Publ. Astron. Soc. Aust., 1999, 16, 12–17

Finding the Bivariate Brightness Distribution of Galaxies from an HI Selected Sample

R. F. Minchin[1,2]

[1] University of Wales, Cardiff, PO Box 913, Cardiff, CF2 5YB, UK
r.minchin@astro.cf.ac.uk

[2] Australia Telescope National Facility, CSIRO, PO Box 76, Epping, NSW 2121, Australia
rminchin@atnf.csiro.au

Received 1998 November 9, accepted 1999 February 1

Abstract: A method is presented that will enable the bivariate luminosity/surface brightness distribution of galaxies to be determined from a relatively small HI selected sample. This will be taken from the HI Parkes All Sky Survey (HIPASS). The advantages of using an HI sample in order to avoid the selection effects that are present at optical wavelengths are discussed. We are developing an algorithm to automatically extract a uniform sample of galaxies from the HIPASS data cubes and to determine the parameters of these galaxies. We have so far conducted tests involving both simulated sources injected into cubes with real noise and data from the Multibeam Deep survey. Results from these tests are encouraging.

Keywords: galaxies: general — galaxies: luminosity function, mass function — galaxies: photometry — radio lines: galaxies

1 Introduction

The luminosity function, usually parametrised as a Schechter (1976) function, is often used to describe a population of galaxies. As surface brightness selection effects are not taken into account by this description, there is an implicit assumption that these can be ignored (McGaugh 1994; Ferguson & McGaugh 1995). However, optically selected samples are known to suffer from serious selection effects that act against low surface brightness objects (e.g. Disney 1976; Impey & Bothun 1997). This means that luminosity functions derived from these optical samples really only describe the way the Universe is populated by relatively high surface brightness galaxies (HSBGs) which are near the peak of the 'visibility function' (Disney & Phillipps 1983; McGaugh, Bothun & Schombert 1995). These galaxies can be seen to much further distances than LSBGs and are therefore preferentially selected in optical surveys.

The bivariate brightness distribution (BBD) will determine the luminosity function as a function of surface brightness. This will describe the population of galaxies more fully than is possible using the luminosity function alone and will determine if there is a correlation between luminosity and surface brightness. If such a correlation does exist then the number of Schechter (1976) $L^\star$ galaxies has probably been determined quite accurately, as the numbers of giant low surface brightness galaxies (LSBGs) will be insignificant. However, this would also suggest that a significant number of dwarf galaxies will have been missed due to SB selection effects, thus adding even greater uncertainty to the poorly determined faint end of the luminosity function. If the correlation is weak or non-existant, then the population of giant LSBGs will be significant and SB selection effects must be taken into account across the whole range of the luminosity function to make an accurate determination.

It is known (Schombert et al. 1992) that LSBGs cover the same range of HI mass as HSBGs. This implies that selection using HI will not be subject to selection effects in the same way as an optical survey. We aim to use HI mass measurements to determine our sample and therefore cover a much wider range of surface brightness than possible with an optical survey. The luminosity and surface brightness of the galaxies will then be determined as part of an optical follow-up programme.

Even though we will avoid optical selection effects, there will still be selection effects inherent in using an HI sample. These include possible biases against LSBGs and high velocity-width galaxies due to profile shape. There is a further possibility of resolving large galaxies and removing them during either bandpass subtraction or baseline removal. These are discussed in more detail in Section 3.

When both HI and optical characteristics of the sample have been determined, we can also investigate relationships other than luminosity–surface brightness. For instance, it has been proposed (McGaugh, Bothun & Schombert 1995)

 1323-3580/99/010012$05.00

https://doi.org/10.1071/AS99012 Published online by Cambridge University Press

that luminosity–scale length may be a more important relationship than luminosity–surface brightness in determining optical selection effects, and it will be possible to test this. We can further examine the Tully–Fisher relationship (Tully & Fisher 1977) for these galaxies and investigate whether it continues to hold true for LSBGs as found by Zwaan et al. (1995).

2 Determining the Sample

There are many advantages in using HI to determine a sample for the BBD. The most important of these is that it is possible to select a sample that spans the full range of luminosity and surface brightness without the need for large numbers of optical observations. In an ordinary, magnitude-limited, optical sample almost all of the galaxies found will have luminosities near $L^\star$ and surface brightnesses near the peak of the visibility function. This makes the construction of a BBD spanning a large range in both luminosity and surface brightness impossible without making observations of thousands of galaxies. By selecting the sample using HI it is possible to cover the required range of luminosity and surface brightness with a much smaller sample due to the lack of optical selection effects. It is therefore possible to determine the BBD with far fewer optical observations than would be needed otherwise.

As the HI masses of the galaxies are known, it is possible to choose the sample in a way that avoids selecting $L^\star$ galaxies preferentially. It appears likely that HI mass works as a tracer of luminosity for spiral galaxies, as larger galaxies will be more luminous and will contain more HI while smaller galaxies will be less luminous and will contain less HI. Previous work on optically selected galaxies has found such a relationship, although LSBGs do tend to have more HI than HSBGs for a given luminosity (de Blok, McGaugh & van der Hulst 1996). If we assume this to be true, then it is possible to select our sample to have equal numbers of dwarfs and giants by binning the galaxies by HI mass and then selecting equal numbers from each bin. This approach means that only a small proportion of the galaxies around $M^\star_{\rm HI}$, analogous to optical $L^\star$ galaxies, need be observed and should enable us to cover two decades in HI mass, from 10^8 to 10^{10} solar masses. This range should be approximately equivalent to a magnitude interval from -17 to -22 in B-band.

The selection in surface brightness is also avoided, to a certain degree, by using HI, as there are no explicit surface brightness selection effects. It is known (de Blok, McGaugh & van der Hulst, 1996) that LSBGs are generally of lower column density than HSBGs, thus Disney & Banks (1997) related column density and optical surface brightness using a simple scaling based on the size of the galaxy. Using this scaling it is estimated that we could detect objects in HIPASS down to a central surface brightness $\mu_{\rm B} = \sim 25{\cdot}5$ mag arcsec^{-2}, so we expect to cover a range of around 5 magnitudes in surface brightness.

In order to get sufficient numbers in these bins to expect statistically meaningful results it will be necessary to obtain data for around 400 galaxies—if these were distributed evenly across a 5×5 grid there would be 16 in each bin, giving 25% accuracy. From the galaxies found so far in HIPASS it appears that approximately 75% are catalogued. Almost all of these galaxies have luminosities and surface brightnesses in the ESO-LV catalogue (Lauberts & Valentijn 1988). This leaves around 100 galaxies in our sample which will require optical follow up. This should be possible in a reasonable time-period.

We hope to use an automatic galaxy finder to determine the catalogue our sample will be taken from (see Section 4). This will enable the sample to be determined objectively and uniformly across different data cubes, which is not possible using the human eye. As full reliability is needed for the optical follow-up involved in this project, spectra selected by the finder will be inspected by eye before the final sample is selected. The primary function of this final screening is to determine which spectra are real galaxies and which are either noise or baseline ripple. While this does introduce some subjectivity into the process, the impact is minimal as the selection of the spectra to be inspected has been made totally objectively and it is normally clear which spectra represent real galaxies.

3 Determining the BBD

Once the sample has been selected and the HI characteristics determined, it will be necessary to find the optical characteristics of the galaxies in the data set. Optical data will be taken from the ESO-LV catalogue and from existing photometric measurements. Where there is no existing data in the literature, the galaxies will be observed in B and R as part of the HIPASS optical follow-up programme. The two main parameters to be determined from the optical follow-up are the luminosity and the surface brightness. Inclination and scale-length will also be determined

The luminosity and surface brightness will be used to place the galaxy onto the BBD. Here each galaxy will be given a weighting based on the volume density of galaxies of similar HI mass, from the HIPASS HIMF (Kilborn, Staveley-Smith & Webster 1999, present issue p. 8), and the number of galaxies in that mass bin. These corrections will give a number per unit volume for galaxies similar to the one being analysed.

Once this analysis has been completed, the pixels on the luminosity–surface brightness plane will all

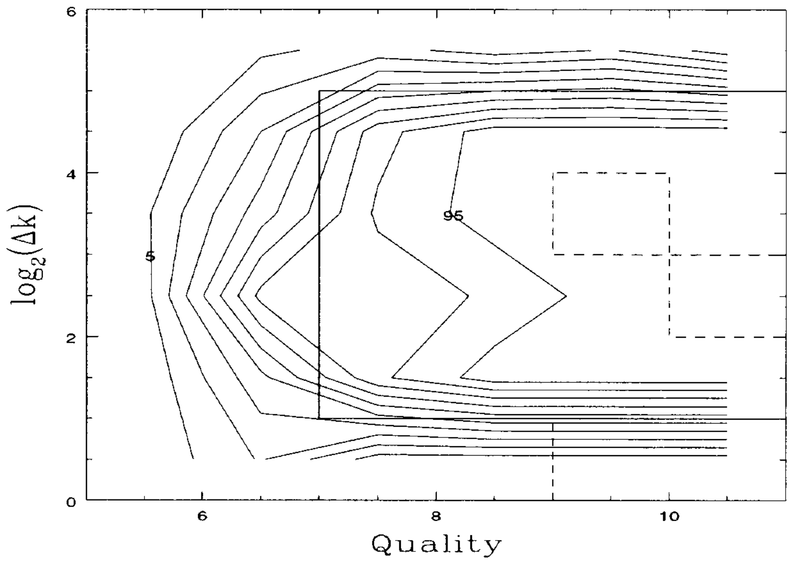

Figure 1—Contour map of reliability from simulation. Contours are from 5% to 95% at intervals of 10%. Dashed-off regions indicate that there were no sources found in this area of parameter space and values have been interpolated. This figure shows how the reliability increases dramatically between Quality 5 and Quality 7 and then increases more slowly above this level, and also how the reliability falls off for low and high velocity width galaxies. The rectangular box shows the 'reliable' region from which the sample will be drawn.

have a total number of galaxies per unit volume associated with them. This will give the BBD for the HI selected sample of galaxies. It should then be possible to see if LSB galaxies do make a significant contribution to the numbers in the Universe, and whether this contribution is limited to low luminosities or spread across the whole range of galaxy sizes.

There are a number of biases that could affect the selection of the sample. The most important of these is that low column-density galaxies will not be found in HIPASS, although if the Disney–Banks scaling is correct then HIPASS will reach a low enough column density to determine the BBD to a very low surface brightness. However, there is the added possibility that the hydrogen in these galaxies will be ionised (e.g. Corbelli & Salpeter 1993; Charlton, Salpeter & Linder 1994; van Gorkom 1993), which would prevent them being found in HIPASS. It is also likely that galaxies with very high linewidths will be missed as their flux will be spread out over a wide velocity range. Another problem is that LSBGs have different rotation curves to HSBGs generally (de Blok, McGaugh & van der Hulst 1996), resulting in these galaxies having lower peak fluxes for the same total flux. This problem should be overcome by the automatic galaxy finder, but it does present a difficulty to detections made using the human eye.

Another problem which can occur is that very large galaxies that overfill the Parkes beam (FWHM 15′) may be removed during the on-line bandpass removal or may be measured as sky by the galaxy finder; in either case such galaxies will not be catalogued. As LSBGs are expected to have a longer scale-length for the same luminosity when compared to HSBGs, they may also have larger HI envelopes. This effect could possibly introduce a bias against finding LSBGs. However, the number of galaxies likely to be affected by the bandpass filter is small, as this will only affect structures greater than 2° across (Barnes et al. 1998) and sky subtraction is much less of a problem for the eye. This has allowed a check to be made on the cube to see if galaxies are being missed due to overfilling the beam. It appears from this that such galaxies do not exist in significant numbers in the velocity range being examined by this project.

4 PICASSO—An Automated Galaxy Finder

We are developing an automated galaxy finder for use on HIPASS data cubes. The results of the tests so far are encouraging. It is currently useable for finding candidate sources in real data cubes, although a by-eye inspection is still necessary to ensure reliability.

The finder, PICASSO, contains a number of sub-processes which are written in FORTRAN 77 and PERL. The most important of these are *finder*, which identifies sources and associates a quality level with them, and *fitter*, which provides accurate 3D positions and linewidths for the galaxies. Other sub-processes remove degenerate sources and parametrise the galaxy using routines from Miriad (Sault, Teuben & Wright 1995).

https://doi.org/10.1071/AS99012 Published online by Cambridge University Press

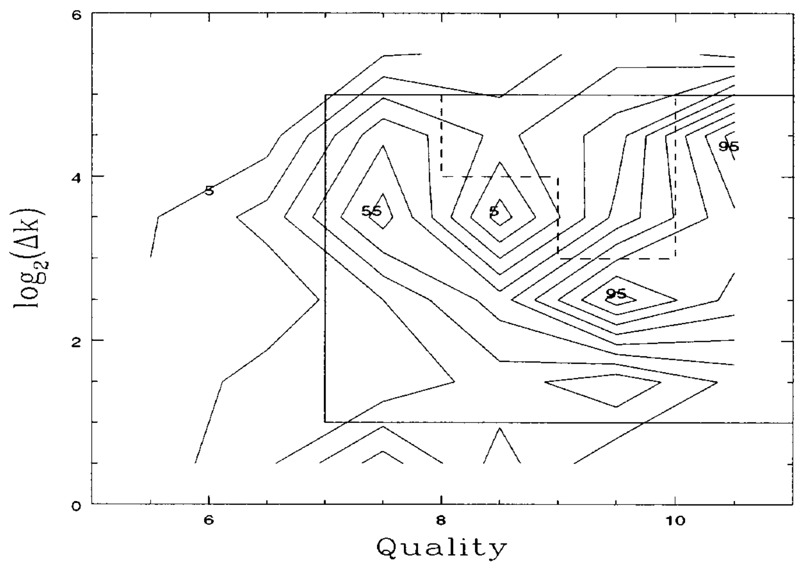

Figure 2—Contour map of reliability from shallow fields before velocity limits were imposed. Contours are from 5% to 95% at intervals of 10%. Dashed-off regions indicate that there were no sources found in this area of parameter space and values have been interpolated. This figure shows that, although there is some increase in reliability as we move towards higher quality, there is very little order and there are large variations. The finder does not become consistently reliable until above Quality 10, which is too high a threshold to be useful. The rectangular box shows the 'reliable' region from which the sample will be drawn.

The quality level output by *finder* is defined as being the signal measured in the pixel being examined, summed over a number of velocity channels, divided by the noise measured in a surrounding annulus, taken over the same number of channels. The quality of a source is therefore its signal-to-noise as seen by the finder.

The quality data from *finder* can be used to determine a maximum distance to which the galaxy would be detected, as the only input into the quality that should change with distance is the flux of the galaxy (the noise on the cubes remains almost constant in the velocity range being examined). The distance to which the galaxy can be seen $d_{\rm max}$ is therefore simply related to the minimum quality that is accepted $Q_{\rm min}$ as

$$\left(\frac{d}{d_{\rm max}}\right) = \frac{Q_{\rm min}}{Q},$$

where Q and d are the detected quality and distance. Therefore $V/V_{\rm max}$ (Schmidt 1968) can be determined from the quality and used to measure the completeness of the sample.

Testing on simulated sources injected onto an HIPASS data cube with real noise indicated that sources could be found reliably above quality 7 and in the range $2 < \Delta k < 32$ (where Δk is the number of channels, each channel being $13 \cdot 2$ km s^{-1} wide), the range over which the finder searches (see Figure 1). Quality 7 was associated, using these simulations, with a signal-to-noise ratio of around 10. It has since been found that pre-processing which was being applied to the cube in order to reduce processing time and memory load was degrading the signal-to-noise. This pre-processing has now been abandoned and quality should now correspond to signal-to-noise. As the pre-processing was degrading the cube before it was searched by the finder, this does not affect the reliability results given here.

The finder was then tested on real data. The data used were from the Deep project (Disney et al. 1999, present issue p. 66) and consisted of three independant cubes at HIPASS coverage (the 'shallow' fields) and the full Deep cube at $12\frac{2}{3}\times$ HIPASS coverage (the 'deep' field). Initially no velocity limit was applied to the source region, and the reliability of the shallow fields was judged by whether a source was also found in the deep field. A contour map of this reliability is given in Figure 2, showing that problems were encountered.

It was found that the density of unreliable sources on the real data was three times higher than on the simulated cube, which also had real noise. It was also found that a large proportion of the 'reliable' sources were actually the peaks of baseline ripple due to strong continuum sources. The large number of unreliable sources was mainly due to noise spikes superimposed on lower continuum peaks. In order to avoid this baseline structure, the region used was limited in velocity to $v < 8000$ km s^{-1}, at the same time a lower limit to the velocity of $v > 200$ km s^{-1} was introduced to reject contamination from HI in our Galaxy. This lower limit is raised to 1000 km s^{-1} for the BBD sample in order to reject nearby galaxies with poorly-defined distances.

The sources in this region were then checked by eye on the deep field cube and a judgement made as to their reliability. In addition to this, the source list from the deep field was compared to an independent by-eye survey of the deep field that had been carried out in Cardiff (Disney et al. 1999). It was determined that, at the $Q > 7$ level, PICASSO had a reliability of just over 70% and found about same number of true galaxies as the by-eye survey, with an overlap between the lists of 75%. At the $Q > 10$ level, PICASSO was found to be around 95% reliable. This region contains just over 60% of the sources found. The reliability for the shallow and deep fields, and an average of the two, is given in Figure 3.

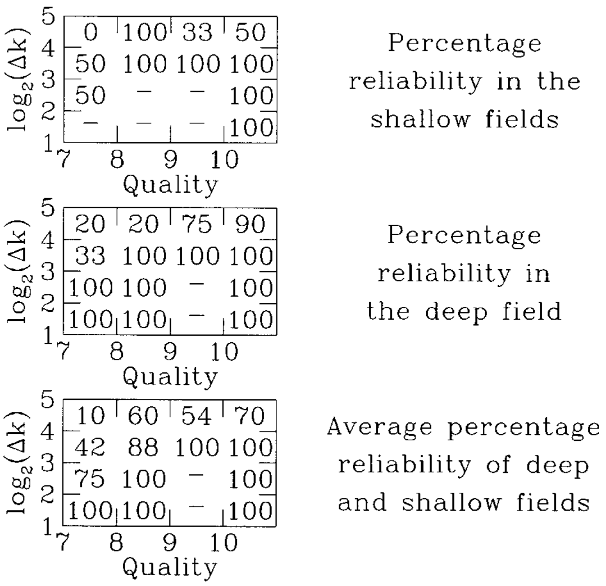

Figure 3—Reliability percentages for velocity limited data. A marked improvement is seen over the reliability without the velocity limit, shown in Figure 2.

As part of the post-processing of the PICASSO data, moment maps are constructed using a $44' \times 44'$ window over a range in velocity of twice the fitted velocity width, centred on the fitted position. The spatial size is chosen to be three times the FWHM of the Parkes beam. There is only a small chance of confusion as galaxies would need to be close in velocity as well as spatially. A gaussian fit and a base-line offset are fitted simultaneously to the zeroth-order moment map using the *imfit* routine in Miriad. As long as the galaxy is a point source, the peak value of the gaussian is the integrated flux of the galaxy, which can be used to determine the mass of the galaxy. The *imfit* routine also gives an error on the peak, which gives an estimate of the error on the mass. It is hoped, although it has not yet been tested, that this will give another discriminator against spurious detections and therefore further reduce the subjectivity of the final sample by cutting out false candidates before the by-eye inspection.

A slight caution should be attached to the reliabilities determined for PICASSO on these cubes. Although real data cubes were used, the observations used to construct them were all made at night time, so these cubes are virtually free of the solar ripple that can affect standard HIPASS cubes. However, the region is fairly close both to the galactic plane and to the strong radio source in NGC 5128 and contains a relatively high number of strong continuum sources. The polarisation subtraction used to construct the real-noise cubes for the injection of the simulated sources means that these cubes are similarly free of solar ripple. The reliability of PICASSO when run on a standard cube has been found to vary widely from cube to cube, although this is not a problem for this project due to the by-eye inspection of the spectra before selection of the final sample.

5 Summary

The methods given above will enable the BBD of an HI selected sample of galaxies from HIPASS to be determined. The sample will be free of optical selection effects, although we may still suffer some discrimination against LSBGs due to column-density limits and profile shapes.

In determining the BBD we will investigate the significance of LSBGs in the population of galaxies as a whole and will determine how the luminosity function varies with surface brightness. This will answer the question as to whether or not HI rich giant LSBGs make up a significant fraction of the population of giant spirals.

We have developed a galaxy finder that can be used to extract a catalogue from the HIPASS cubes in order to obtain an objectively selected sample of galaxies. The finder is still under development but is giving acceptable reliability and similar numbers of true galaxies to by-eye searching. It also delivers a parametrisation of the galaxies found, including the HI mass that is of prime importance in determining the sample and $V/V_{\max}$.

Acknowledgments

RM acknowledges the support of a PPARC postgraduate research award and the help of the Multibeam Science Working Group in making the HIPASS observations. PICASSO was written in association with Alan Wright. I would like to thank Mike Disney, Erwin de Blok, Helmut Jerjen and Peter Boyce for useful discussions on the BBD project and Erwin de Blok, David Barnes, Lister Staveley-Smith and Ron Ekers for discussions on PICASSO and sample selection. I would also like to thank Mike Disney, Gareth Banks, Erwin de Blok, Peter Boyce and the referee for useful comments on this paper.

https://doi.org/10.1071/AS99012 Published online by Cambridge University Press

References

Barnes, D. G., Staveley-Smith, L., Ye, T., & Oosterloo, T. 1998, in ADASS VII (San Francisco: ASP), p. 89

Charlton, J. C., Salpeter, E. E., & Linder, S. M. 1994, ApJ, 430, L29

Corbelli, E., & Salpeter, E. E. 1993, ApJ, 419, 104

de Blok, W. J. G., McGaugh, S. S., & van der Hulst, J. M. 1996, MNRAS, 283, 18

Disney, M. J. 1976, Nature, 263, 573

Disney, M. J., & Banks, G. 1997, PASA, 14, 69

Disney, M. J., & Phillipps, S. 1983, MNRAS, 205, 1253

Disney, M. J., Boyce, P. J., Banks, G. D., Minchin, R. F., & Wright, A. E. 1999, PASA, 16, 66

Ferguson, H. C., & McGaugh, S. S. 1995, ApJ, 440, 470

Impey, C., & Bothun, G. 1997, ARA&A, 35, 267

Kilborn, V., Staveley-Smith, L., & Webster, R. 1999, PASA, 16, 8

Lauberts, A., & Valentijn, E. A. 1988. Astronomy from Large Databases: Scientific Objectives and Methodological Approaches (Garching: ESO), p. 37

McGaugh, S. S. 1994, Nature, 367, 538

McGaugh, S. S., Bothun, G. D., & Schombert, J. M. 1995, AJ, 110, 573

Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, in ADASS IV (San Francisco: ASP), p. 433

Schechter, P. 1976, ApJ, 343, 94

Schmidt, M. 1968, ApJ, 151, 393

Schombert, J. M., Bothun, G. D., Schneider, S. E., & McGaugh, S. S. 1992, AJ, 103, 1107

Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661

van Gorkom, J. H. 1993, in The Environment and Evolution of Galaxies, ed. J. M. Shull & H. A. Thronson (Dordrecht: Kluwer), p. 345

Zwaan, M. A., van der Hulst, J. M., de Blok, W. J. G., & McGaugh, S. S. 1995, MNRAS, 273L, 35