

The HIPASS Catalogue: I - Data Presentation

M. J. Meyer,^{1,2*} M. A. Zwaan,^{1,3} R. L. Webster,¹ L. Staveley-Smith,⁴
E. Ryan-Weber,^{1,4} M. J. Drinkwater,⁵ D. G. Barnes,^{1,6} M. Howlett,⁶
V. A. Kilborn,⁷ J. Stevens,¹ M. Waugh,¹ M. J. Pierce,^{6,8} R. Bhathal,⁸
W. J. G. de Blok,⁹ M. J. Disney,⁹ R. D. Ekers,⁴ K. C. Freeman,¹⁰
D. A. Garcia,⁹ B. K. Gibson,⁶ J. Harnett,¹¹ P. A. Henning,¹²
H. Jerjen,¹⁰ M. J. Kesteven,⁴ P. M. Knezek,¹³ B. S. Koribalski,⁴
S. Mader,⁴ M. Marquarding,⁴ R. F. Minchin,⁹ J. O’Brien,¹⁰
T. Oosterloo,¹⁴ R. M. Price,¹² M. E. Putman,¹⁵ S. D. Ryder,¹⁶
E. M. Sadler,¹⁷ I. M. Stewart,¹⁸ F. Stootman,⁸ and A. E. Wright⁴

¹ *School of Physics, University of Melbourne, VIC 3010, Australia*

² *Space Telescope Science Institute, 3700 San Martin Drive, Baltimore MD 21218, U.S.A.*

³ *European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching b. München, Germany*

⁴ *Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, NSW 1710, Australia*

⁵ *Department of Physics, University of Queensland, QLD 4072, Australia*

⁶ *Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122 Australia.*

⁷ *Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire, SK11 9DL, U.K.*

⁸ *Department of Physics, University of Western Sydney Macarthur, P.O. Box 555, Campbelltown, NSW 2560, Australia*

⁹ *Department of Physics & Astronomy, University of Wales, Cardiff, P.O. Box 913, Cardiff CF2 3YB, U.K.*

¹⁰ *Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia*

¹¹ *University of Technology Sydney, Broadway NSW 2007, Australia*

¹² *Institute for Astrophysics, University of New Mexico, 800 Yale Blvd, NE, Albuquerque, NM 87131, USA.*

¹³ *WIYN, Inc. 950 North Cherry Avenue, Tucson, AZ, U.S.A.*

¹⁴ *ASTRON, P.O. Box 2, 7990 AA Dwingeloo, The Netherlands*

¹⁵ *CASA, University of Colorado, Boulder, CO 80309-0389, U.S.A.*

¹⁶ *Anglo-Australian Observatory, P.O. Box 296, Epping, NSW 1710, Australia*

¹⁷ *School of Physics, University of Sydney, NSW 2006, Australia*

¹⁸ *Department of Physics & Astronomy, University of Leicester, Leicester LE1 7RH, U.K.*

ABSTRACT

The HI Parkes All-Sky Survey (HIPASS) Catalogue forms the largest uniform catalogue of HI sources compiled to date, with 4,315 sources identified purely by their HI content. The catalogue data comprise the southern region $\delta < +2^\circ$ of HIPASS, the first blind HI survey to cover the entire southern sky. RMS noise for this survey is 13 mJy beam^{-1} and the velocity range is $-1,280$ to $12,700 \text{ km s}^{-1}$. Data search, verification and parametrization methods are discussed along with a description of measured quantities. Full catalogue data are made available to the astronomical community including positions, velocities, velocity widths, integrated fluxes and peak flux densities. Also available are on-sky moment maps, position-velocity moment maps and spectra of catalogue sources. A number of local large-scale features are observed in the space distribution of sources including the Super-Galactic plane and the Local Void. Notably, large-scale structure is seen at low Galactic latitudes, a region normally obscured at optical wavelengths.

Key words: methods: observational – surveys – catalogues – radio lines: galaxies

1 INTRODUCTION

The neutral hydrogen (HI) content of galaxies provides a unique and fundamental perspective into the nature and evolution of galaxies, being intimately linked to both the dark matter component of galaxies along with their ability to form stars.

Surveys for HI in galaxies, however, are limited by the relative weakness of HI photons. Despite a cosmological mass density ratio of HI to luminous matter of only $1:10$,¹ the improbability of the HI hyperfine transition and weakness of photons emitted mean that the received power ratio favours optical photons by many orders of magnitude. In combination with observational limitations, this has traditionally necessitated either blind surveys of relatively small volumes (eg. Shostak 1977, Krumm & Brosch 1984, Kerr & Henning 1987, Henning 1992, Sorar 1994, Zwaan et al. 1997, Spitzak & Schneider 1998, Rosenberg & Schneider 2000), or targeted surveys across larger regions through the use of existing, typically optical, catalogues (eg. Fisher & Tully 1981, Mathewson et al. 1992, Giovanelli

* E-mail: martinm@stsci.edu (MM), mzwaan@eso.org (MZ), rwebster@ph.unimelb.edu.au (RW), lister.staveley-smith@csiro.au (LSS)

¹ $\Omega_{\text{HI}} = (3.8 \pm 0.6) \times 10^{-4} h_{75}^{-1}$ (Zwaan et al. 2003) cf. $\Omega_{\text{S}} = (39 \pm 5.7) \times 10^{-4} h_{75}^{-1}$ (Salpeter initial mass function, Cole et al. 2001)

et al. 1997, Haynes et al. 1999). Parameters for existing blind HI surveys are summarized in Table 1.

With the advent of Multibeam receivers on the Parkes and Jodrell Bank radio telescopes, along with the forthcoming ALFA instrument at Arecibo, large untargeted surveys are now practicable for the first time. In this work we discuss results from the HI Parkes All-Sky Survey (HIPASS), a blind survey of the entire sky $\delta < +25^\circ$. Various extragalactic catalogue subsets from this survey have already been completed including: Kilborn et al. (2002, $\delta < -62$); Koribalski et al. (2004, $\delta < 0^\circ$ and $S_p > 116$ mJy) and companion analysis papers Zwaan et al. (2003, HI mass function) and Ryan-Weber et al. (2002, newly catalogued galaxies); Waugh et al. (2002, Fornax region); and Banks et al. (1999, Centaurus A group). Presented here are results from the southern region of the survey ($\delta < +2^\circ$); these data were searched for extragalactic sources to full survey depth.

A complete database of HI-selected galaxies is important for a number of important astrophysical studies. Amongst these is a measurement of the HI cosmic mass density, Ω_{HI} . At higher redshifts, Ω_{HI} can be measured through damped Ly α systems, however at $z=0$, the only way to accurately measure Ω_{HI} is through large-scale surveys for HI emission. Galaxies under-represented in optical surveys such as low surface brightness galaxies, those with elusive optical counterparts and Malin 1 type systems (Bothun et al. 1987) are also best detected via blind HI surveys. Gaining an accurate census of such objects is important for constraining models of galaxy formation, although not contributing a significant fraction of the overall gas mass density (Zwaan et al. 2003).

Other important applications for a large uniform catalogue of HI sources include the measurement of large-scale structure, 2-point correlation functions, the Tully-Fisher relation, the role of hydrogen gas in the evolution of galaxy groups and the study of environments around HI-rich galaxies compared to those of optically-selected galaxies.

In Section 2 we discuss the observational methods, data processing and basic properties of HIPASS. Methods used to compile the source database are described in detail in Section 3, including the identification, verification and parametrization of sources, confused- and extended-source processing, the effects of radio-frequency interference (RFI) and recombination lines, and follow-up observations. Access to catalogue data is described in Section 4. Finally, basic property distributions are presented in Section 5 and the large-scale distribution of galaxies in Section 6. Reliability, completeness and accuracy of catalogue parameters are discussed in the companion paper by Zwaan et al. (2004, Paper II). The identification

of optical counterparts to HIPASS Catalogue galaxies is made in Drinkwater et al. (2004, Paper III).

2 THE HI PARKES ALL-SKY SURVEY

HIPASS was completed at the Parkes² 64m radio telescope using the Multibeam receiver, an hexagonal array of 13 circular feed-horns installed at the focal plane of the telescope. Observations and data reduction for this survey are described in detail in Barnes et al. (2001) and Staveley-Smith et al. (1996). A brief summary is provided below.

2.1 Observations

HIPASS observations were carried out from February 1997 to March 2000 for the southern portion of the survey ($\delta < +2^\circ$) and to December 2001 for the northern extension ($\delta < +25^\circ$). For these observations, the telescope was scanned across the sky in 8° strips at a rate of 1° per minute, with scans separated by 7 arcmin in RA, giving a total effective integration time of 450s beam^{-1} . To obtain spectra, the Multibeam correlator was used with a 64 MHz bandwidth and 1024 channel configuration, giving a velocity range of $-1,280$ to $12,700\text{ km s}^{-1}$ and a channel separation of 13.2 km s^{-1} at $z = 0$. Parameters for the survey are summarized in Table 2.

2.2 Data Processing

Initial processing of HIPASS data was completed in real-time at the observatory. To correct spectra for the standard bandpass effects, spectra were reduced using the package LIVEDATA. This package also performed the conversion to heliocentric rest frame velocities by shifting spectra using Fourier techniques.

Bandpass correction is done by dividing the signal target spectrum by a reference off-source spectrum, this spectrum representing the underlying spectral shape caused by signal filtering, as well as the temperature of the receivers, ground and sky. For the HIPASS data, the reference spectrum for each receiver is obtained by taking signals immediately before and after a particular integration as the telescope scans across the sky. To provide a robust measure of the bandpass, the value for each channel in the reference spectrum is taken to

² The Parkes telescope is part of the Australia Telescope which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO

be the median rather than the mean of available signals. However, one artifact that can be created by this process is the generation of negative side lobes north and south of particularly bright HI sources, since actual HI emission is included in the bandpass determination. Finally, spectra are smoothed with a Tukey filter to suppress Gibbs ringing, resulting in a velocity resolution of 18 km s^{-1} .

To create the three-dimensional position-position-velocity cubes that form the main data product of HIPASS, the bandpass-corrected spectra are gridded together using the package GRIDZILLA. The spectrum at each RA-Dec pixel is taken as the median of all data within 6 arcmin of the pixel centre, multiplied by 1.28 to correctly scale point source flux densities. This procedure thus corrupts data for extended sources in the standard HIPASS data, with peak flux densities³ for such sources over-estimated (by 1.28 for an *infinitely* extended source). Extended sources nevertheless represent only a small fraction of overall sources ($\sim 2\%$, see Section 3.5) and the corresponding corrections are small. We therefore apply no correction to measured catalogue parameters for this effect. The resulting cubes from the gridding process are $8^\circ \times 8^\circ$ with $4' \times 4'$ pixels in the on-sky directions and extend over the full HIPASS velocity range in the third axis (an example cube is shown in Figure 1). Overlap regions between cubes vary, but in general are $\sim 1^\circ$ in RA and Dec. In total, 388 cubes are required to cover the southern sky.

Lastly, a correction is applied to the spectra in each HIPASS data cube to minimize baseline distortion caused by the Sun and other continuum sources. This is applied using the program LUTHER and involves the fitting and subtraction of a template distortion spectrum.

3 THE CATALOGUE

The HIPASS Catalogue (HICAT) is compiled using a combination of automatic and interactive processes, with candidate detections first generated through automated finder scripts and then manually verified. Detections are finally parametrized using semi-automated routines. These are discussed below; a summary of the number of detections remaining at each stage is given in Table 3.

³ ‘peak flux density’ is hereafter referred to as ‘peak flux’

3.1 Candidate Generation

To generate candidate detections, the results of two separate scripts are combined. The first of these, a modified version of `MULTIFIND` (Kilborn 2001), uses a peak flux threshold method, while the second, `TOPHAT`, cross-correlates spectra with tophat profiles of various scales. The two methods are used together as neither is able to recover fully the real sources detected by the other (`MULTIFIND` locates $\sim 83\%$ of final catalogue sources, while `TOPHAT` finds $\sim 90\%$). Both candidate lists are then merged and automatically cleaned to remove high velocity clouds (HVCs) and suspect detections at known interference frequencies as detailed below.

3.1.1 `MULTIFIND`

The first of the finder algorithms is based on `MIRIAD` (Sault et al. 1995) routines. Each of the velocity planes in a data cube is searched for detections rising above a specified peak flux density and a Gaussian fitted to each pixel clump. For `HICAT`, the peak flux threshold is set to four times the calculated noise. Cube noise is calculated robustly using the relation:

$$\text{cube noise} = s \times \sqrt{\pi/2}, \quad (1)$$

where s is the median absolute deviation from the median. A median noise estimate is used to provide a method insensitive to outliers, with the applied conversion equivalent to estimating RMS of the normally distributed component of cube noise. Detections are filtered to include only those detections present in adjacent velocity planes with a separation < 5 arcmin.

This detection process is repeated twice, each time Hanning smoothing the data cube to increase detection signal-to-noise, and cube noise re-measured. Finally, detection lists are combined and duplicated detections removed.

3.1.2 `TOPHAT`

In the second finder algorithm, each spectrum in the cubes is searched for emission on a variety of velocity width scales between 1 and 40 channels. An initial list of detections is created as follows. First, solar and continuum ripple in the spectral baselines is reduced by using a moving median filter with a width dependent on the current search scale. Spectra are then cross-correlated with a top hat of the appropriate size, with the convolution weighted

by the noise in each velocity plane. A feature is detected in a convolved spectrum if it rises above a threshold proportional to the interquartile range values in that spectrum.

In the second stage of candidate identification, detections from all scales are island grouped, listing together all detections that are either adjacent or overlapping (the majority of features will be detected on multiple scales). The bounding box of these groups is then found, and the original data re-fitted to create the entry in the final detection list for each group.

3.1.3 Cleaning

Results from the automated finders are then merged and cleaned. First, all candidates with $v_{\text{gsr}} < 300 \text{ km s}^{-1}$ are removed to avoid Galactic HI emission and HVCs (HVCs in HIPASS are catalogued separately, see Putman et al. 2002). Additionally, an initial filter is applied to remove false detections arising from hydrogen recombination lines and radio frequency interference. All detections having velocity widths $< 25.0 \text{ km s}^{-1}$ in the ranges $2,638 < v_{\text{hel}} < 2,670 \text{ km s}^{-1}$, $4,355 < v_{\text{hel}} < 4,397 \text{ km s}^{-1}$, $4,505 < v_{\text{hel}} < 4,559 \text{ km s}^{-1}$ and $7,428 < v_{\text{hel}} < 7,482 \text{ km s}^{-1}$ are removed. We note that many more recombination line and radio frequency interference detections are discarded in subsequent verification procedures than by this simple and conservative filter (see Sections 3.6 & 3.7 for further discussion). As no sources in the final catalogue have a velocity width $< 25 \text{ km s}^{-1}$, these cuts are not expected to have any effect on the final catalogue completeness. A declination limit of $\delta < +2^\circ$ is also applied. Following this process, $\sim 140,000$ candidates remain to be manually checked.

3.2 Candidate Verification

Verification of candidate sources is a multi-stage process. First, two independent checks are made on each candidate, visually inspecting the full HIPASS spectrum summed from a 12×12 arcmin box around the position of the detection. The aim of this process is to quickly remove obviously spurious detections without rejecting any real sources.

Next, as a preliminary pass at removing multiple detections of the same object, candidates are grouped using an angular separation limit of 5 arcmin and a velocity separation of 100 km s^{-1} . In this process a detection is added to an existing group if it is within the separation limits of any of its members. One detection from each group is then selected

for further examination, giving preference first to those not rejected in both of first checks, then to the detection with the highest peak flux and finally, in instances of equal peak flux, to the detection with the greatest velocity width. Group size warning limits are set to 10 arcmin and 200 km s^{-1} . Groups with sizes exceeding these warning limits are later inspected to ensure no nearby galaxies are removed (see Section 3.4).

In the third stage of detection verification, candidates not rejected by both of the first checks are examined in spectral, position, and position-velocity space. Third checks are carried out by one of only three selected observers to improve consistency. Positions are fitted using automatic iterative fitting routines based on original detection coordinates. To generate spectra, candidates are treated as unresolved, with pixels in a 28×28 arcmin box used to calculate source flux in a given velocity plane via:

$$S(v) = \sum_i \frac{f_i S_i}{f_i^2} \quad (2)$$

where S_i is the flux at a given pixel (and velocity v), and f_i is the value of the Gaussian beam at that pixel, relative to the previously fitted centre of emission.

While every effort is made to either reject or confirm detections, any detections unable to be evaluated either way are left in the catalogue at this stage. Spurious detections are flagged in the parametrization stage for follow-up observations.

3.3 Parametrization

The parametrization of HICAT sources is done interactively using standard MIRIAD routines, together with a modified version of MBSPECT that uses Gaussian smoothing for baseline fitting. All parametrization is done by M.M. and M.Z.

3.3.1 Baseline Fitting

Baseline fitting is carried out by first linearly interpolating the spectrum over the galaxy profile, as specified with the parameters v_{lo} and v_{hi} (see Table 4). The average values of the ten channels on either side of the masked profile are taken as the endpoints for the interpolation. The resulting spectrum is then convolved with a Gaussian of an interactively specified width. A Gaussian is used to provide the smoothest estimate of the spectral baseline. It is important in choosing the width of the Gaussian kernel to ensure that the scale is large enough to fit smoothly beneath the galaxy profile without creating spurious small

scale variations. The Gaussian also needs to be wide enough not to reduce artificially the noise of the observed spectrum. At the same time, the Gaussian should be narrow enough to fit the baseline without missing the peaks and troughs of the spectrum. Example spectra comparing Gaussian smoothing baseline fitting to polynomial baseline fitting are shown in Figure 2.

3.3.2 Parameter Measurement

To parametrize a detection, the position of the candidate is first refined by fitting an elliptical Gaussian to a 28×28 arcmin moment map around galaxy finder coordinates. Provided fitting is successful, the coordinates of the detection are taken as those of the ellipse and the 28×28 arcmin moment map regenerated. Entering an interactive loop for parameter measurement, an ellipse is re-fitted to the new detection location, and the measurement region again moved to the ellipse position if the fit is successful. Plots of the detection are now generated in RA-Dec, Dec-velocity and spectral spaces, along with initial parameter measurements (see Figure 3). Spectra are Hanning smoothed for parameter measurement to improve signal-to-noise, giving a final velocity resolution of 26.4 km s^{-1} . Spectrum and profile limits, masked regions, initial fitting position and final position can all be varied using the interactive windows at this point. Checks can be made on the veracity of the detection by changing grey-scales, box size, Hanning smoothing and the velocity range displayed. The size of the Gaussian kernel used in baseline fitting can also be adjusted. For each adjustment made, images and moment maps are regenerated, where appropriate, with fitting redone accordingly. When a satisfactory fit is achieved, the detection is given a final comment to confirm or reject it, with a follow-up flag available if necessary. By default, all fluxes are measured using the point source method, treating sources as non-extended with the central high signal-to-noise pixels preferentially weighted according to Parkes' beam parameters. The treatment of extended sources is discussed in Section 3.5. A full list of parameters measured is given in Table 4, along with an example spectrum illustrating various measured quantities in Figure 4.

3.4 Confused Source Identification and Duplicate Removal

To combine the lists from individual cubes, the full set of results is again island-grouped and cleaned to remove duplicate detections in the cube overlap regions. Accurate positions

enable much tighter matching limits, now chosen at 1.5 arcmin and 20 km s⁻¹. Tight limits also avoid accidentally removing real source pairs.

As the scales of duplicate detections and confused source pairs overlap at this point, all galaxies with velocity profiles overlapping and on-sky position within 6σ of another galaxy (as defined by the gridded Parkes beamwidth) are examined manually. Also included in this process are any detections previously flagged as confused during parametrization, as well as galaxies originating from cleaned groups that exceeded size warning limits in either of the island grouping stages. Confused sources flagged in the final catalogue are either those manually identified at this stage, or detections with velocity profiles overlapping and source separations $< 4\sigma$.

Sources that could reasonably be distinguished (ie. measured fluxes largely due to the galaxy itself rather than a close neighbour), are identified and measured separately, whereas those for which this was not possible are grouped together into a single source and the detection manually flagged as confused.

3.5 Extended Sources

The minimum expected number of extended sources in HICAT can be roughly estimated using the relation between HI mass and HI diameter found for an optically selected sample of galaxies (Broeils & Rhee 1997):

$$\log(M_{HI}) = (1.96 \pm 0.04) \log(D_{HI}) + (6.52 \pm 0.06), \quad (3)$$

where M_{HI} is HI mass in M_\odot and D_{HI} is the HI diameter defined at a surface density of $1 M_\odot pc^{-2}$ in kpc. Combining this with the relation between HI mass (in solar masses), distance (in Mpc) and observed flux $\int S dV$ ($= S_{int}$, in Jy km s⁻¹):

$$M_{HI} = 2.356 \times 10^5 D^2 \int S dV \quad (4)$$

and making the approximation $1.96 \log(D) \approx \log(D^2)$, this gives a relation for source size that depends solely on source flux and is independent of distance (θ_{HI} is the source diameter in arcminutes):

$$\int S dV \approx 1.2 \theta_{HI}^2 \quad (5)$$

The relationship between $\int SdV$ and diameter can be empirically confirmed by examining the difference between summed integrated flux in a 44×44 arcmin box and the standard non-extended source integrated flux in a 28×28 arcmin box (pixels weighted by the Parkes beam parameters). This difference is plotted as a function of the summed 44×44 arcmin integrated flux in Figure 5. As expected, the average difference is close to zero at small integrated fluxes (point sources) and positive for large integrated fluxes (extended sources), a positive value indicating that some flux was missed through use of the point-source method. The large amount of scatter in this graph is indicative of the uncertainties in measuring integrated fluxes by summing the value of all pixels in a given region, compared to the point source method.

To identify extended sources in HICAT, all sources greater than $7'$ in size are taken as potentially extended, this limit corresponding to 57 Jy km s^{-1} from Equation 5. Assuming a uniform galaxy flux distribution, 93% of source flux is retrieved for a $7'$ diameter source if treated as unresolved, and higher for a centrally biased distribution. In total, there are 188 candidate sources brighter than this flux limit, where flux is measured using an 44×44 arcmin ‘summed’ box. Each source is examined manually, re-fitted and flagged if necessary. Where extended sources are confused or only marginally extended, detections remain fitted as point sources if this best removes erroneous flux contributions from neighbouring galaxies or cube noise. In the final catalogue, 90 sources are flagged as extended.

3.6 Interference

To minimize the potential effects of interference, scans in HIPASS were taken in 5 separate sets, each 35 arcminutes apart, at well separated times. Nevertheless, some narrow-band interference remains in the data, the most prominent of which is a line at 1408 MHz (corresponding to $\sim 2640 \text{ km s}^{-1}$) caused by the 11th harmonic of the 128 MHz multibeam correlator sampler clock (Barnes et al. 2001), and another line at 1400 MHz (corresponding to $\sim 4370 \text{ km s}^{-1}$) again caused by local interference. The presence of these, along with a number of other narrow interference lines, is clearly evident in the distribution of HICAT candidates following the first two checks of the cataloguing process (top panel, Figure 6). Also apparent in the velocity distribution are detections associated with the GPS L3 beacon at 1381 MHz ($\sim 8550 \text{ km s}^{-1}$). These detections cover a range of velocities due to intrinsic signal spreading of the GPS beacon. The presence of these interference detections following

the first two verification checks is consistent with the purely spectral nature of these checks and their deliberately conservative approach to candidate rejection. However, the examination of position-velocity maps in the third stage of checking enables the removal of these false detections as shown by the final velocity distribution of the catalogue in Figure 9. The time varying nature of the GPS signal, combined with the HIPASS observation and data processing methods, results in a signal that is dominated by only a few on-sky pixels, much narrower than the signal of an unresolved source (see Figure 7). Similarly, narrow-line RFI can be distinguished by the narrow-line signature spread across large spatial regions (Figure 8).

The lack of features in the final velocity distribution at RFI frequencies shows that HICAT completeness and reliability are not significantly degraded at these locations. This point is discussed further in Section 3.5 of Paper II.

3.7 Recombination lines

Another source of potentially erroneous HI detections in HICAT is the recombination of warm ionized gas in the Milky Way. The full list of hydrogen $\Delta n=1$ to $\Delta n=5$ recombination lines appearing in the HIPASS velocity range is given in Table 5 (taken from Lilley & Palmer 1968) and plotted against the distribution of detections following the first two verification checks in Figure 6). Given the increased line strengths of $\Delta n=1$ transitions compared to those with higher principal quantum number differences, the erroneous detection spikes in Figure 6 most likely due to hydrogen recombination are H167a and H210b. These also coincide with known RFI frequencies. Once again, these lines are successfully removed in the subsequent verification stages as demonstrated by the final source velocity distribution.

3.8 Narrow-Band Follow-up

As detailed in Paper II, an extensive program of Parkes narrow-band follow-up observations has been carried out to test the reliability of HICAT. To additionally remove as many false detections as possible, low peak flux sources and those flagged as uncertain were preferentially targetted. In total, 1082 HICAT source were confirmed in this process and 119 removed. Detections confirmed by narrow-band observations have been flagged as such in the final catalogue.

4 DATA

The full data of HICAT is available on-line at:

<http://hipass.aus-vo.org>

This database is searchable in a number of ways, including by position and velocity. Returned parameters can be individually selected, along with any of the image products, including detection spectra, on-sky moment maps and position-velocity moment maps. The format of the returned catalogue data can also be chosen, with both HTML and plain text available. An extract of the HIPASS Catalogue is given in Table 6.

We encourage other researchers to make use of this database. For optimum utility, researchers also need to be aware of the completeness, reliability and accuracy of the measured parameters. These are described in detail in Paper II. Users are also encouraged to be familiar with the full processing of HIPASS data (Barnes et al. 2001).

5 BASIC PROPERTY DISTRIBUTIONS

To show the basic properties of the HIPASS galaxies, Figure 9 plots the bivariate distributions of velocity, velocity width, peak flux and integrated flux. Plotted along the diagonal are the one-dimensional histograms of each quantity.

The large-scale structure of the local universe is clearly visible in the velocity distribution, with two prominent over-densities appearing at $\sim 1,600$ and $\sim 2,800$ km s^{-1} . This diagram also shows the noise-limited nature of the survey, with no sharp cutoff in galaxy numbers observed at the highest observed velocities. From the peak flux distribution, it is clear that most galaxies are close to the detection limit of the survey, with the distribution peaking at ~ 50 mJy.

The sharp decline at the upper end of the galaxy mass function can be seen by examining the peak flux-velocity and integrated flux-velocity distributions. This effective upper limit to HI mass results in galaxies at high distances all having low peak and integrated fluxes. The steepness of the decline is apparent in the relatively sharp boundary between the flux-velocity regions in which galaxies are, and are not, observed. The low space density of the most massive galaxies is also visible in the width-velocity distribution, with small-velocity-width galaxies preferentially found nearby due to the small volumes surveyed.

6 LARGE SCALE STRUCTURE

Figures 10 and 11 plot the equatorial coordinate polar and Aitoff projections of the full on-sky distribution of HICAT sources respectively. Points are shaded according to the observed recessional velocities. From these plots, the inhomogeneous distribution of HI sources in the local universe is clearly apparent. Most obvious are the Super-Galactic plane (marked with a dark solid line), the Fornax wall (comprising the over-densities around Dorado, Fornax and Eridanus; passes through RA=3.5 hrs, Dec= -30°), and a third filament extending from near the Super-Galactic plane, through Antlia, Puppis and Lepus (passes through RA=8 hrs, Dec= -30°). Also notable are significantly under-dense regions, the most prominent of which is the Local Void (around RA=18 hrs, Dec= -30°). The location of large-scale structures is important when determining the representative nature of galaxy surveys given their search region.

HICAT also maps structures behind the Milky Way, with relatively little obscuration apparent in the observed distribution. Many of these structures and filaments have been previously identified (Kraan-Korteweg et al. 2000; Henning et al. 1998), however their uninterrupted path through the Zone of Avoidance is traced out here for the first time in a single survey.

In Figures 12 and 13, the velocity distribution of HICAT galaxies is plotted in two different projection sets. In Figure 12, the southern sky hemisphere is divided like segments of an orange into six wedges, each with a common base axis along the line RA=0-12 in the equatorial plane. The location of these wedges with respect to the southern celestial hemisphere is shown in Figure 14. The shape of these regions results in volume increasing in the vertical rather than radial direction. Thus, more galaxies are present in the central horizontal regions when compared to similar velocity regions further out. Figure 13 also divides the southern hemisphere into six wedges, this time using a common base axis along RA=6-18 in the equatorial plane.

These figures again show the Super-Galactic plane and the Local Void. Figure 12(d) contains the bulk of the Super-Galactic plane in face-on projection and Figures 13(c)-(e) show the plane in edge-on cross sections. From this second set of diagrams it is also clear that local regions of the plane are in fact comprised of two main parallel components. The lower right regions of these diagrams show the significantly under-dense region of the Local Void. Figure 12(f) most clearly shows this in the orthogonal projection in the centre of

the diagram. Also marked in the velocity diagrams are the locations of four major nearby clusters (F=Fornax, C=Centaurus, H=Hydra and N=Norma), their locations coinciding with over-dense regions in the HICAT data.

7 SUMMARY

HIPASS is the first blind HI survey of the entire southern sky. Using the Multibeam receiver on the Parkes Telescope, this survey covers $\delta < +25^\circ$ with a velocity range $-1,280$ to $12,700 \text{ km s}^{-1}$. Mean noise for the survey is 13 mJy beam^{-1} . Taking data from the southern region of this survey $\delta < +2^\circ$, we have compiled a database of 4,315 sources selected purely on their HI content. Candidates are found using automatic finder algorithms, before being subjected to an extensive process of manual checking and verification. Issues examined in compilation of the catalogue include RFI and recombination lines, along with additional processes for confused and extended sources. A large program of follow-up observations is also used to confirm catalogue sources. Completeness, reliability and parameter accuracy of HICAT sources is addressed in Paper II. Optical counterparts to HIPASS galaxies are examined in Paper III.

The large-scale distribution of sources in the catalogue covers a number of major local structures, including the Super-Galactic plane along with filaments passing through the Fornax and Puppis regions. Also notable is the probing of structure in the Zone of Avoidance, a region significantly obscured in optical bands. The survey also surveys a number of under-dense local regions including the Local Void.

This database offers a unique opportunity to study a number of important astrophysical issues. Future papers will address topics such the cosmic mass density of cool baryons, the properties of HI selected galaxies at other wavelengths, the comparative clustering of HI-rich galaxies, the Tully-Fisher relation, the role of hydrogen in galaxy group evolution and the study of HI-rich galaxy environments.

8 ACKNOWLEDGMENTS

The Multibeam system was funded by the Australia Telescope National Facility (ATNF) and an Australian Research Council grant. The collaborating institutions are the Universities of Melbourne, Western Sydney, Sydney and Cardiff, Research School of Astronomy and Astrophysics at Australian National University, Jodrell Bank Observatory and the ATNF.

The Multibeam receiver and correlator was designed and built by the ATNF with assistance from the Australian Commonwealth Scientific and Industrial Research Organisation Division of Telecommunications and Industrial Physics. The low noise amplifiers were provided by Jodrell Bank Observatory through a grant from the UK Particle Physics and Astronomy Research Council. The Multibeam Survey Working Group is acknowledged for its role in planning and executing the HIPASS project. This work makes use of the AIPS++, MIRIAD and KARMA software packages. We would also like to acknowledge the assistance of Richard Gooch, Mark Calabretta, Warwick Wilson and the ATNF engineering development group. Finally, we thank Brett Beeson and the Australian Virtual Observatory for development of the SKYCAT service which hosts the HICAT data.

REFERENCES

- Banks G. D., et al., 1999, *ApJ*, 524, 612
- Barnes D. G., et al., 2001, *MNRAS*, 322, 486
- Bothun G. D., Impey C. D., Malin D. F., Mould J. R., 1987, *AJ*, 94, 23
- Braun R., Thilker D., Walterbos R. A. M., 2003, *AAP*, 406, 829
- Broeils A. H., Rhee M.-H., 1997, *A&A*, 324, 877
- Cole S., et al., 2001, *MNRAS*, 326, 255
- de Vaucouleurs G., de Vaucouleurs A., Corwin H. G., Buta R. J., Paturel G., Fouque P., 1991, *Third Reference Catalogue of Bright Galaxies*, Springer-Verlag: Berlin Heidelberg New York
- Drinkwater M. J., et al., 2004, Paper III, in prep.
- Fisher J. R., Tully R. B., 1981, *ApJS*, 47, 139
- Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, *ApJ*, 473, 576
- Giovanelli R., Avera E., Karachentsev I. D., 1997, *AJ*, 114, 122
- Haynes M. P., Giovanelli R., Chamaraux P., da Costa L. N., Freudling W., Salzer J. J., Wegner G., 1999, *AJ*, 117, 2039
- Henning P. A., 1992, *ApJS*, 78, 365
- Henning P. A., Kraan-Korteweg R. C., Rivers A. J., Loan A. J., Lahav O., Burton W. B., 1998, *AJ*, 115, 584
- Henning P. A., et al. 2000, *AJ*, 119, 2686

- Karachentsev I. D., Makarov D. A., 1996, *AJ*, 111, 794
- Kerr F. J., Henning P. A., 1987, *ApJ*, 320, L99
- Kilborn V. A., 2001, PhD thesis, The University of Melbourne
- Kilborn V. A., et al., 2002, *AJ*, 124, 690
- Koribalski B., et al, 2004, submitted
- Kraan-Korteweg R. C., Henning P. A., Andernach H., eds, 2000, Mapping the Hidden Universe: The Universe behind the Milky Way - The Universe in HI
- Krumm N., Brosch N., 1984, *AJ*, 89, 1461
- Lang R. H., et al., 2003, *MNRAS*, 342, 738
- Lilley A. E., Palmer P., 1968, *ApJS*, 16, 143
- Mathewson D. S., Ford V. L., Buchhorn M., 1992, *ApJS*, 81, 413
- Putman M. E., et al., 2002, *AJ*, 123, 873
- Rosenberg J. L., Schneider S. E., 2000, *ApJS*, 130, 177
- Ryan-Weber E., et al., 2002, *AJ*, 124, 1954
- Sault R. J., Teuben P. J., Wright M. C. H., 1995, in *ASP Conf. Ser. 77: Astronomical Data Analysis Software and Systems IV.*, 433
- Shostak G. S., 1977, *A&A*, 54, 919
- Sorar E., 1994, PhD thesis, Pittsburgh University
- Spitzak J. G., Schneider S. E., 1998, *ApJS*, 119, 159
- Staveley-Smith L., et al., 1996, *PASA*, 13, 243
- Waugh M., et al., 2002, *MNRAS*, 337, 641
- Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, *ApJ*, 490, 173
- Zwaan M. A., et al., 2004, *MNRAS*, Paper II, submitted
- Zwaan M. A., et al., 2003, *AJ*, 125, 2842

Table 1. Parameters of existing blind HI surveys

Name	Area (deg ²)	Beam Size (arcmin)	Velocity (km s ⁻¹)	δv^* (km s ⁻¹)	RMS (mJy beam ⁻¹)	$N_{HI}^{lim\dagger}$ (cm ⁻²)	Telescope	Sources
Shostak ¹	85, 70, 11	10.8	~ -800 to $\sim 2,835$	11	32, 44, 18	1.6×10^{19} , 2.3×10^{19} , 9.2×10^{18}	91m Green Bank	1 [‡]
Krumm & Brosch ²	~ 35 , ~ 44	10.8	6,300 to 9,600, 5,300 to 8,500	~ 45	22	2.2×10^{19}	91m Green Bank	0
Henning ³	7,204 pointings	10.8	-400 to $7,500$	22	-	-	91m Green Bank	37
AHISS ⁴	65	3.3	$< 7,400$	16	0.75	4.3×10^{18}	305m Arecibo	66
Slice Survey ⁵	55	3.3	100 to 8,340	16	1.7	9.6×10^{18}	305m Arecibo	75
ADBS ⁶	430	3.3	$< 7,980$	34	3.5	2.8×10^{19}	305m Arecibo	265
HIZSS ⁷	1,840	15.5	$-1,200$ to $12,700$	27	15	5.0×10^{18}	64m Parkes	110
SCC ^{8§}	~ 2400	15.5	$-1,280$ to $12,700$	18	13	3.5×10^{18}	64m Parkes	536
WRST Wide Field Survey ⁹	1,800	49	$-1,000$ to $6,500$	17	18	4.8×10^{17}	Westerbork Array	155
HIJASS ^{10¶}	1,115	12	$-3,500$ to $10,000$	18	16	7.3×10^{18}	76m Jodrell Bank	222
BGC ^{11§}	21,346	15.5	$-1,200$ to $8,000$	18	13	3.5×10^{18}	64m Parkes	1,000
HICAT ^{12§}	22,091	15.5	300 to $12,700$	18	13	3.5×10^{18}	64m Parkes	4,315

*Velocity resolution

[†]Limiting HI column density for gas filling the beam. Limits are 5σ and calculated with $\Delta V = 100 \text{ km s}^{-1}$, scaling RMS appropriately according to the velocity resolution.[‡]Not certainly extragalactic.[§]SCC, BGC and HICAT all use HIPASS data. The SCC sample searches $\delta < -62^\circ$ and the BGC searches $\delta < 0^\circ$ and $S_p > 116 \text{ mJy}$.[¶]Survey in progress.^{||}Data velocity resolution. Parametrization was carried out with additional Hanning smoothing, giving a velocity resolution of 26.4 km s^{-1} .¹Shostak (1977), ²Krumm & Brosch (1984), ³Kerr & Henning (1987), Henning (1992), ⁴Sorar (1994), Zwaan et al. (1997),⁵Spitzak & Schneider (1998), ⁶Rosenberg & Schneider (2000), ⁷Henning et al. (2000), ⁸Kilborn et al. (2002),⁹Braun, Thilker & Walterbos (2003), ¹⁰Lang et al. (2003), ¹¹Koribalski et al. (2004), ¹²this paper.**Table 2.** HIPASS parameters

Sky Coverage	$\delta < +25^\circ$ *
Integration time per beam	450s
Average FWHM	14.3 arcmin
Gridded FWHM	15.5 arcmin
Pixel size	4 arcmin
Velocity Range	$-1,280$ to $12,700 \text{ km s}^{-1}$
Channel separation [†]	13.2 km s^{-1}
Velocity Resolution	18.0 km s^{-1}
Positional accuracy [‡]	1.5 arcmin
RMS noise	13 mJy beam^{-1}

* The HIPASS Catalogue covers southern regions $\delta < +2^\circ$ [†] at $z = 0$ [‡] 1σ accuracy at 99% completeness flux limit
(Zwaan et al. 2004)**Table 3.** Detection numbers at different stages of catalogue compilation

Stage	Number
Generated by MULTIFIND	137,060
Generated by TOPHAT	17,232
Examined in first two verification stages	142,276
Examined in third verification stage	61,276
Final catalogue	4,315

Table 4. HIPASS parameter descriptions

Parameter	Database Name	Units	Description ¹
Name	hipass_name	—	names are of the form HIPASS JXXXX±YY[a-z], where XXXX is the unrounded source RA in hrs and min, and YY is the unrounded source declination in degrees. An additional letter a-z is added if needed to distinguish sources. Names are also made consistent with existing published sources where necessary.
RA	ra	hrs	right ascension (J2000 hexadecimal format)
Dec	dec	deg	declination (J2000 hexadecimal format)
v_{50}^{\max}	vel_50max	km s^{-1}	average of velocities at which profile reaches 50 per cent of peak flux density (using width maximization procedure, ² recommended velocity)
v_{50}^{\min}	vel_50min	km s^{-1}	average of velocities at which profile reaches 50 per cent of peak flux density (using width minimization procedure ²)
v_{20}^{\max}	vel_20max	km s^{-1}	average of velocities at which profile reaches 20 per cent of peak flux density (using width maximization procedure ²)
v_{20}^{\min}	vel_20min	km s^{-1}	average of velocities at which profile reaches 20 per cent of peak flux density (using width minimization procedure ²)
v_{mom}	vel_mom	km s^{-1}	flux weighted velocity average between v_{lo} and v_{hi}
v_{Sp}	vel_Sp	km s^{-1}	velocity at which profile peak flux occurs
v_{gsr}	vel_gsr	km s^{-1}	Galactic standard of rest velocity (converted from heliocentric v_{mom} using de Vaucouleurs et al. 1991)
v_{lg}	vel_lg	km s^{-1}	Local Group standard of rest velocity (converted from heliocentric v_{mom} using Karachentsev & Makarov 1996)
v_{cmb}	vel_cmb	km s^{-1}	Cosmic Microwave Background standard of rest velocity (converted from heliocentric v_{mom} using Fixsen et al. 1996)
v_{lo}	vel_lo	km s^{-1}	manually specified minimum profile velocity
v_{hi}	vel_hi	km s^{-1}	manually specified maximum profile velocity
v_{speclo}	vel_speclo	km s^{-1}	manually specified minimum velocity for spectral plots and RMS measurement
v_{spechi}	vel_spechi	km s^{-1}	manually specified maximum velocity for spectral plots and RMS measurement
v_{mask}	vel_mask	km s^{-1}	manually specified mask regions over which to interpolate baseline fitting and to exclude values from RMS measurement (in velocity pairs)
W_{50}^{\max}	width_50max	km s^{-1}	difference of velocities at which profile reaches 50 per cent of peak flux density (using width maximization procedure ²)
W_{50}^{\min}	width_50min	km s^{-1}	difference of velocities at which profile reaches 50 per cent of peak flux density (using width minimization procedure ²)
W_{20}^{\max}	width_20max	km s^{-1}	difference of velocities at which profile reaches 20 per cent of peak flux density (using width maximization procedure ²)
W_{20}^{\min}	width_20min	km s^{-1}	difference of velocities at which profile reaches 20 per cent of peak flux density (using width minimization procedure ²)
S_{p}	Sp	Jy	peak flux density of profile
S_{int}	Sint	Jy km s^{-1}	integrated flux of source (within region v_{lo} , v_{hi} and box size)
RMS	rms	Jy	RMS of masked spectrum between v_{speclo} and v_{spechi}
RMS_{clip}	rms_clip	Jy	clipped RMS of masked spectrum between v_{speclo} and v_{spechi}
RMS_{cube}	rms_cube	Jy	unsmoothed RMS of cube (median absolute deviation from median multiplied by $\sqrt{\pi/2}$)
cube	cube	—	cube number
σ	sigma	km s^{-1}	standard deviation of Gaussian used in baseline smoothing
box size	boxsize	arcmin	box size used for parameter measurements
comment	comment	—	comment (1=real, 2=have concerns)
follow-up	nb_flg	—	narrow band follow-up status flag (1=real)
confused	cfstd_flg	—	confused source flag (1=confused)
extended	ext_flg	—	extended source flag (1=extended). For extended sources S_{int} is summed flux within region rather than default point-source weighting.

¹ Unless otherwise stated, all velocities are in the heliocentric restframe and cz convention² Maximization procedure starts at profile velocity limits and searches inward until the required per cent of peak flux is reached. The minimization procedure starts at v_{Sp} and searches outward.

Transition	Apparent H I Velocity (km s^{-1})
H166a	-911
H209b	36
H239c	438
H263d	1434
H283e	1540
H240c	4199
H210b	4340
H167a	4507
H284e	4718
H264d	4857
H285e	7918
H241c	7991
H265d	8306
H211b	8685
H168a	9990
H286e	11140
H266d	11781
H242c	11815

Table 5. Hydrogen $\Delta n=1$ to $\Delta n=5$ recombination lines in the range of HIPASS observations (Lilley & Palmer 1968) and their apparent H I velocities.

Landscape table to go here.

Table 6. Extract of the HIPASS Catalogue. The catalogue is presented in its entirety in the electronic version of this journal. Parameter descriptions are given in Table 4.

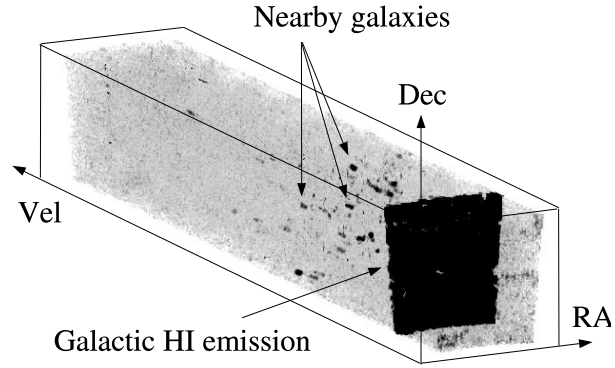


Figure 1. Orthogonal projection of HIPASS cube 93 showing emission from the Milky Way as well as other nearby galaxies. Velocity range is truncated at $10,000 \text{ km s}^{-1}$.

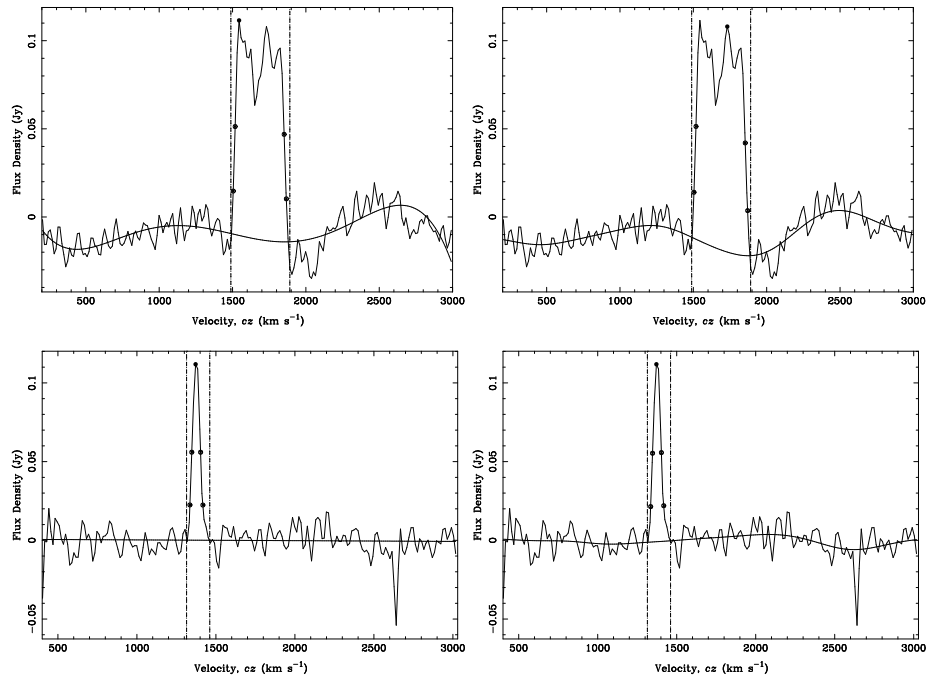


Figure 2. Examples of baseline fitting using polynomials (left) and Gaussian smoothing (right). Upper polynomial fit uses a 5th order polynomial and lower fit uses 1st order polynomial.

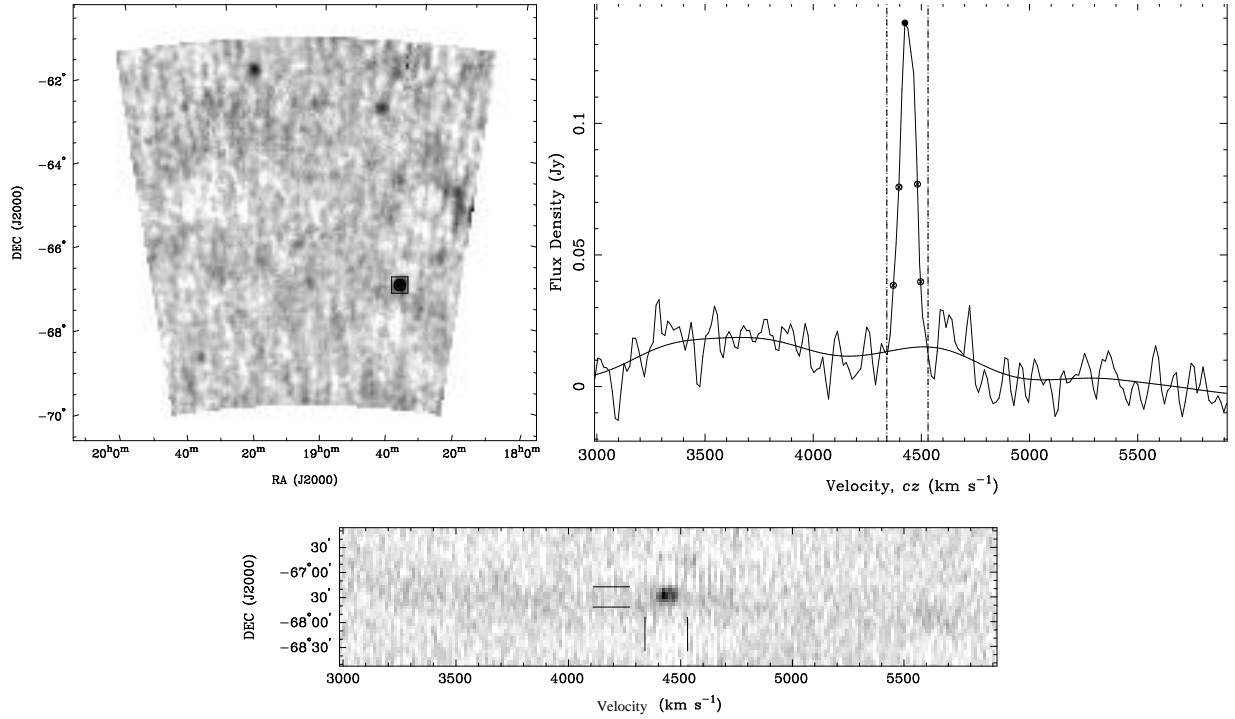


Figure 3. Example plots examined during the final checking and parametrization processes: (top-left) RA-Dec moment map showing the default 28×28 arcmin data box used for position fitting; (top-right) spectrum within the 28×28 arcmin box, treating the detection as a point source and weighting pixels by Parkes' beam parameters - vertical lines show manually specified velocity limits within which profile parameters are measured and the RA-Dec moment map generated; (bottom) Dec-velocity moment map showing 28 arcmin box and profile velocity limits.

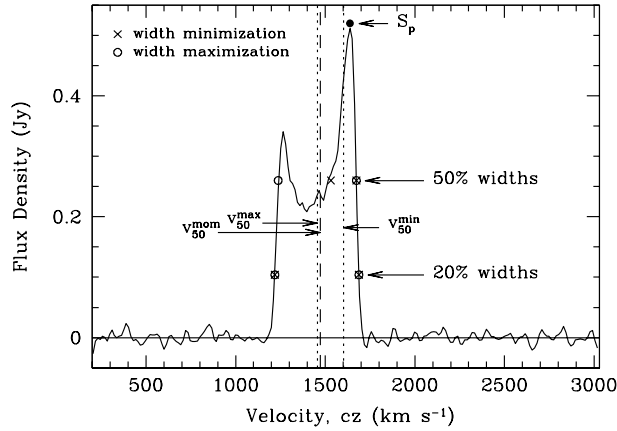


Figure 4. Example spectrum following baseline fitting and removal showing the locations of S_p , W_{50}^{\max} , W_{50}^{\min} , W_{20}^{\max} , W_{20}^{\min} , v_{50}^{\max} , v_{50}^{\min} , v_{mom} , v_{lo} and v_{hi} . The parameters v_{20}^{\max} and v_{20}^{\min} are not shown but are defined similarly to v_{50}^{\max} and v_{50}^{\min} . The offset of v_{mom} toward higher velocities from v_{50}^{\max} is due to the asymmetric nature of profile.

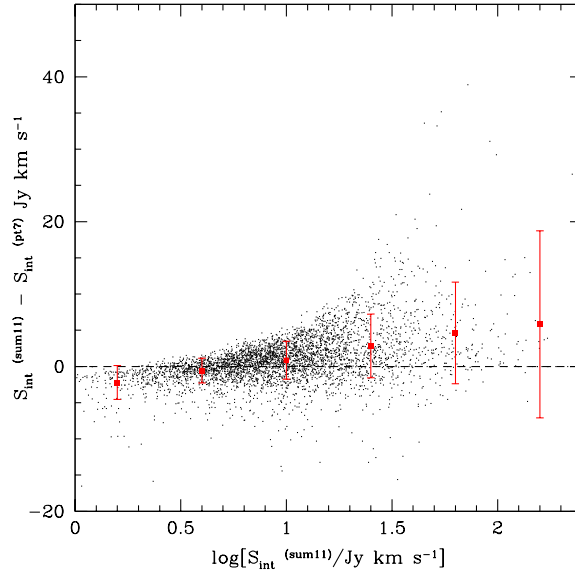


Figure 5. Difference between summed flux in 44×44 arcmin box and flux in 28×28 arcmin box weighted by Parkes beam parameters assuming a point source. Values increase toward higher integrated fluxes as a result of extended sources.

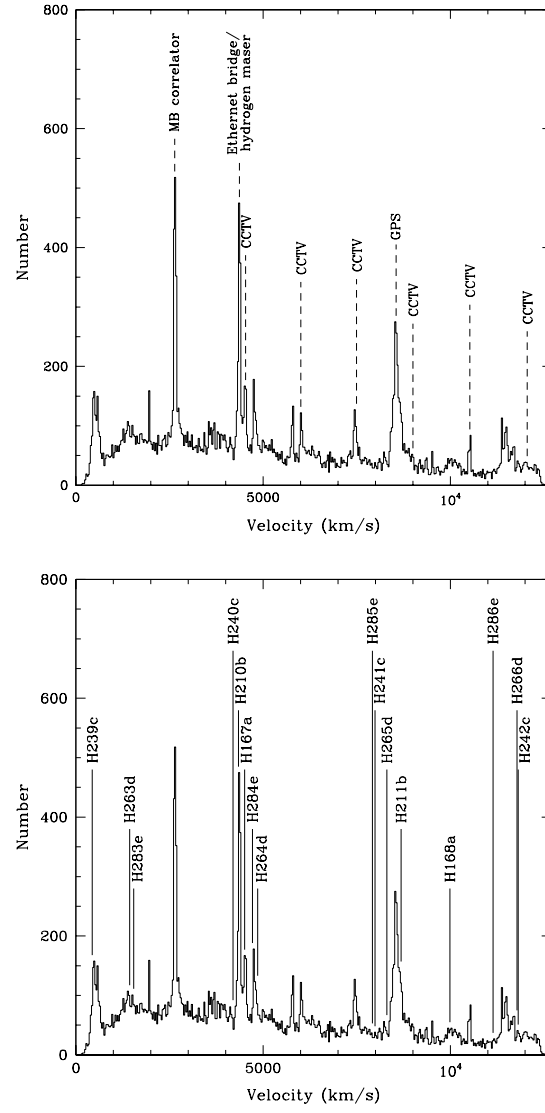


Figure 6. Velocity distribution of HIPASS galaxies after the second verification check, with labels corresponding to potential RFI frequencies (top) and hydrogen recombination frequencies (bottom).

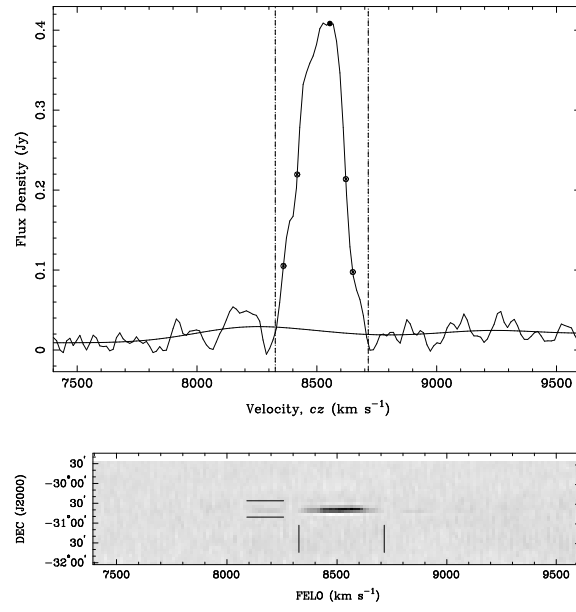


Figure 7. Example spectra and position-velocity maps of GPS L3 beacon

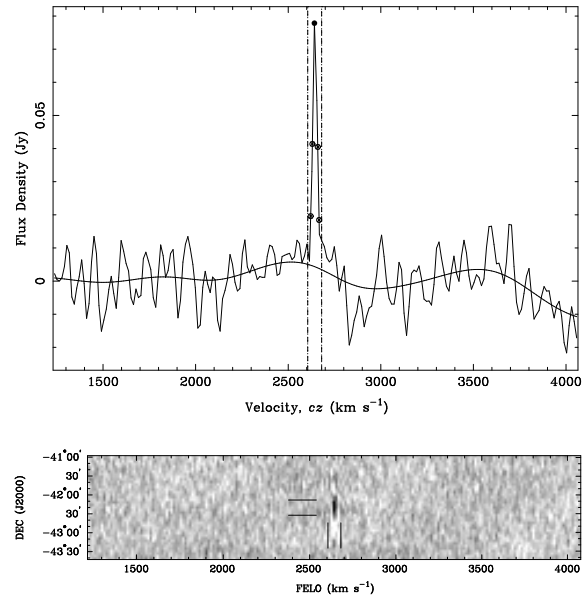


Figure 8. Example spectra and position-velocity maps of narrow-line interference

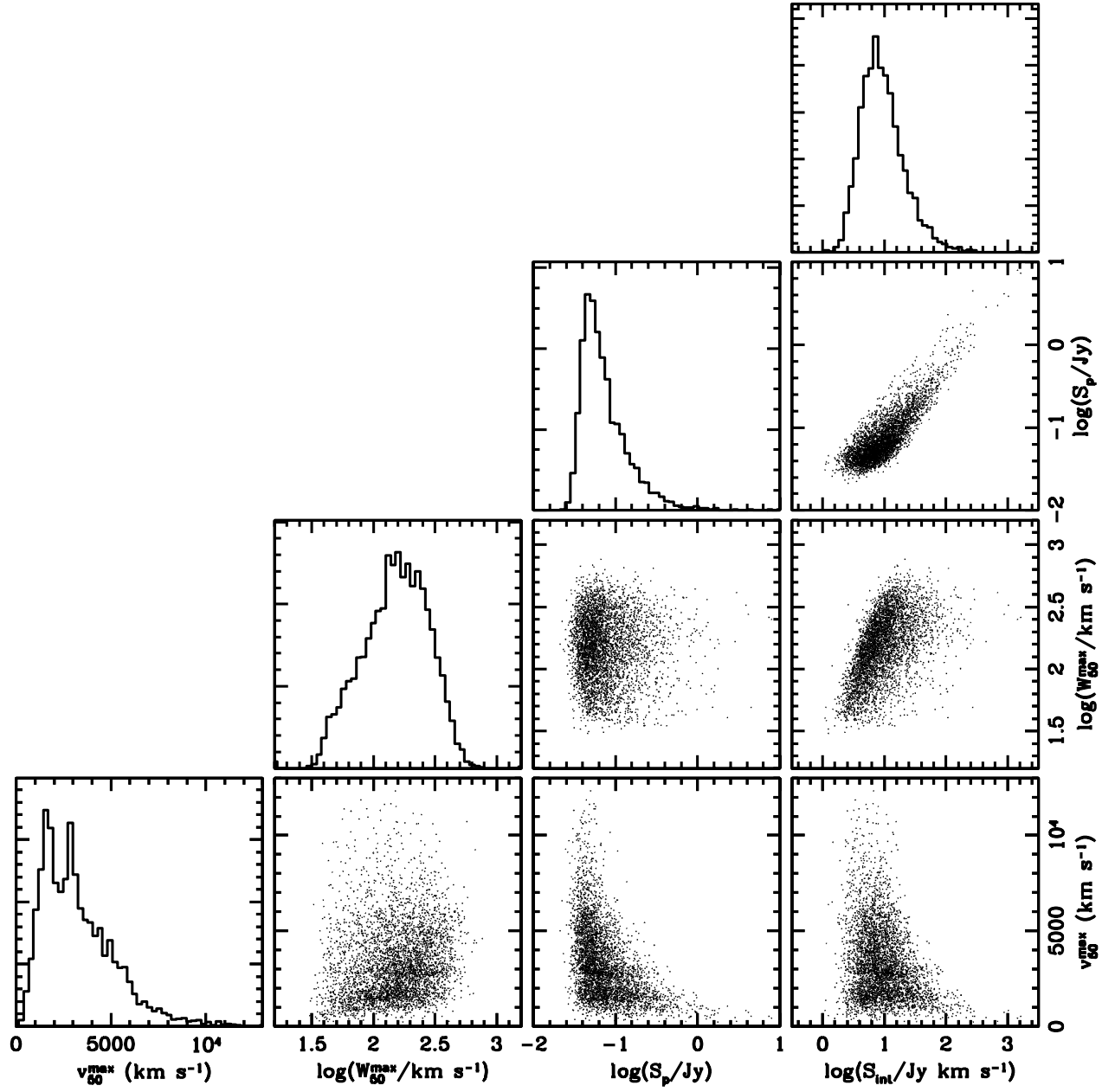


Figure 9. Log-log bivariate distributions of observed parameters: velocity (v_{50}^{\max}), velocity width (W_{50}^{\max}), peak flux (S_p) and integrated flux (S_{int}). Single parameter distributions are also given along the diagonal.

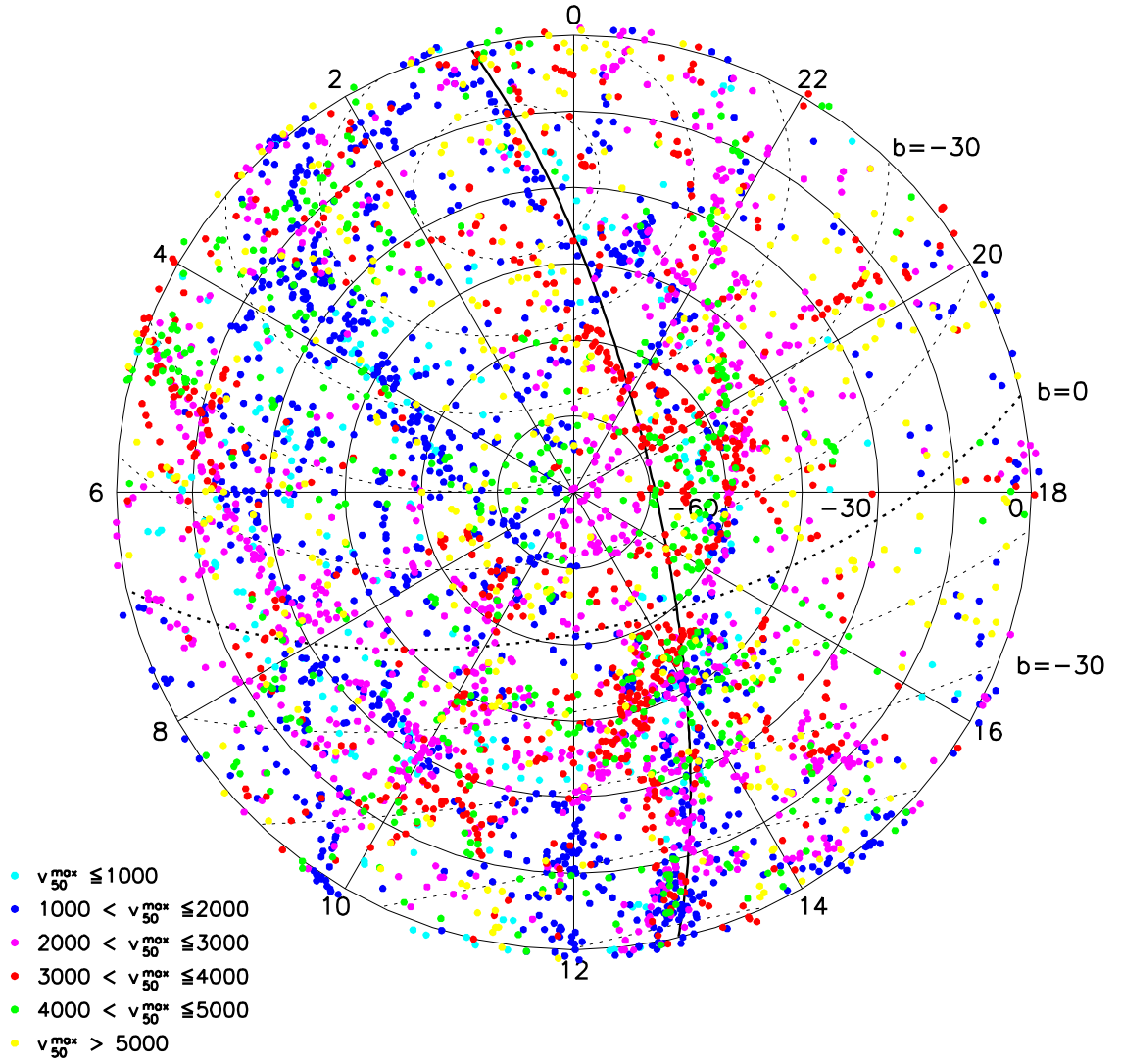


Figure 10. Polar projection of HiCAT detections. Objects are colour-coded by velocity as noted in the key. Coordinate systems are marked as follows: light solid lines - equatorial coordinates; dotted lines - galactic coordinate; dark solid line - super-galactic plane.

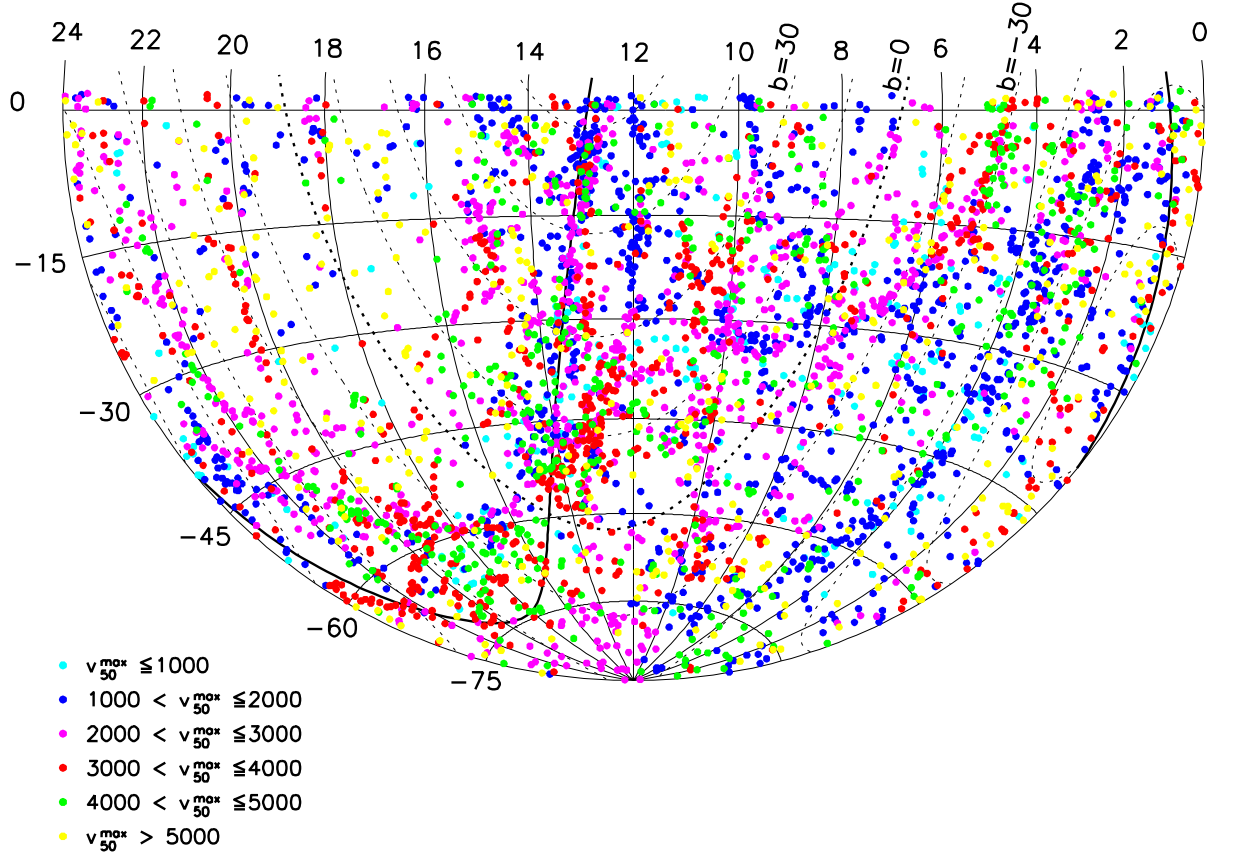


Figure 11. Aitoff projection of HICAT detections. Objects are colour-coded by velocity as noted in the key. Coordinate systems are marked as follows: light solid lines - equatorial coordinates; dotted lines - galactic coordinate; dark solid line - super-galactic plane.

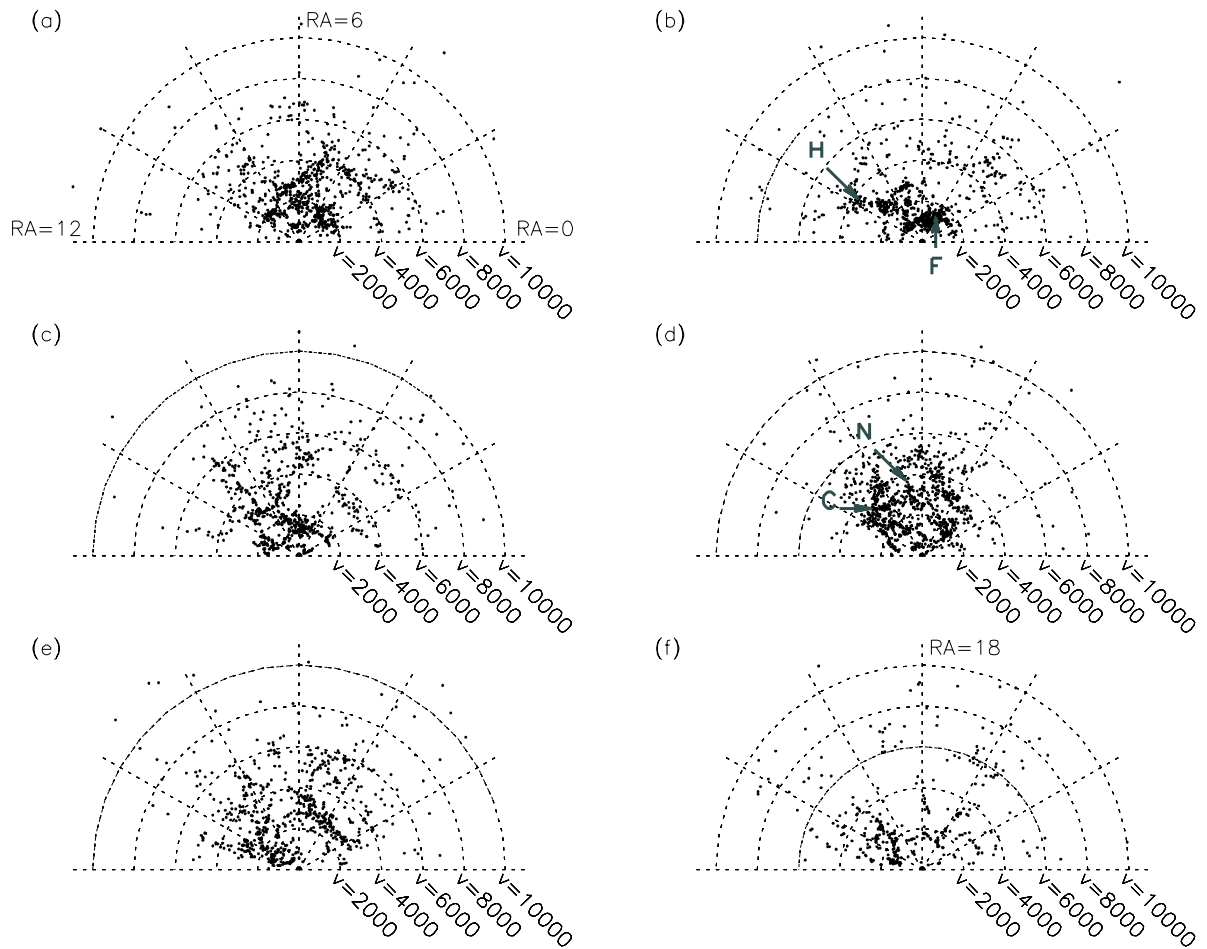


Figure 12. Velocity distribution of HICAT sources divided into segments all hinged on the equatorial plane at the 0-12 hour line starting with segment (a) with one side lying in that plane centred at $RA=6\text{hr}$, (b) above that and so on with (f) again with one side on the plane, but centred at $RA=18\text{hr}$. The location of these wedges in the southern sky is shown in Figure 14.

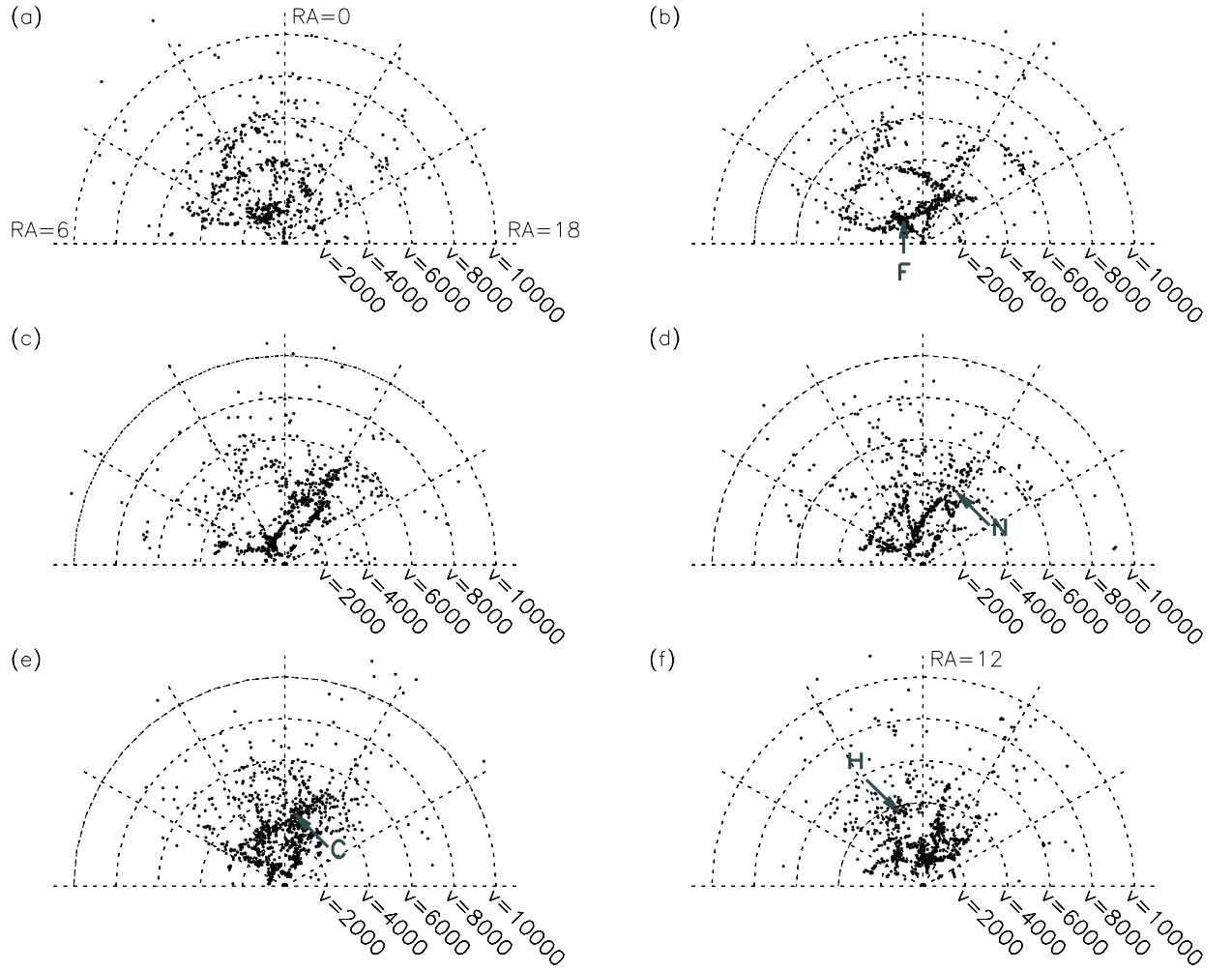


Figure 13. Velocity distribution of HICAT sources divided into segments all hinged on the equatorial plane at the 6-18 hour line starting with segment (a) with one side lying in that plane centred at RA=0hr, (b) above that and so on with (f) again with one side in the plane, but centred at RA=12hr.

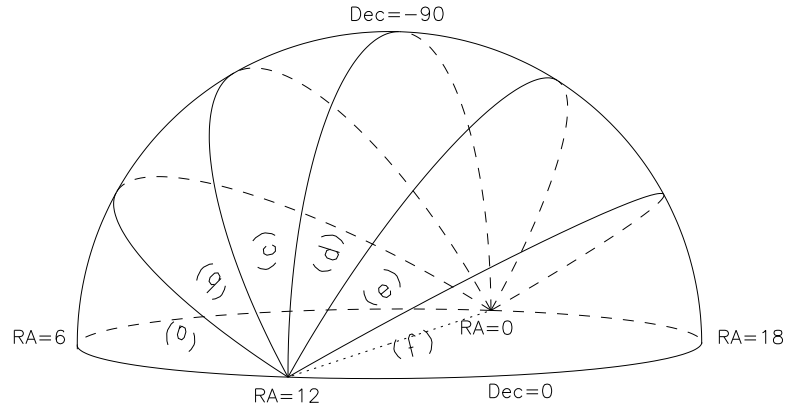


Figure 14. Location in the southern sky of the velocity distribution wedges shown in Figure 12.