determined by the THOR survey (Wang et al. 2018) for compact common sources. We have found a good consistency between the spectral index determination by both surveys. The THOR survey covered the full L -band (1.0–2.0 GHz) with the VLA, and they also split their observed bandwidth into smaller frequency chunks to estimate the in-band spectral index (see Wang et al. 2020, for details). Though THOR observed at lower frequencies than GLOSTAR, the resulting angular resolution of their images is similar to that of the GLOSTAR-VLA D-configuration images (18'' for GLOSTAR and 25'' for THOR). Thus, following Medina et al. (2019), we now proceed to determine the source spectral indices from the peak flux densities measured by GLOSTAR and THOR ($\alpha_{\text{GLOSTAR-THOR}}$). We consider mid-frequencies of 1.5 GHz and 5.8 GHz for THOR and GLOSTAR, respectively. As only two flux density values are considered, the spectral index can be determined as

$$\alpha_{\text{GLOSTAR-THOR}} = \frac{\ln(S_{\text{GLOSTAR}}/S_{\text{THOR}})}{\ln(5.8/1.5)},$$

where S_{THOR} and S_{GLOSTAR} are the peak flux densities from the respective survey. The spectral index uncertainty is calculated using:

$$\sigma_{\alpha_{\text{GLOSTAR-THOR}}} = \frac{\sqrt{(\sigma_{S_{\text{THOR}}}/S_{\text{THOR}})^2 + (\sigma_{S_{\text{GLOSTAR}}}/S_{\text{GLOSTAR}})^2}}{\ln(5.8/1.5)},$$

which is obtained following standard error propagation theory. $\sigma_{S_{\text{GLOSTAR}}}$ and $\sigma_{S_{\text{THOR}}}$ are the source flux density uncertainties for

GLOSTAR and THOR. In the case of the THOR survey, their catalog does not contain direct values for flux density uncertainties. However, since their S/N ratio is given, the THOR flux density uncertainty can be estimated as

$$\sigma_{S_{\text{THOR}}} = \frac{S_{\text{THOR}}}{S/N}.$$

With this procedure, we estimate $\alpha_{\text{GLOSTAR-THOR}}$ for 4 127 radio sources, of which 1 308 do not have an estimated value of α_{GLOSTAR} .

The total number of radio sources with an estimated radio spectral index, either the GLOSTAR-inband or GLOSTAR-THOR, is 5 276. Results of both α_{GLOSTAR} and $\alpha_{\text{GLOSTAR-THOR}}$ on individual sources are listed in Table 6 and they will be compared in Sect. 5.3.

4.5. Counterparts in other surveys

During the last few decades, Galactic Plane surveys have addressed many aspects of high-mass star ($>8 M_\odot$) formation. As these surveys image large areas of the Galaxy, they provide unbiased samples of star-forming regions with different properties (see, for example Urquhart et al. 2018, 2022; Elia et al. 2017).

As we have seen in previous sections, information from other radio surveys, such as CORNISH and THOR, can corroborate and complement the source parameters from our catalog. Additionally, information at shorter wavelengths can be used to give insights into the nature of the observed radio sources. In the context of GLOSTAR, particularly interesting radio sources are those related to star formation. CORNISH and THOR were briefly described in previous sections and in the following we describe the other Galactic plane surveys used to characterize the GLOSTAR radio sources.

Recently, we have presented the GLOSTAR VLA B-configuration catalog of the Galactic plane using higher angular resolution (1'') images (Yang et al. 2023). The imaged field covers the area $|b| \leq 1.0^\circ$, and the Galactic longitude ranges $2^\circ < \ell < 40^\circ$ and $56^\circ < \ell < 60^\circ$; that is, 32 square degrees less than the GLOSTAR VLA D-configuration. About 5500 radio sources were reported and classified. Using a crossmatch radius of 10'', we found a match for 2497 compact radio sources.

The RMS survey was a multiwavelength project aiming to identify massive young stellar objects (MYSOs) in the Galactic plane (Hoare et al. 2005; Lumsden et al. 2013). Using a multiwavelength classification scheme, RMS identifies the nature of radio sources; particularly within the northern hemisphere they identified 79 PNe and 391 H II regions with radio emission (Urquhart et al. 2009). We retrieved 150 of the RMS sources within a search radius of 10''.

Emission at sub-millimeter wavelengths is dominated by dense cool dust and gas, which is intimately related to star formation. At these wavelengths, ATLASGAL (Schuller et al. 2009) is the first high-resolution ($\approx 20''$ FWHM) ground-based sub-millimeter (870 μm) survey of the thermal dust emission in the entire inner Galactic plane. The ATLASGAL survey has presented $>10\,000$ dense clumps in the Galactic plane (Contreras et al. 2013; Csengeri et al. 2014; Urquhart et al. 2014). Correlating radio sources (such as those detected in GLOSTAR) with ATLASGAL clumps is an excellent way to identify embedded or dust-enshrouded objects such as UC-H II regions (see, for example, Medina et al. 2019; Irabor et al. 2018; Urquhart et al. 2007, 2009, 2013; Purcell et al. 2013). Correlating our radio source

catalog with the ATLASGAL compact source catalog (Urquhart et al. 2014), we find cross match 281 sources within 10".

HiGAL is a photometric survey in five far-infrared (FIR) bands between 70 and 500 μm (Molinari et al. 2016). Its observations covered the whole Galactic plane with a varying latitude range (Elia et al. 2021). The FWHM beam sizes range from 6" to 35", and the mean position uncertainty is 1".2 (Molinari et al. 2016). Using a cross matching radius of 10" we found 1527 sources in both our GLOSTAR and the HiGAL catalogs.

WISE was a NASA IR-wavelength astronomical space telescope mission (Wright et al. 2010) that mapped the entire sky in four MIR bands W1, W2, W3, and W4 centered at 3.4, 4.6, 12, and 22 μm , respectively, using a 40 cm telescope feeding an array with a total of 4 million pixels; these wavelengths correspond to angular resolutions of 6".1, 6".4, 6".5, and 12". The WISE All-sky release source catalog contains the source properties of ~563 million sources. The WISE All-sky release source catalog has been filtered out for sources with $S/N < 5$, spurious detections, and image artifacts⁹. Using a radius of 10" for cross matching GLOSTAR radio sources and sources from the WISE All-sky release source catalog, we found 7298 common sources. However, as was mentioned earlier, because of the high density of WISE detected sources, mostly foreground field stars, there is a significant probability of false matches and we have reduced the cross matching criteria to 6" and found 3 034 matching sources.

Finally, we have also used the GLIMPSE point source catalog¹⁰ (Churchwell et al. 2009). This legacy project was conducted with the Spitzer space telescope and observed the shorter wavelength part of the MIR range; that is, at 3.6, 4.5, 5.8, 8.0 μm . GLIMPSE is composed of several surveys, and particularly interesting for GLOSTAR are GLIMPSEI and GLIMPSEII which, combined, effectively observed the Galactic plane from $-70^\circ < \ell < 65^\circ$ and $|b| < 1^\circ$. Using a cross matching criterion of 10", we found 7 326 GLIMPSE counterparts to radio sources. However, similar to the WISE catalog, a large number of these are expected to be foreground objects. Thus, we reduced the cross match radius to 6" for the final counterpart search and found 5 525 counterparts.

4.6. Source classification

The present catalog is composed of 11 211 radio sources, and thus one-by-one source classification, as performed in our previous work (i.e., Medina et al. 2019), is a time-consuming task. Furthermore, in our previous catalog work, we have shown that the classification of radio sources detected within the GLOSTAR-VLA project is largely (>90%) compatible with the classifications made by other radio surveys (Medina et al. 2019; Dzib et al. 2023; Yang et al. 2023). Notably, in the GLOSTAR-VLA pilot region, the source classifications in the high-resolution and low-resolution maps are also highly consistent (Dzib et al. 2023). Thus, our source classification approach for the present catalog is slightly different and is presented below. We focus the source classification analysis on the 9227 sources that are not related to LSSs (839 sources) or SNRs (1126 sources) or Unclear (19).

⁹ Detailed information on WISE data processing and source catalogs (including rejected detections) can be found in <https://wise2.ipac.caltech.edu/docs/release/allsky/>

¹⁰ The GLIMPSE point Source Catalog (GLMC) is the most reliable of GLIMPSE catalogs with a reliability $\geq 95\%$ (Churchwell et al. 2009). The GLMC can be found at the following sites: <http://ssc.spitzer.caltech.edu/legacy/glimpsehistory.html> and <http://irsa.ipac.caltech.edu/data/SPITZER/GLIMPSE>

4.6.1. Source classification from other surveys

First, we identify sources with counterparts in other catalogs and use their classification. We prioritize the catalogs in the following order: GLOSTAR-VLA B-configuration (Yang et al. 2023), CORNISH (Purcell et al. 2013), RMS (Urquhart et al. 2009), WISE H II regions (Anderson et al. 2014), and THOR (Wang et al. 2020). We also check the GLOSTAR-VLA 6.7 GHz methanol maser catalog (Nguyen et al. 2022), as this maser line is only detected toward MYSOs.

The above catalogs use different names for different classes of sources. To homogenize them, we have used the classifications chosen in previous GLOSTAR-VLA catalogs. These classifications are *HII* for H II regions and H II region candidates, *PN* for planetary nebulae and planetary nebula candidates, *EgC* for background extragalactic candidate sources, *Psr* for pulsars, *Radio-star* for stars with radio emission, and *Unclassified* for sources that could not be classified within our classification scheme. The CORNISH survey used the terms UCHII and HII-Region, which, in our final classification, are renamed as HII. Sources CORNISH classified as Radio Galaxy (Central Source), Radio Galaxy (lobe), Galaxy, and IR-Quiet are renamed as EgC. Sources classified as H II region and HII/YSO in the RMS survey are categorized as HII, while YSO and Evolved star classifications are renamed as Radio-star in our catalog. Regardless of the H II region sub-classification from Anderson et al. (2017), we rename these sources as HII. From the THOR survey, we discarded the classification X-ray (which only accounts for a counterpart at another wavelength) and renamed their *jet* classification to EgC (see also Wang et al. 2018). Sources associated with methanol masers are labeled as HII, as they represent objects in the very early phases of high-mass star formation. While most class II methanol masers are not associated with compact radio emission, some of them are: in the GLOSTAR D array data, Nguyen et al. (2022) find that toward 12% of the 554 detected methanol masers also radio continuum is detected, while 97% are associated with dust emission.

The number of classified radio sources from other surveys is 3302. The numbers of adopted classifications are 2286 from GLOSTAR-VLA B-configuration, 537 from CORNISH, 57 from RMS, 167 from WISE H II regions, 249 from THOR, and six sources from the 6.7 GHz methanol maser catalog. Classifications and references to the classification origins are given in Table 6. We note that a large fraction of GLOSTAR radio sources could not be classified by only using the information from other surveys.

4.6.2. Source classification from information at shorter wavelengths

After the cross match classification, there are still 5 925 unclassified sources. To classify them, we use the information obtained in surveys conducted at other than radio wavelengths. We follow similar criteria as used in the previous GLOSTAR-VLA catalogs.

First, compact ($Y_{\text{factor}} \leq 2.0$) radio sources that do not have counterparts at submm, FIR, and MIR wavelengths are likely to be background extragalactic sources (Hoare et al. 2012; Lucas et al. 2008; Marleau et al. 2008; Medina et al. 2019), and thus they are classified as EgC. By matching these criteria, we classified 1569 radio sources as EgC.

The *HII* region candidate classification is assigned to radio sources that have a counterpart at submm wavelengths or that have a counterpart only at FIR wavelengths (Hoare et al. 2012; Anderson et al. 2012; Urquhart et al. 2013; Yang et al. 2021).

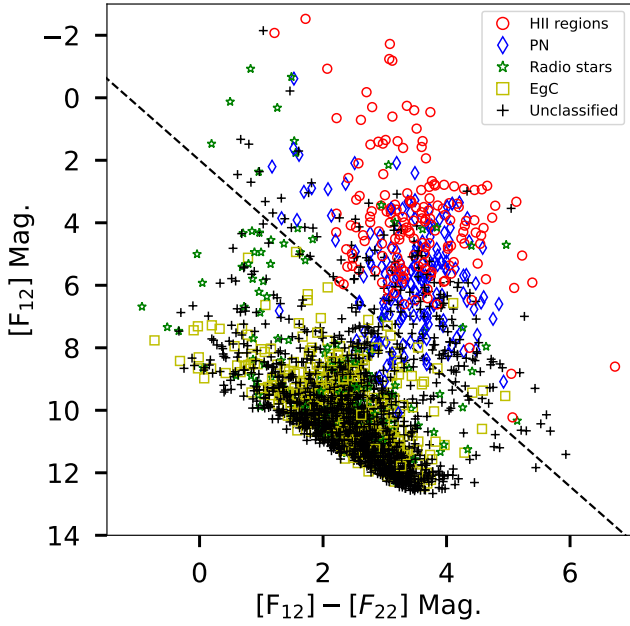


Fig. 8. Mid-infrared color-magnitude diagram of GLOSTAR radio sources with a WISE counterpart. The dashed line indicates the constraint used by Medina et al. (2019) to distinguish between different types of objects.

There are 32 radio sources with counterparts in the ATLASGAL submm survey, and 82 sources with a counterpart only in the HiGAL FIR survey. All of these 114 sources are classified as H II region candidates.

If radio sources have counterparts only at FIR and MIR wavelengths, additional criteria are needed to classify them. Extended sources can, for example, represent photo dissociated regions (PDRs), parts of SNRs, or parts of extended H II regions and complicate the classification procedure. While these may include interesting cases, for the purpose of this paper we concentrate on compact radio sources with the signatures of early phases of star formation. Thus, in the following, we focus on the classification of compact radio sources ($Y_{\text{factor}} \leq 2.0$). The 1 463 sources with $Y_{\text{factor}} > 2.0$ that have not been previously classified, are classified as *Unclassified* in our final catalog.

The 2 779 remaining unclassified compact radio sources have counterparts at FIR or MIR wavelengths. Medina et al. (2019) constructed [12] vs. [12]–[22] color-magnitude diagram using mid IR magnitudes from the WISE catalog and showed that in the resulting diagram the H II regions and PNe occupy a different area than radio stars and other unclassified objects (which have more likely an extragalactic origin, see discussion by Medina et al. 2019). As is discussed by Medina et al. (2019), H II regions and PNe fulfill the constraint (hereafter the color-magnitude constraint):

$$[12]\text{Mag.} < 1.74 \times ([12] - [22]) + 2, \quad (4)$$

while Radio-stars and EgC do not.

To illustrate the situation for the GLOSTAR radio sources discussed in this work, we use the flux information from WISE. A total of 2 401 GLOSTAR radio sources have counterparts detected in the 12 and 22 μm WISE bands. These include 178 H II regions, 175 PNe, 452 EgC, 159 radio-stars, and 1437 unclassified sources. In Fig. 8, we plot the color-magnitude diagram [12] vs. [12]–[22] of these sources, where the source types are

distinguished with different markers and colors. The line that follows the color-magnitude constraint is also plotted as a dashed line. Fig. 8 clearly shows that H II regions and PNe occupy different areas in this diagram than the radio stars and EgC objects.

From the already classified sources, we notice that 360 lie above the color-magnitude constraint line, and these are divided into 178 (50%) H II regions, 152 (42%) PNe, 7 (2%) EgC, and 23 (6%) radio stars. Thus, we can assume that sources that fall above the color-magnitude constraint are either H II regions or PNe. There are 153 unclassified sources fulfilling these criteria, and in this work we classify them as H II region candidates, labeled as HII in the final catalog; we notice that up to 50% will likely be PNe; however, given the current analysis, it is not possible to definitively characterize them.

On the other hand, 604 classified sources do not comply with the color-magnitude constraint and these are divided into 23 (4%) PNe, 445 (74%) EgC, and 136 (22%) radio stars. We notice that EgC sources constitute the majority in this region of the color-magnitude diagram. Following these counts, we classify the 1 284 previously unclassified sources in this area as EgC, noting that about 25% of them may likely be galactic sources; however, it is not possible to differentiate them with the current data.

The IR information of the remaining sources is scarcer than for the previous cases, and MIR colors cannot be obtained as fluxes are only reported in a single band. We noticed that 86 of these sources have at least a counterpart in one of the FIR bands. As FIR emission may trace cold dust, radio sources associated with FIR emission might be related to H II regions, and accordingly we classify them as H II region candidates (labeled as HII in the final catalog). However, further studies are required to confirm the nature of these sources. The 1 256 remaining sources have counterparts only in the MIR range. These can be either EgC or radio stars; however, our previous work (Medina et al. 2019; Dzib et al. 2023; Yang et al. 2023) has shown that most sources with only MIR and near-infrared counterparts are background extragalactic sources. Accordingly, these 1256 sources are classified as EgC. A summary of the classified sources per method is given in Table 7. Class and classification methods are given in the catalog and Table 6.

5. Summary of full catalog

The final catalog consists of 11 211 entries, 1965 of which are related to extended structures (for example, SNRs or extended H II regions) and 19 are sources of uncertain nature that could not be identified as real or spurious sources. Source classification was possible for most of the remaining sources, but 1468 extended sources ($Y_{\text{factor}} > 2$) were classified as *Unclassified* as their extended structures require a deeper analysis to classify them, which goes beyond the scope of the present work. In the following, we discuss the remaining classified sources. This discussion will focus on the spatial distribution, flux densities, sizes, and spectral indices of the sources. Also, we discuss the identified H II region candidates.

5.1. Spatial distribution

The Galactic disk hosts most of its star-forming regions in its inner parts. Basic source properties, namely intensity and angular size, depend on their distance, which is determined by their location in the Galaxy.

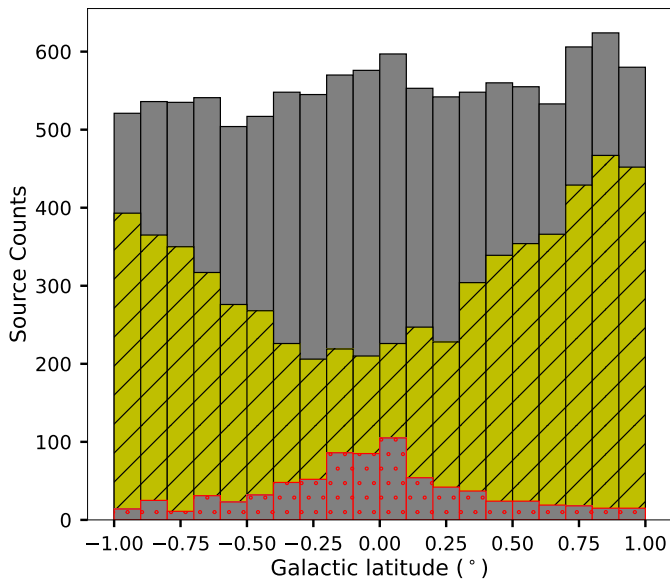
Table 6. GLOSTAR source catalog for the observed region.

GLOSTAR name	ℓ	b	S/N	S_{peak} (mJy beam ⁻¹)	$\sigma_{5\text{ peak}}$	S_{int} (mJy)	$\sigma_{5\text{ int}}$	Y_{factor}	Radius			Spectral index					Radio				submm				IR			Class	Ref.	Method (26)							
									α	$\Delta\alpha$	$\alpha^{\#}$	$\Delta\alpha^{\#}$	GB	C	R	R	T	MM	A	H	W	G	α	$\Delta\alpha$	$\alpha^{\#}$	$\Delta\alpha^{\#}$	GB				C	R	R	T	MM	A	H
G002.005+00.348	02.0047	+0.3479	09.2	1.29	0.16	1.34	0.16	1.0	13.10	EgC	1	No SMM or IR counterpart		
G002.007-00.681	02.0072	-0.6806	170.3	26.07	1.41	121.27	6.07	4.7	52.00	HII	1	Only FIR counterpart		
G002.011+00.108	02.0111	+0.1076	05.4	0.75	0.14	1.10	0.15	1.5	12.50	EgC	1	No SMM or IR counterpart		
G002.012+00.744	02.0120	+0.7438	178.3	15.20	0.82	13.67	0.69	0.9	21.10	PN	2	Previously known source		
G002.013+00.896	02.0125	+0.8959	05.3	0.46	0.09	0.57	0.09	1.2	11.10	EgC	1	Color-magnitude constrain		
G002.018+00.669	02.0177	+0.6688	32.2	3.12	0.19	3.52	0.20	1.1	20.20	Other	2	Previously known source		
G002.021-00.824	02.0210	-0.8243	19.0	1.92	0.14	1.65	0.13	0.9	14.40	HII	1	Color-magnitude constrain		
G021.103-00.976	21.1032	-0.9757	15.3	2.89	0.24	2.67	0.23	0.9	14.20	0.26	0.15	0.04	0.24	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Radio-star	2	Previously known source				
G021.103+00.664	21.1034	+0.6639	07.4	0.60	0.09	0.51	0.09	0.8	10.60	EgC	2	Previously known source	
G021.114+00.872	21.1137	+0.8723	05.1	0.43	0.09	0.44	0.09	1.0	9.70	-0.96	0.20	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Radio-star	2	Previously known source				
G021.121-00.956	21.1206	-0.9563	05.6	0.91	0.17	1.29	0.17	1.4	12.10	EgC	2	Previously known source	
G021.126-00.294	21.1258	-0.2944	17.2	5.42	0.43	204.65	10.24	37.7	106.60	0.11	0.12	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	LSS	1	Visual Inspection				
G021.130-00.247	21.1300	-0.2465	05.4	1.63	0.31	17.96	0.95	11.0	31.80	LSS	1	Visual Inspection	
G021.145+00.585	21.1445	+0.5854	05.9	0.49	0.09	0.68	0.09	1.4	12.60	EgC	1	No SMM or IR counterpart	
G047.145-00.166	47.1451	-0.1660	04.1	0.25	0.06	0.23	0.06	0.9	7.70	EgC	1	Only MIR counterparts	
G047.147+00.730	47.1471	+0.7299	20.1	1.36	0.10	11.02	0.56	8.1	46.90	SNR	7	Previously known SNR	
G047.147-00.113	47.1472	-0.1132	05.1	0.31	0.06	0.28	0.06	0.9	9.00	EgC	1	Only MIR counterparts	
G047.154+00.890	47.1540	+0.8903	24.6	1.52	0.10	1.72	0.11	1.1	18.90	-0.22	0.07	-0.64	0.30	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	EgC	1	Only MIR counterparts				
G047.162-00.513	47.1617	-0.5132	04.5	0.28	0.06	0.21	0.06	0.8	7.50	-0.55	0.22	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	No SMM or IR counterpart			
G047.171-00.621	47.1708	-0.6208	04.0	0.24	0.06	0.16	0.06	0.6	6.30	EgC	1	Color-magnitude constrain	
G047.172-00.481	47.1722	-0.4806	08.5	0.52	0.07	0.39	0.06	0.8	10.60	-0.52	0.13	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	Only MIR counterparts			
G059.932+00.837	59.9320	+0.8368	55.7	5.88	0.33	7.02	0.37	1.2	21.60	-0.47	0.04	-0.66	0.11	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	No SMM or IR counterpart			
G059.937-00.353	59.9365	-0.3528	05.2	0.88	0.18	0.93	0.18	1.1	10.00	-0.43	0.16	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	SNR	7	Previously known SNR			
G059.940-00.157	59.9399	-0.1569	39.4	6.92	0.41	7.06	0.39	1.0	18.60	-0.03	0.05	0.36	0.12	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	3	Previously known source			
G059.941-00.042	59.9413	-0.0417	28.6	4.86	0.31	5.58	0.33	1.1	18.80	-0.52	0.05	-0.88	0.23	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	No SMM or IR counterpart			
G059.948+00.555	59.9478	+0.5549	14.6	1.83	0.16	1.67	0.15	0.9	14.00	-0.13	0.09	-0.28	0.28	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	Color-magnitude constrain			
G059.963+00.036	59.9629	+0.0361	04.6	0.94	0.21	1.26	0.21	1.3	10.20	EgC	1	Color-magnitude constrain
G059.968+00.660	59.9680	+0.6597	07.1	1.17	0.18	0.89	0.17	0.8	11.00	-0.04	0.14	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	...	EgC	1	Color-magnitude constrain			

Notes. The columns are, from left to right, the source name, Galactic coordinates, the S/N, the peak and integrated flux densities and their errors, the Y-factor, the effective radius, the estimated spectral indices determined from the GLOSTAR and THOR flux densities and from only the in-band GLOSTAR sub-images (*). Columns 15 to 23 list if the source has a counterpart in other surveys. Radio: GB for GLOSTAR B-configuration catalog (Yang et al. 2023), C for CORNISH (Hoare et al. 2012; Purcell et al. 2013), R for RMS (Urquhart et al. 2009), T for THOR (Beuther et al. 2016; Wang et al. 2020), and MM for methanol maser surveys (Nguyen et al. 2022; Green et al. 2009). Submillimeter/FIR: A for ATLASGAL (Urquhart et al. 2018). Infrared: H for HiGAL (Molinari et al. 2016), W for WISE (Wright et al. 2010), and G for GLIMPSE (Churchwell et al. 2009). Column 24 lists the final classification of the radio sources. References in column 25 are as follows: (1) This work, (2) Yang et al. (2023), (3) Purcell et al. (2013), (4) Urquhart et al. (2009), (5) Anderson et al. (2014), (6) Wang et al. (2020), (7) SNR remnants catalogs by Dokara et al. (2021) or Green (2019), and (8) Methanol maser catalogs by Nguyen et al. (2022) or Green et al. (2009). Finally, column 27 lists the method used for the source classification. Only a small portion of the data is provided here, and the full table is available at the dedicated GLOSTAR webpage at <https://glostar.mpi.fr-bonn.mpg.de/glostar/>

Table 7. Summary of classification categories.

Class	Method	Number of sources
HII	Other catalog	410
	Methanol maser	6
	SMM counterpart	32
	Only FIR counterpart	82
	Color-magnitude constrain	153
	FIR counterpart, no color with MIR	86
	Total	769
PNe	Previously known source	240
Radio-star	Previously known source	425
PSR	Previously known source	6
PDR	Previously known source	1
EgC	Previously known source	2203
	No SMM or IR counterpart	1569
	Color-magnitude constrain	1284
	Only MIR counterparts	1256
[2pt]	Total	6312
Unclassified	Previously known source	19
	$Y_{\text{factor}} > 2$ & no previous class	1463
[2pt]	Total	1482


Fig. 9. Source distribution of GLOSTAR D-configuration sources in the Galactic latitude. Color bars are: full sample in gray, yellow for EgC sources and red for H II regions. The bin size is 0.1° .

Figures 9 and 10 show the distribution of all detected GLOSTAR sources. The overall distribution in the observed Galactic latitude range is robustly uniform (Fig. 9). When we only consider the EgC sources (yellow bars), it can be noticed that their number diminishes towards the Galactic mid-plane ($b = 0^\circ$). Moreover, as can be seen from Fig. 1, most of the extended sources are found around this latitude line, and it is also where most of the star formation is expected. The extended

structures have two effects on source identification: they confuse background sources and, in interferometric images with no zero-spacing information, they also increase the noise levels. Thus, background extragalactic sources that are intrinsically uniformly distributed in the sky show a decreasing number towards the mid-plane. On the contrary, H II regions (red bars) are more numerous towards low Galactic latitudes; that is, where most of the star formation in the Galaxy is occurring.

The source distribution in Galactic longitude, Fig. 10, also looks uniform for $\ell < 56^\circ$, with the source counts increasing at larger longitudes. The area with $\ell > 56^\circ$ contains a small number of LSSs, resulting in an improvement of the noise level (see also Table 2) and less compromised sources. As a consequence, a higher number of sources are detected at these longitudes. However, the number of H II regions in this area is also lower, indicating a lower number of star-forming sites. This is in concordance with the dearth of MYSOs that Zhang et al. (2019) noted in the segment of the Perseus arm stretching between $\ell \sim 50^\circ$ and 90° .

Finally, in Fig. 11, we show the complete spatial distribution of radio sources detected in the course of this work. We have included the sources with LSS and SNRs. This figure shows that the number of sources is lower in the presence of extended sources because of the effects mentioned earlier.

5.2. Fluxes and angular sizes

The source flux density is one of the most critical parameters in the GLOSTAR source catalog. By comparing our results with the CORNISH survey, we have shown that the flux density parameter from the GLOSTAR survey is accurate within 6% (see Section 4.2). The CORNISH survey, however, had a nominal noise level of ~ 0.4 mJy beam $^{-1}$, significantly higher than the GLOSTAR D-configuration images. In the top panel of Fig. 12, we show the source peak flux density distribution for GLOSTAR and CORNISH for the Galactic plane area covered by this work. The better sensitivity of the GLOSTAR survey resulted in a larger number of detected sources, indicating that most sources with a brightness lower than a few mJy beam $^{-1}$ are new detections.

The ratio $S_{\text{int}}/S_{\text{peak}}$ is also known as the Y -factor, a parameter that can be used to infer the size of a source. In the GLOSTAR survey, we have used it to define an unresolved source when the $Y_{\text{factor}} \leq 1.2$, a compact source when $1.2 < Y_{\text{factor}} \leq 2.0$, and an extended source when $Y_{\text{factor}} > 2.0$. In Fig. 13, we plot the Y -factor distribution of the GLOSTAR sources. Similarly, Fig. 14 shows the distribution of source effective radius. Both results show that our catalog mainly comprises compact and unresolved sources.

5.3. Spectral indices

Sources that exhibit radio emission produced from ionized thermal gas with temperature $> 10^4$ K (such as H II regions and PNe) have spectral indices ranging from -0.1 to 2.0 , representing optical thin and thick bremsstrahlung, respectively. On the other hand, non-thermal radio emission is produced in high-energy processes. The most common, resulting in synchrotron radiation, is produced by relativistic electrons spiralling around magnetic field lines and reaches brightness temperatures above 10^6 K. Active galactic nuclei, for example, are common nonthermal radio emitters. Other compact Galactic nonthermal radio sources are pulsars (PSRs) and micro-quasars. Interestingly, also in the star formation context nonthermal radio emission may arise from

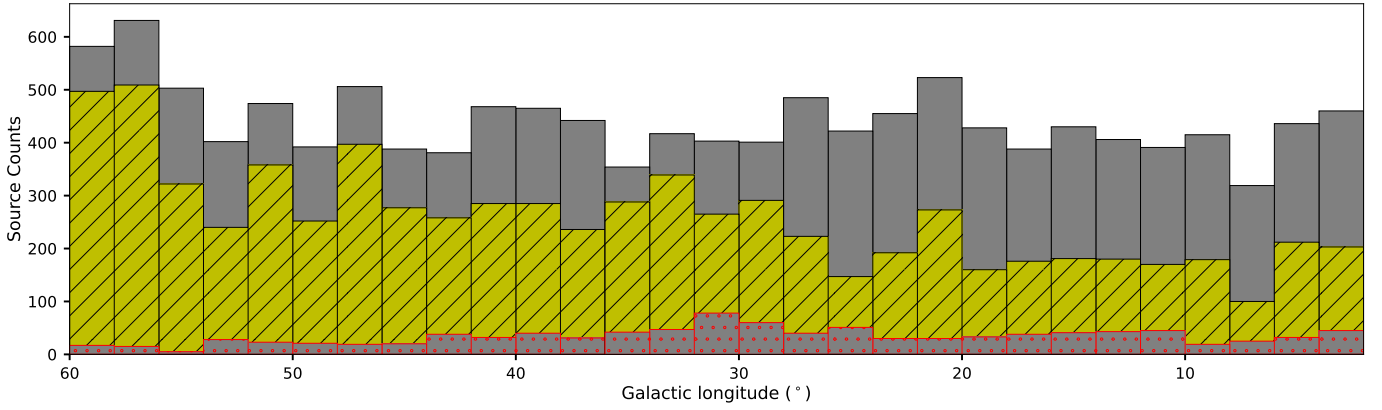


Fig. 10. Source distribution of GLOSTAR D-configuration sources in Galactic longitude. Sources in the range $28^\circ < \ell < 36^\circ$ are taken from Medina et al. (2019). The color of the bars is the same as for Fig. 9. The bin size is 2° .

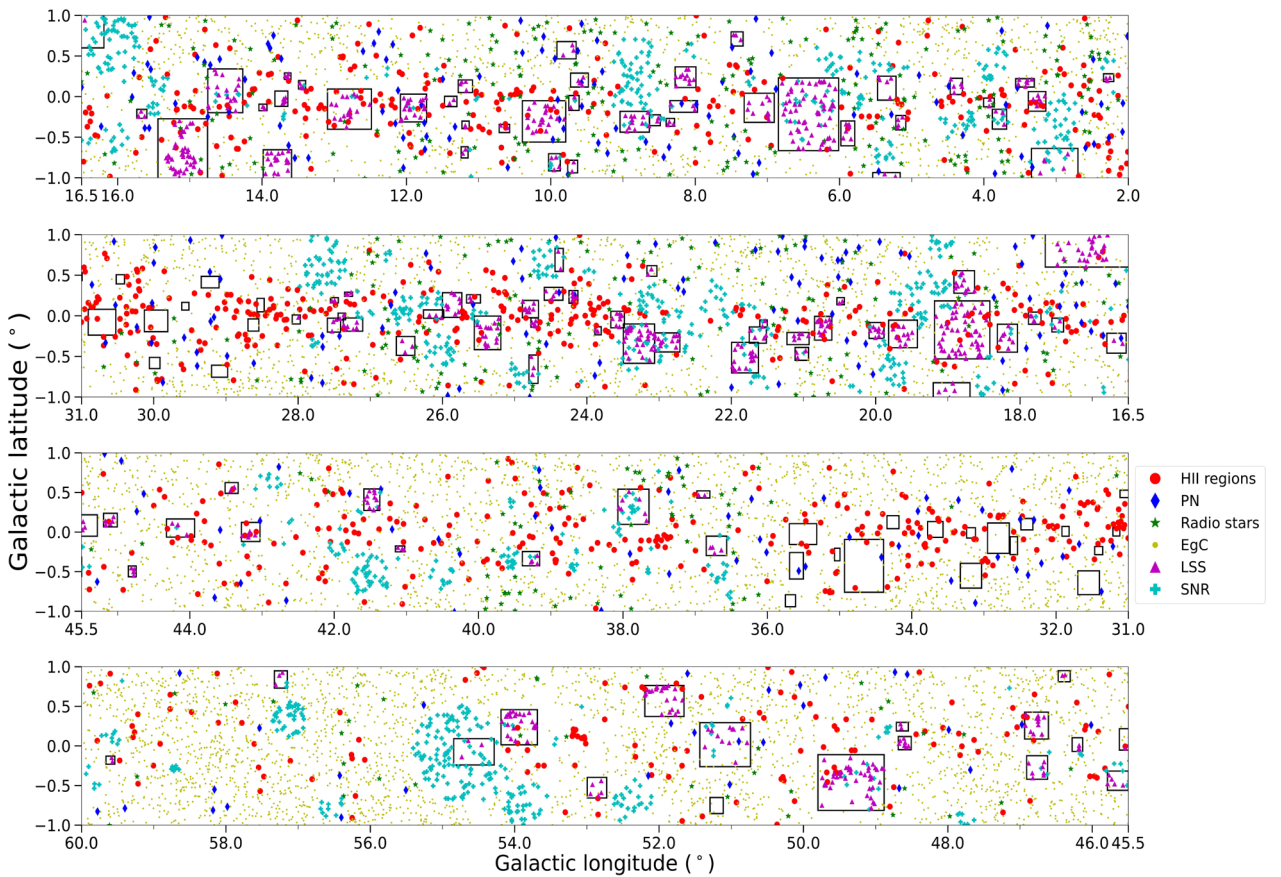


Fig. 11. Spatial distribution of GLOSTAR radio sources reported in this work. Different markers are used for the different resulting source classifications and are labeled on the right side of the plot. Squares delimit the area of the LSSs also identified in this work (see Sect. 3.2). Sources in the range $28^\circ < \ell < 36^\circ$ are taken from Medina et al. (2019).

compact sources, such as strong wind collision regions in massive multiple stellar systems (Dzib et al. 2013; Yanza et al. 2024, MNRAS, submitted), strong shocks of jets of very young massive stars (Reid et al. 1995; Carrasco-González et al. 2010), and magnetically active young low-mass stars (Forbrich et al. 2021; Dzib et al. 2021). Typical spectral index values for non-thermal radio sources related to massive stars are about -0.5 (e.g., Reid et al. 1995; Dougherty et al. 2003; Mohan et al. 2022). However, the spectral index of radio emission from magnetically active stars ranges from -2.0 to $+2.0$ (Dulk 1985; Dzib et al. 2015).

In this work, we have measured the spectral index of over 5000 sources. First, by using the upgraded capabilities of the VLA interferometer that offer a total bandwidth of 4 GHz in C-band, we have produced images in smaller frequency bins and determined in-band spectral indices of compact sources detected with S/N ratio >10.0 . Second, we have used the flux densities from the THOR survey and computed GLOSTAR-THOR spectral indices of compact sources detected in both surveys. Using the flux densities reported by THOR has two advantages. First, it increases the frequency baseline of the spectral index

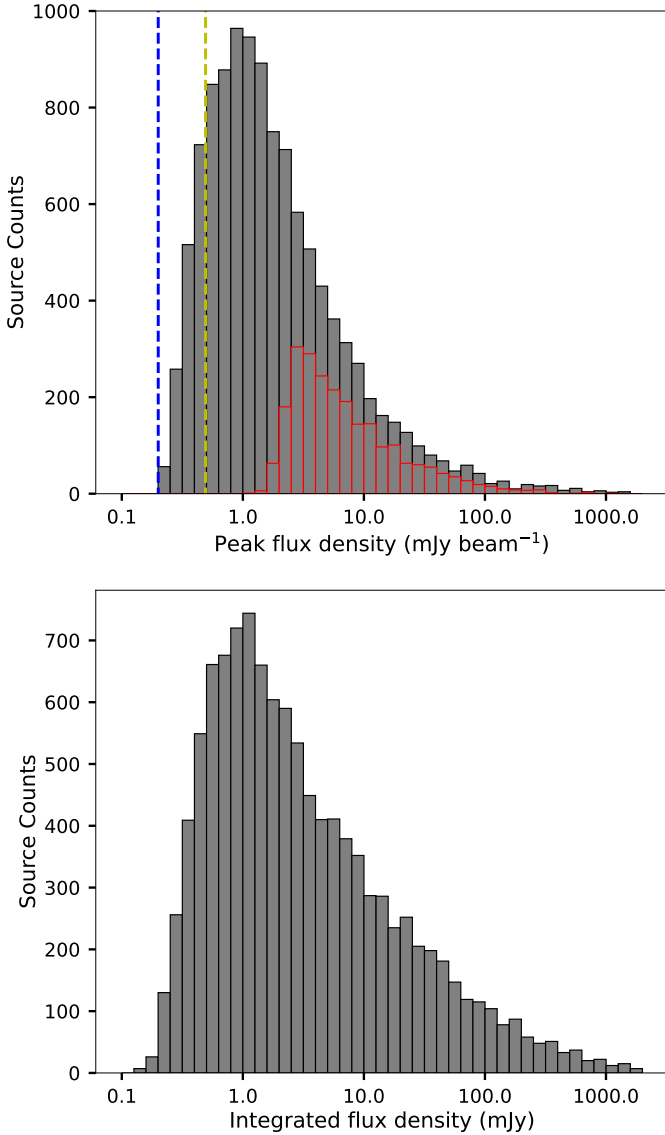


Fig. 12. Peak (top) and integrated (bottom) flux distribution of GLOSTAR sources. The dashed blue and yellow lines in the upper panel delimit the $4\sigma_{\text{noise}}$ detection limit in areas free of extended emission ($\sigma_{\text{noise}} \approx 50 \mu\text{Jy beam}^{-1}$) and for the nominal noise level of the D-configuration images ($\sigma_{\text{noise}} \approx 123 \mu\text{Jy beam}^{-1}$). The red histogram in the upper panel shows the flux distribution of sources previously identified by CORNISH. The bin width is 0.1 dex.

determination. Second, the total flux densities of both surveys can be used without splitting the data into smaller and less sensitive frequency bins. By using the full-band images of both surveys, we optimize the (u, v) -coverage of both surveys. This results in smaller errors in the spectral index determination. Quantitatively, this is reflected by the fact that we obtain mean error values for the in-band and the GLOSTAR-THOR spectral indices of 0.2 and 0.07, respectively. We also notice the recent findings by [Rashid et al. \(2024\)](#), whose results indicate that broadband spectral indices are more reliable than in-band spectral indices. For the 2 819 sources in which both spectral indices could be determined, we plot their values in the top panel of Fig. 15. In the middle panel of Fig. 15 we show the distribution of the differences between these two spectral index determinations, and in the low panel we plot this difference against the GLOSTAR source effective radius. We found a mean difference

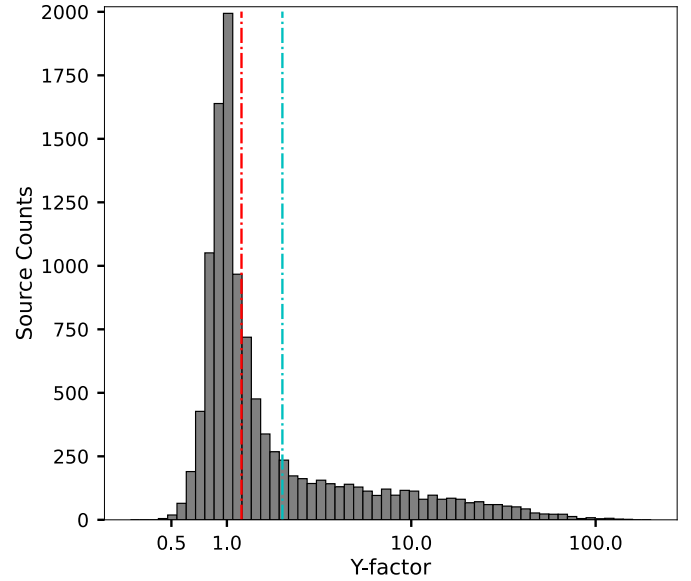


Fig. 13. Distribution of Y-factor of GLOSTAR sources. The dash-dotted red ($Y_{\text{factor}} = 1.2$) and cyan ($Y_{\text{factor}} = 2.0$) lines indicate the criterion used to distinguish between unresolved, compact, and extended sources, respectively. The bin width is 0.05 dex.

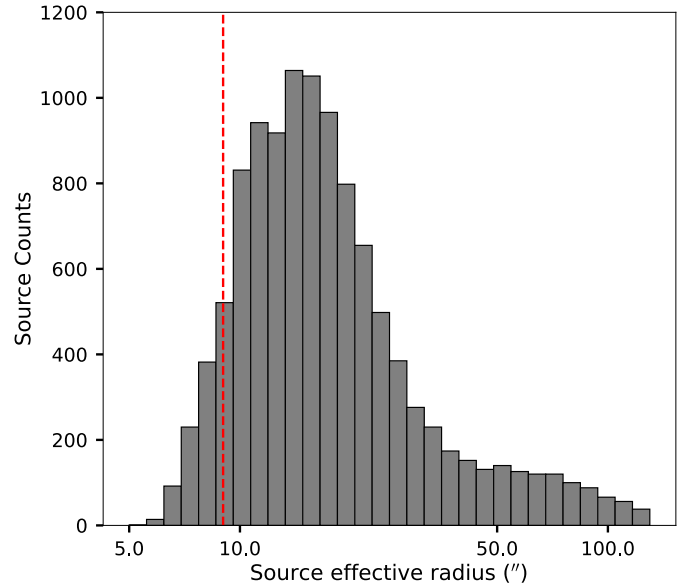


Fig. 14. Distribution of the source radii. The dashed red line indicates the resolution of the observation (radius of the synthesized beam). The bin size used is 0.05 dex.

of -0.08 ± 0.01 and a standard deviation of 0.35. To check if the small negative difference is correlated with size, we ran a Pearson correlation test between the source effective radius and the difference of spectral indices. We obtain a correlation coefficient $r = -0.14$, with a significance value of $p = 6 \times 10^{-14}$. This indicates a weak negative correlation (larger sources have more negative values) and could explain the small negative difference between the two spectral index determinations. While the in-band spectral indices are slightly more negative than the GLOSTAR-THOR spectral index, the independent spectral index determinations are consistent considering the mean values of the errors obtained for both methods. Because the error bars are

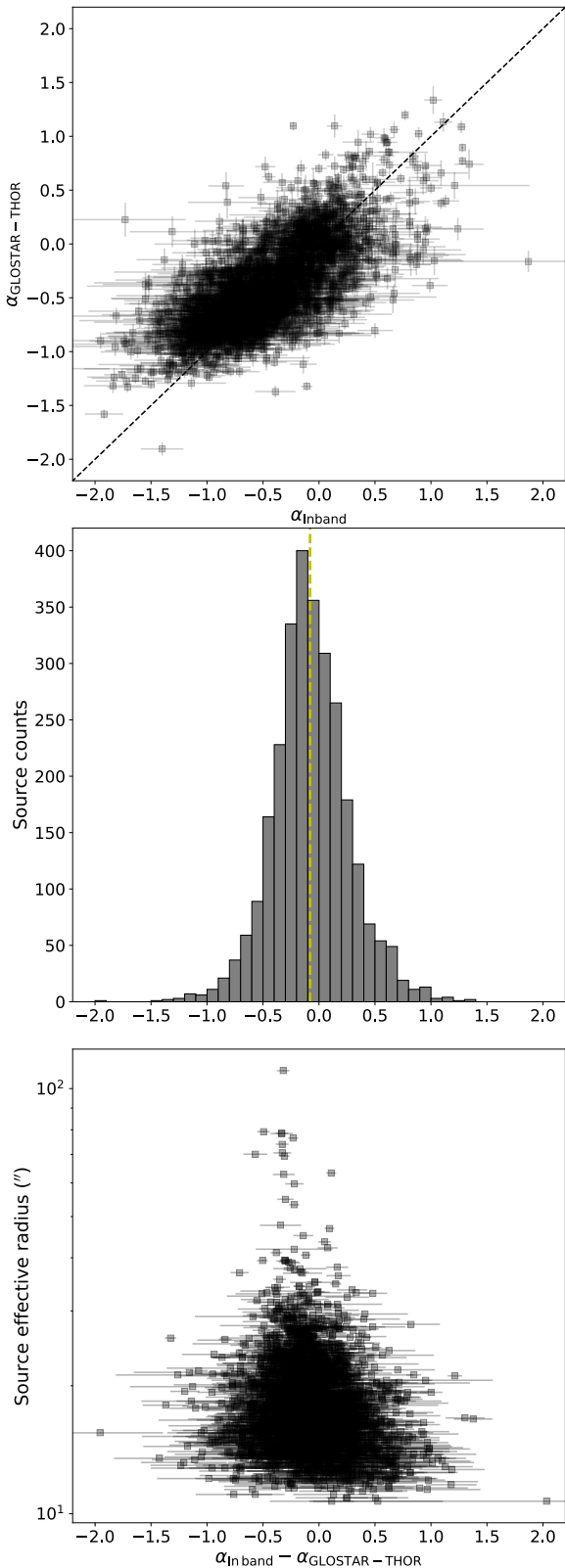


Fig. 15. Comparison of GLOSTAR in-band and GLOSTAR-THOR spectral indices. Top: scatter plot of α_{Inband} vs. $\alpha_{\text{GLOSTAR-THOR}}$. The dashed black line is the equality line. Middle: Distribution of the differences of the spectral indices. The dashed yellow line indicates the mean value of -0.08 ± 0.01 . The standard deviation is 0.35. The histogram used a bin size of 0.1. Bottom: source effective radius vs. $\alpha_{\text{Inband}} - \alpha_{\text{GLOSTAR-THOR}}$.

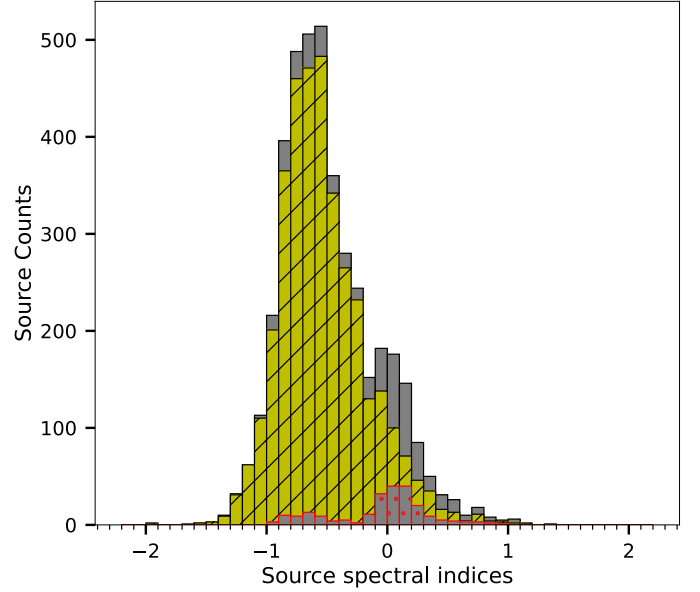


Fig. 16. GLOSTAR-THOR spectral index distribution of radio sources reported in this work. In gray is the full sample, yellow is for sources related to EgC and red for H II region candidates. The histogram used a bin size of 0.1.

smaller, however, a preference is given to the GLOSTAR-THOR spectral indices.

Figure 16 shows the GLOSTAR-THOR spectral index distribution of GLOSTAR sources. The distribution shows two peaks, the first at a value approximately -0.6 and the second at a value ~ 0.0 , pointing to two populations of radio sources. The distribution of EgC sources (yellow histogram) shows only one peak and the mean is approximately -0.6 , similar to the canonical value of -0.7 for extragalactic nonthermal radio sources (Condon 1984). For the 225 H II regions with measured spectral index (red histogram), the distribution also shows a single peak at $\alpha = 0.14 \pm 0.02$, consistent with the values for (almost) optically thin free-free emission (thermal radio emission). We discuss H II regions in more detail in the following section.

5.4. H II regions

Following the classification criteria discussed in Section 4.6, we have identified 769 H II regions, of which 359 are new H II region candidates. Most of them are located close to the mid-plane of the Galactic disk ($b \simeq 0.0$, see Figure 9). Most of the newly identified H II region candidates are compact, with an effective radius smaller than $20''$ (see Fig. 17).

We have estimated the spectral index for 225 H II region candidates, and their mean spectral index value is $+0.14 \pm 0.02$ (Fig. 16). Surprisingly, 55 (24%) of the H II region candidates have a spectral index < -0.2 , which is smaller than -0.1 , the minimum spectral index value expected for optically thin free-free radio emission. In the GLOSTAR higher angular resolution images, we have also found such sources (Dzib et al. 2023; Yang et al. 2023). In those cases, we have speculated that in H II regions associated with radio sources with negative spectral indices, the radio sources do not trace radio emission from the ionized gas of the H II region, which will be resolved out, but rather trace stellar processes producing non-thermal radio emission or a mixture of emission produced by thermal and

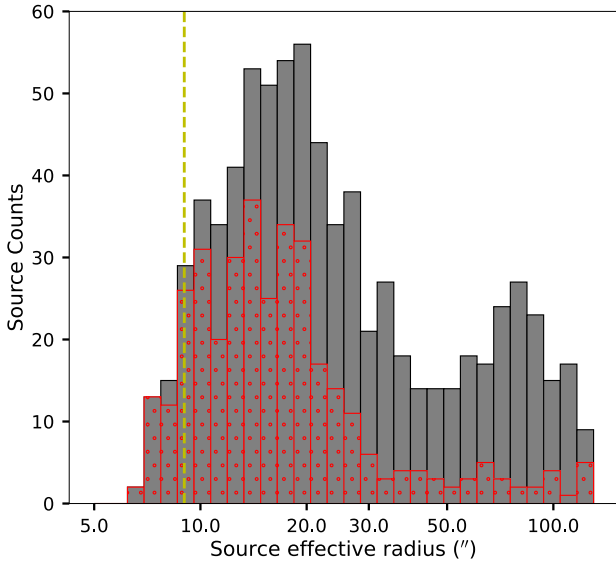


Fig. 17. Distribution of H II region effective radius. The gray bars show the distribution for all identified H II regions. Red bars are the new H II regions candidates identified in this work. The dashed yellow line indicates the resolution of the observation (radius of the synthesized beam). The bin size used is 0.05 dex.

non-thermal processes. Sensitive radio observations have indeed revealed compact non-thermal radio sources in the vicinity of the H II regions, up to several tens in some cases (e.g., Wilner et al. 1999; Medina et al. 2018; Yanza et al. 2022). Wilner et al. (1999) argue that the non-thermal compact radio sources they find around the archetypal ultracompact H II region W3(OH), several of which are time variable, represent low-mass stars in the stellar cluster that surrounds the MYSO that excites the H II region.

In our VLA D-configuration images, we do not expect that H II regions related to compact radio sources ($Y_{\text{factor}} \leq 2.0$) have resolved out emission. This excludes the possibility of imaging artifacts.

In Fig. 18 we plot the 225 spectral indices measured for H II region candidates as a function of their S/N from the GLOSTAR maps. From this plot, it can be seen that most of the sources with negative spectral index also have a low S/N ratio. Most of them also have a low S/N ratio in the THOR survey. The low brightness of these sources could have biased the spectral index determination. However, there are still some sources with a S/N ratio >10 in both GLOSTAR and THOR that clearly have a negative spectral index. These radio sources require further studies to confirm or refute their classification as H II regions.

5.5. Extragalactic background sources

Along with Galactic sources, our observations are expected to detect a large number of extragalactic radio sources. In fact, it has been shown that most of the unclassified sources in our previous catalog can indeed be related to galaxies with radio emission (Medina et al. 2019; Chakraborty et al. 2020). While selection effects could play a role on their distribution, the early studies by Fomalont et al. (1991) show that, at 5 GHz, the rough number of expected background sources per square arcminute, with flux levels above S [μJy], is described by:

$$N(S[\mu\text{Jy}]) = (0.42 \pm 0.05)(S[\mu\text{Jy}]/30)^{-1.18 \pm 0.19}.$$

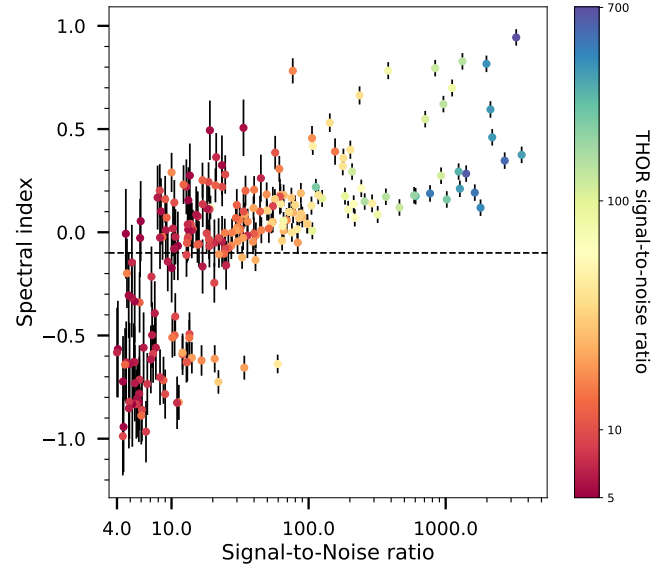


Fig. 18. Spectral indices of H II regions as a function of the S/N. The circles are color-coded with their S/N in the THOR survey. The dashed line shows the spectral index value of -0.1 , the minimum expected for optically thin free-free radio emission.

Considering a mean noise level of $\sigma = 123 \mu\text{Jy}$, and a 4σ threshold for our detected sources, the number of expected extragalactic radio sources in our images is $6\,400 \pm 3\,750$. This number is in excellent agreement with the 6 312 sources that we have classified as EgC.

6. Summary and conclusions

The GLOSTAR-VLA Galactic plane survey (Medina et al. 2019; Brunthaler et al. 2021) is currently the most sensitive mid-radio wavelength survey covering a large fraction of the Galactic plane observed from the northern hemisphere. Its main objective is unveiling signatures of recent massive star formation. However, as many of these sources are expected to be extragalactic background sources, it is necessary to classify the detected sources, a task that has many challenges.

In this work, we have presented radio images of a 100 square degree area of the Galactic plane, with an angular resolution of $18''$. The mapped region covers the area delimited by the coordinates $2^\circ < \ell < 28^\circ$, $36^\circ < \ell < 60^\circ$ and $|b| < 1^\circ$. We have used a combination of the BLOBCAT software and visual inspection procedures to identify 11 211 radio sources in the GLOSTAR map. We have identified 1965 that are part of very extended sources (H II region complexes or SNRs). The radio source catalog presents the results of the source extraction performed with BLOBCAT, such as positions, S/N, flux densities, Y-factor, and effective radius. We have also obtained the spectral indices of 5 276 radio sources, which are also listed in the catalog. We have cross-matched the GLOSTAR radio sources with the radio sources reported from other radio surveys; for example, THOR (Beuther et al. 2016), CORNISH (Hoare et al. 2012), and RMS (Urquhart et al. 2009). These radio surveys were used to, first, verify the measured source parameters and, second, classify sources. Source classification was also performed with the information on (sub)millimeter and IR wavelength counterparts. Source classes are also listed in the catalog, in which a large number (6312) are extragalactic background sources.

With the performed multiwavelength analysis, we identified 769 H II region candidates. Previous works have reported 410 of these as H II regions or H II region candidates, and the remaining 359 we identify as H II candidates for the first time. The spatial distribution of these sources is concentrated around the Galactic mid-plane and their numbers decrease in the outer parts of the Galactic disk ($\ell > 56^\circ$), indicating zones with higher and lower star formation, respectively. Using additional flux density measurements at 1.4 GHz from the THOR survey, we have also determined spectral indices of 225 H II region candidates. Their mean spectral index is ~ 0.1 , consistent with thermal free-free radio emission. Interestingly we found several H II region candidates with negative spectral index, although this could be partly due to their low S/N ratio in GLOSTAR and THOR. However, there is an interesting sample of H II region candidates with $S/N > 10$ and negative spectral indices that need to be studied further to establish their true nature.

Combining the information from large surveys is an excellent way to obtain an unbiased look for tracers of early star formation. The derived properties and classification of several thousands of new and known radio sources are invaluable information and truly show the legacy nature of the GLOSTAR radio survey.

Data availability

The full version of Table 6, and Fig. 1 are available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](ftp://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/689/A196> or via <https://glostar.mpifr-bonn.mpg.de>

Acknowledgements. This research was partially funded by the ERC Advanced Investigator Grant GLOSTAR (247078). S.A.D. acknowledge the M2FINDERS project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant No 101018682). AYY acknowledges support from the NSFC grants No. 11988101 and No. NSFC 11973013. M.R.R. is a Jansky Fellow of the National Radio Astronomy Observatory. This work uses information from the GLOSTAR databases at <http://glostar.mpifr-bonn.mpg.de> supported by the MPIfR (Max-Planck-Institut für Radioastronomie), Bonn, which is based on observations with the Karl G. Jansky Very Large Array (VLA) of NRAO (The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.) and 100-m telescope of the MPIfR at Effelsberg. It also made use of information from the ATLASGAL database at http://atlasgal.mpifr-bonn.mpg.de/cgi-bin/ATLASGAL_DATABASE.cgi supported by the MPIfR, Bonn, as well as information from the CORNISH database at <http://cornish.leeds.ac.uk/public/index.php> which was constructed with support from the Science and Technology Facilities Council of the UK. This work has used data from GLIMPSE survey of the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This publication also makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This paper used the data products from the Hi-GAL survey of the Herschel telescope which is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. This document was prepared using the collaborative tool Overleaf available at: <https://www.overleaf.com/>.

References

Anderson, L. D., Zavagno, A., Barlow, M. J., García-Lario, P., & Noriega-Crespo, A. 2012, *A&A*, 537, A1
 Anderson, L. D., Bania, T. M., Balsler, D. S., et al. 2014, *ApJS*, 212, 1
 Anderson, L. D., Wang, Y., Bihr, S., et al. 2017, *A&A*, 605, A58
 Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
 Beuther, H., Bihr, S., Rugel, M., et al. 2016, *A&A*, 595, A32
 Brunthaler, A., Menten, K. M., Dzib, S. A., et al. 2021, *A&A*, 651, A85

Carrasco-González, C., Rodríguez, L. F., Anglada, G., et al. 2010, *Science*, 330, 1209
 Chakraborty, A., Roy, N., Wang, Y., et al. 2020, *MNRAS*, 492, 2236
 Chini, R., & Hoffmeister, V. 2008, in *Handbook of Star Forming Regions*, ed. B. Reipurth (Belgium: ASP Books), 5, 625
 Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
 Condon, J. J. 1984, *ApJ*, 287, 461
 Contreras, Y., Schuller, F., Urquhart, J. S., et al. 2013, *A&A*, 549, A45
 Cotton, W. D. 2008, *PASP*, 120, 439
 Csengeri, T., Urquhart, J. S., Schuller, F., et al. 2014, *A&A*, 565, A75
 Dokara, R., Brunthaler, A., Menten, K. M., et al. 2021, *A&A*, 651, A86
 Dokara, R., Gong, Y., Reich, W., et al. 2023, *A&A*, 671, A145
 Dougherty, S. M., Pittard, J. M., Kasian, L., et al. 2003, *A&A*, 409, 217
 Dulk, G. A. 1985, *ARA&A*, 23, 169
 Dzib, S. A., Rodríguez, L. F., Loinard, L., et al. 2013, *ApJ*, 763, 139
 Dzib, S. A., Loinard, L., Rodríguez, L. F., et al. 2015, *ApJ*, 801, 91
 Dzib, S. A., Forbrich, J., Reid, M. J., & Menten, K. M. 2021, *ApJ*, 906, 24
 Dzib, S. A., Yang, A. Y., Urquhart, J. S., et al. 2023, *A&A*, 670, A9
 Elia, D., Molinari, S., Schisano, E., et al. 2017, *MNRAS*, 471, 100
 Elia, D., Merello, M., Molinari, S., et al. 2021, *MNRAS*, 504, 2742
 Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *ApJS*, 154, 10
 Fomalont, E. B., Windhorst, R. A., Kristian, J. A., & Kellerman, K. I. 1991, *AJ*, 102, 1258
 Forbrich, J., Dzib, S. A., Reid, M. J., & Menten, K. M. 2021, *ApJ*, 906, 23
 Ghosh, S. K., Iyengar, K. V. K., Rengarajan, T. N., et al. 1989, *ApJ*, 347, 338
 Ginsburg, A., Goddi, C., Kruijssen, J. M. D., et al. 2017, *ApJ*, 842, 92
 Gong, Y., Ortiz-León, G. N., Rugel, M. R., et al. 2023, *A&A*, 678, A130
 Green, D. A. 2019, *J. Astrophys. Astron.*, 40, 36
 Green, J. A., Caswell, J. L., Fuller, G. A., et al. 2009, *MNRAS*, 392, 783
 Greisen, E. W. 2003, *Astrophys. Space Sci. Lib.*, 285, 109
 Hales, C. A., Murphy, T., Curran, J. R., et al. 2012, *MNRAS*, 425, 979
 Helfand, D. J., Becker, R. H., White, R. L., Fallon, A., & Tuttle, S. 2006, *AJ*, 131, 2525
 Hoare, M. G., Lumsden, S. L., Oudmaijer, R. D., et al. 2005, *IAU Symp.*, 227, 370
 Hoare, M. G., Purcell, C. R., Churchwell, E. B., et al. 2012, *PASP*, 124, 939
 Holwerda, B. W. 2005, arXiv e-prints [arXiv:astro-ph/0512139]
 Irabor, T., Hoare, M. G., Oudmaijer, R. D., et al. 2018, *MNRAS*, 480, 2423
 Kerton, C. R., Arvidsson, K., & Alexander, M. J. 2013, *AJ*, 145, 78
 Khan, S., Rugel, M. R., Brunthaler, A., et al. 2024, *A&A*, 689, A81 (Paper X)
 Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, *MNRAS*, 391, 136
 Lumsden, S. L., Hoare, M. G., Urquhart, J. S., et al. 2013, *ApJS*, 208, 11
 Marleau, F. R., Noriega-Crespo, A., Paladini, R., et al. 2008, *AJ*, 136, 662
 Medina, S. N. X., Dzib, S. A., Tapia, M., Rodríguez, L. F., & Loinard, L. 2018, *A&A*, 610, A27
 Medina, S. N. X., Urquhart, J. S., Dzib, S. A., et al. 2019, *A&A*, 627, A175
 Messineo, M., Davies, B., Figer, D. F., et al. 2011, *ApJ*, 733, 41
 Mohan, S., Vig, S., & Mandal, S. 2022, *MNRAS*, 514, 3709
 Molinari, S., Swinyard, B., Bally, J., et al. 2010, *A&A*, 518, L100
 Molinari, S., Schisano, E., Elia, D., et al. 2016, *A&A*, 591, A149
 Motte, F., Bontemps, S., & Louvet, F. 2018, *ARA&A*, 56, 41
 Nguyen, H., Rugel, M. R., Menten, K. M., et al. 2021, *A&A*, 651, A88
 Nguyen, H., Rugel, M. R., Murugesan, C., et al. 2022, *A&A*, 666, A59
 Oliveira, J. M. 2008, in *Handbook of Star Forming Regions*, Volume II, ed. B. Reipurth (Belgium: ASP Books), 5, 599
 Ortiz-León, G. N., Menten, K. M., Brunthaler, A., et al. 2021, *A&A*, 651, A87
 Purcell, C. R., Hoare, M. G., Cotton, W. D., et al. 2013, *ApJS*, 205, 1
 Rashid, M., Roy, N., Pandian, J. D., et al. 2024, arXiv e-prints [arXiv:2405.18978]
 Reid, M. J., Argon, A. L., Masson, C. R., Menten, K. M., & Moran, J. M. 1995, *ApJ*, 443, 238
 Schuller, F., Menten, K. M., Contreras, Y., Wyrowski, F., & Schilke, E. 2009, *A&A*, 504, 415
 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., Thompson, M. A., Moore, T. J. T., et al. 2013, *MNRAS*, 435, 400
 Urquhart, J. S., Csengeri, T., Wyrowski, F., et al. 2014, *A&A*, 568, A41
 Urquhart, J. S., König, C., Giannetti, A., et al. 2018, *MNRAS*, 473, 1059
 Urquhart, J. S., Wells, M. R. A., Pillai, T., et al. 2022, *MNRAS*, 510, 3389
 Wang, Y., Bihr, S., Rugel, M., et al. 2018, *A&A*, 619, A124
 Wang, Y., Beuther, H., Rugel, M. R., et al. 2020, *A&A*, 634, A83
 Wilner, D. J., Reid, M. J., & Menten, K. M. 1999, *ApJ*, 513, 775
 Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
 Yang, A. Y., Urquhart, J. S., Thompson, M. A., et al. 2021, *A&A*, 645, A110
 Yang, A. Y., Dzib, S. A., Urquhart, J. S., et al. 2023, *A&A*, 680, A92
 Yanza, V., Masqué, J. M., Dzib, S. A., et al. 2022, *AJ*, 163, 276
 Yanza, V., Dzib, S. A., Palau, A., et al. 2024, *MNRAS*, submitted
 Zhang, B., Reid, M. J., Zhang, L., et al. 2019, *AJ*, 157, 200