*Statistical Challenges in 21st Century Cosmology*
*Proceedings IAU Symposium No. 306, 2014*
*A. F. Heavens, J. -L. Starck & A. Krone-Martins, eds.*

doi:10.1017/S1743921314011053

# Deep Source-Counting at 3 GHz

**Tessa Vernstrom, Jasper Wall and Douglas Scott**

Department of Physics and Astronomy, University of British Columbia
6224 Agricultural Road, Vancouver, Canada V6T 1Z1
emails: tvern@phas.ubc.ca, jvw@phas.ubc.ca, dscott@phas.ubc.ca

**Abstract.** We describe an analysis of 3-GHz confusion-limited data from the Karl J. Jansky Very Large Array (VLA). We show that with minimal model assumptions, *P(D)*, Bayesian and Markov-Chain Mone-Carlo (MCMC) methods can define the source count to levels some 10 times fainter than the conventional confusion limit. Our verification process includes a full realistic simulation that considers known information on source angular extent and clustering. It appears that careful analysis of the statistical properties of an image is more effective than counting individual objects.

**Keywords.** source count, confusion limit, *P(D)*

## 1. Introduction

Counts of objects as a function of apparent intensity have been important in understanding the cosmology of our Universe since Herschel's day (Herschel 1785, "On the Construction of the Heavens"). *Evolution* of our Universe was first claimed by Ryle & Scheuer (1955), but the claim was tangled with the simultaneous and unwelcome discovery of *confusion*. Confusion is the integration of the myriad of faint sources into a lumpy background, the "confusion limit" (Fig. 1). The subsequent fight over the interpretation of radio source counts – the Big Bang vs. Steady State controversy – obscured the momentous implications of evolution. In the heat of the battle, Scheuer (1957) showed how confused data could be used to extract information on the source count to levels down to and below this limit. His elegant paper describes *P(D)*, Probability-of-Deflection analyis, now known in terms of one-point statistics.

There are pressing reasons for measuring deep radio source counts. Firstly there is serious disagreement between different direct counts to faint intensities at or near frequencies of 1.4 GHz, as shown in Fig. 2. These are not explicable by sample variance or clustering (Condon 2007). Resolution of this issue is important in view of observations from a balloon experiment (Fixsen *et al.* 2011, ARCADE 2) finding excess radio emission, possibly due to broad-beam integration of discrete sources constituting a newly intruding faint-intensity population (Vernstrom *et al.* 2011; Seiffert *et al.* 2011). *Inter alia*, accurate counts to faint intensities provide powerful data for deriving the cosmic evolution of star-formation. In addition, design of the Square Kilometre Array (SKA) is in progress, as well as SKA pathfinder-telescope surveys. Deep counts, confusion issues and consequent survey plans play into the final design of this billion-dollar experiment.

## 2. Observations: the Survey Map

To resolve the disagreements Condon *et al.* (2012, CO12) used the VLA in a single-pointing 57h integration at the position of the Owen-Morrison (2008) survey in the Lockman Hole. The 1.4-GHz count from this survey is the most discrepant at faint intensities (Fig. 2). The field is covered in many other wavebands (e.g. *Spitzer, Chandra,*

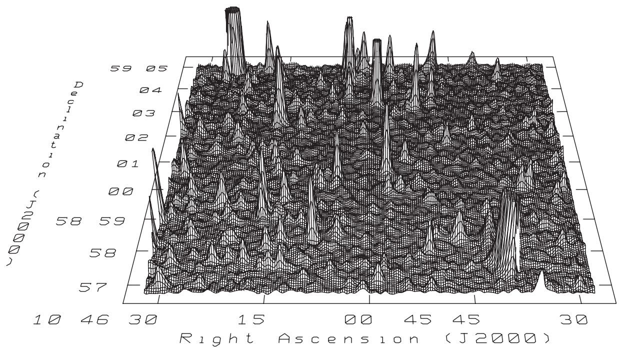


**Figure 1.** The 3-GHz image from a 57h integration with the VLA in C-configuration (beam 8 arcsec FWHM). The instrumental noise is 1.02 $\mu$Jy/beam. All features in the image, positive and negative, are on beamwidth scales and these cover the image, indicating that the "confusion limit" has been reached. The large positive deflections represent discrete radio sources which produce the long positive tail of the *P(D)* distribution (Fig. 3).

*Herschel*, GMRT, CFHT) to allow cross-identifications. The central frequency was 3.02 GHz with the full bandwidth of 2 GHz; the synthesized beam was 8 arcsec FWHM. The survey field is governed by primary beam size, which changes from 22 arcmin FWHM at 2.0 GHz to 11 arcmin at 4 GHz. The broad bandwidth meant rethinking standard analysis procedures and in particular we split the bandwidth into 16 sub-bands and processed these independently. Details of the reduction procedures are in CO12. The resulting survey image (Fig. 1) was confusion-limited as designed, with an instrumental noise of 1.02 $\mu$Jy/beam.

CO12 used this image with a single-beam *P(D)* analysis (Condon 1974) to estimate the source count at 3 GHz down to 1 $\mu$Jy. CO12 found a monotonic decrease in the (Eudlidean-normalized) 3-GHz source count, agreeing with predictions from models of luminosity functions and evolution (Massardi *et al.* 2010), and ruling out the appearance of any new population at flux densities above 1 $\mu$Jy. The CO12 analysis considered only a single short power law over the range 1—10 $\mu$Jy to describe the counts [best value log(dN/dS) = log(9000)-1.7 log(S)], but the process showed that far more information is present in the *P(D)* distribution. It is the statistical approach to extract this information that we briefly describe here, a process fully described by Vernstrom *et al.* (2014a, V14).

## 3. Modelling the Source Count

All stages in this detailed analysis became iterative. Instrumental noise must be accurately assessed, as it is of similar central width to the *P(D)* distribution. The *P(D)* distribution itself was compiled with several different sampling procedures, correcting each pixel height for the (frequency-dependent) primary-beam response. There is a trade-off between increasing the field-radius over which *P(D)* is measured to yield more data, while noise rises with radius inversely as the primary-beam response until it overwhelms confusion signal. We performed multiple tests to find the optimum pixel sampling to reduce correlations between binned data points, as well as to test weighting schemes that take into account the radial-dependent noise. Model-selection also folded into this process. We wanted a model that makes minimal assumptions about the form of the count; and eventually – again after trials – we adopted the multi-power-law model of Patanchon *et al.* (2009) in which the count is approximated by a series of short power-law sections, the variables then being the node heights and their selected positions along the

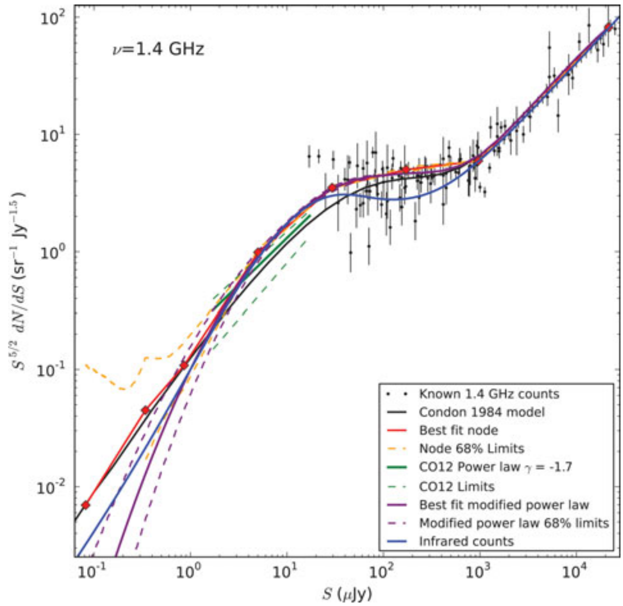


**Figure 2.** Source counts at 1.4 GHz, shown in relative differential form, i.e counts per bin of intensity divided by the counts expected in a static Euclidean universe, so that a horizontal line indicates an integral power-law count of index -3/2. Points and error bars show the previous badly scattered estimates of the counts. The short straight line plus dashed lines above and below it show results from CO12, the early interpretation of the current deep survey data, while joined points plus dashed error limits show the present results, translated from 3.0 to 1.4 GHz with a spectral index of -0.7. Smooth curves represent counts calculated from models of radio and IR luminosity functions and cosmic evolution.

flux-density axis. Such a model loses no cosmological information provided that the nodes are sufficiently closely spaced in log(S). Uncertainties in all trials were assessed with MCMC sampling, using a generic sampler from the COSMOMC package (Lewis & Bridle 2002). The final noise and *P(D)* distributions used apply to the central 5 arcmin radius of the image (Fig. 3); the noise, weighted following the primary beam response, is accurately Gaussian with $\sigma_n = 1.255\ \mu$Jy.

To test our chain of precedures, we used the SKADS simulation (Wilman *et al.* 2008), which computes a source count at the nearby freqeuncy of 1.4 GHz using best estimates of luminosity functions and their evolutions. The simulation data extend to 1 nanoJy, as we require. The simulation includes realistic angular-size and clustering information, and it enabled us to demonstrate retrieval of the exact source count of Wilman *et al.* In the end we fixed the two highest flux-density nodes at positions very well defined by direct source counts (Fig. 2), and floated 6 other nodes down to 1 nanoJy. These two fixed nodes may be considered as priors; in addition we set weak constraints for the node heights to limit MCMC search time, but none of these priors constrained the results significantly. We found it unecessary to iterate the simulation; the Wilman analysis and model count proved to be relatively accurate. Our testing procedure included allowing the noise to be a Gaussian of indeterminate width; the results came back gratifyingly close to our measurement. We tested node frequency and position. The Pearson correlation matrix showed that there is degeneracy between the node heights, telling us that adding any further nodes would not lead to improved count definition.

https://doi.org/10.1017/S1743921314011053 Published online by Cambridge University Press

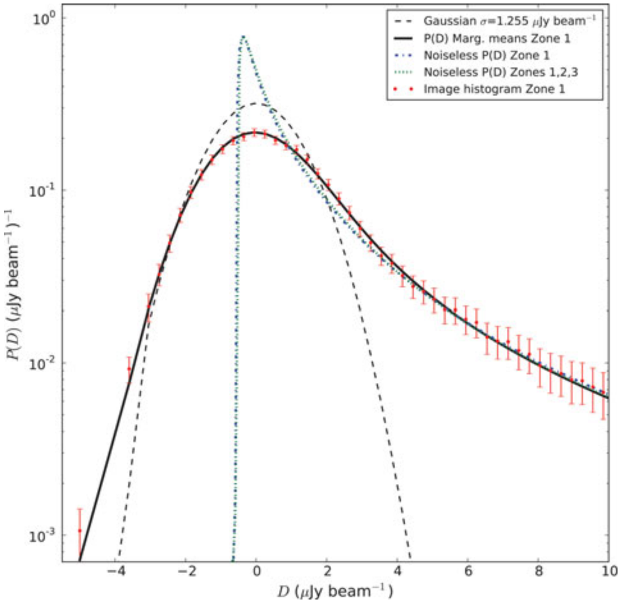


**Figure 3.** Oversampled *P(D)* distribution for the central 5 arcmin radius of the 3-GHz image (dots and $\sqrt{N}$ error bars), with the best-fitting 8-node model (solid curve) of Fig. 1. The dashed line represents a Gaussian of $\sigma$ =1.255 $\mu$Jy, accurately describing the instrumental noise. The dot-dashed line is the noise-free *P(D)* distribution corresponding to the best-fit count model.

## 4. Results

These are shown in Fig. 2. The *P(D)* distribution contains infomation on the 3-GHz source count down to a level of about 100 nanoJy, at least 10 times below the accepted confusion limit. Integration of this count indicates a source surface density of $2.4 \times 10^5$ deg$^{-2}$ at $1.0 \times 10^{-7}$ Jy. A resolution of 8 arcsec should encompass the extent of most star-forming galaxies at this level, but there remains the possibility of extended-emission sources, such as cluster halos. In a further project (Vernstrom *et al.* 2014b), we have obtained deep data with the Australia Telescope Compact Array (ATCA) in a 7-pointing mosaic at 1.75 GHz, now with a beam on 1 arcmin scales. We are searching for any contribution of extended emission, perhaps low-level AGN emission resolved out with our 8 arcsec VLA beam, or synchrotron radiation from cluster halos, possibilites here including relics and mini-halos (Feretti *et al.* 2012) or WIMP Dark Matter annihilation (Fornengo *et al.* 2011). We have tentative evidence for an excess count corresponding to such low surface-brightness products. If the result stands, the count numbers are nearly a factor of two above the count of Fig. 2. Such a population cannot be from star-formation processes, because the radio – FIR correlation would result in an FIR background well above that observed. Cluster halos from relic radiation or from Dark Matter annihilation are the more likely.

**In summary**, we have shown how the marriage of confusion-limited imaging, *P(D)*, and modern statistical techniques (Bayesian likelihood, MCMC) yields results on source counts of major astrophysical/cosmological significance. We have looked in detail at statistical *P(D)* issues for broad-band radio interferometry, including noise and sampling. We have shown that the *P(D)* distribution contains information on the source count down to intensities 10 times below the standard "confusion limit". The 3-GHz count (Fig. 2), now defined down to 50 nanoJy, demonstrates a monotonic sinking to below this level.

https://doi.org/10.1017/S1743921314011053 Published online by Cambridge University Press

The results clean up the ambiguities from direct source counts, rule out the intrusion of "new" populations above 1 nanoJy, and indicate that current models of radio emission from star-forming populations require only minor modification. Lastly, we have estimates of source surface densities down to nanoJy levels, vital data for the SKA project.

## References

Condon, J. J. 1974, *Astrophys. J.*, 188, 279
Condon, J. J. 2007, in *ASPCS*, eds J. Afonso, *et al.* 380, 189
Condon, J. J., Cotton, W. D., Fomalont, E. B., *et al.* 2012 [CO12], *Astrophys. J.*, 758, 23
Feretti, L., Giovannini, G., Govoni, F., & Murgia, M. 2012, *Astron. Astrophys. Rev.*, 20, 54
Fixsen, D. J., Kogut, A., Levin, S., *et al.* 2011, *Astrophys. J.*, 734, 5
Fornengo, N., Lineros, R., Regis, M., & Taoso, M. 2011, *Phys. Rev. Lett.*, 107, A261302
Herschel, W. 1785, *Phil. Trans. R. Soc. Lond.*, 75, 213
Lewis, A. & Bridle, S. 2002, *Phys. Rev. D*, 66, 103511
Massardi, M., Bonaldi, A., Negrello, M., *et al.* 2010, *Mon. Not. R. astr. Soc.*, 404, 532
Patanchon, G., Ade, P. A. R., Bock, J. J., *et al.* 2009, *Astrophys. J.*, 707, 1750
Ryle, M. & Scheuer, P. A. G. 1955, *Proc. Roy. Soc. Lond. A*, 230, 448
Scheuer, P. A. G. 1957, *Proc. Camb. Phil. Soc.*, 53, 764
Seiffert, M., Fixsen, D. J., Kogut, A., *et al.* 2011, *Astrophys. J.*, 734, 6
Vernstrom, T., Scott, D., & Wall, J. V. 2011, *Mon. Not. R. astr. Soc.*, 415, 3641
Vernstrom, T., Scott, D., Wall, J. V., *et al.* 2014a [V14], *Mon. Not. R. astr. Soc.*, 440, 2791
Vernstrom, T., Norris, R. P., Scott, D., & Wall, J. V. 2014b, submitted
Wilman, R. J., Miller, L., Jarvis, M. J., *et al.* 2008, *Mon. Not. R. astr. Soc.*, 388, 1335