Astronomical Image Processing

David Christopher Ragusa

22/4/2015

Abstract

An optical CCD image of the extragalactic sky was computationally processed to identify galaxies. A catalogue was produced containing coordinates, magnitude and other information for the galaxies identified. From this, number counts of galaxies were derived and the results compared to the standard formula for uniform galaxy distribution $\log N(m) = 0.6\,m + const$. The data was found to have two major deviations from the formula, caused by foreground contamination and aperture limitations for bright objects, and hitting the sensitivity limit for dimmer objects. Ignoring these deviations, a linear relation was found with a gradient of 0.654 ± 0.002, which agrees quite well with the expected value. No evidence of galaxy evolution was found in the data.

Theory

The standard formula for uniform galaxy distribution can be derived as follows.

Assuming a flat (Euclidean) universe, number of galaxies $N = \frac{4}{3}\pi r^3$.

The flux-luminosity relationship gives flux $F = \frac{L}{4\pi} r^{-2}$.

Hence $N = const \times F^{-3/2}$, which leads to $\log N = -1.5\,F + const$.

Flux can be converted to magnitude via $m_1 - m_2 = -2.5 \log\left(\frac{F_1}{F_2}\right)$.

This gives $\log N = \frac{-1.5}{-2.5} m + const$ which is equal to the expected:

$$\log N(m) = 0.6\,m + const \quad (1)$$

The original goal of performing galaxy surveys was to confirm the assumption that space is flat – the formula assumes a flat Euclidean universe, so any deviation from this would infer a different volume element of the universe.

However, because observations of galaxies over a range of distances also means observations over a range of times in the past, any deviations from the formula could also give indications on galaxy evolution, such as number density or overall brightness changing over time. This effect is much stronger than any effects on the volume element of the universe, so galaxy surveys unfortunately cannot be used for that purpose.

Method

The `astropy` module was used to read the fits file into Python for processing. The very first thing done was to make a histogram of count values in order to get a feel for what the background count level might be. Figure 1 shows a large peak at around 3000 counts (this goes completely off the scale of the y-axis).

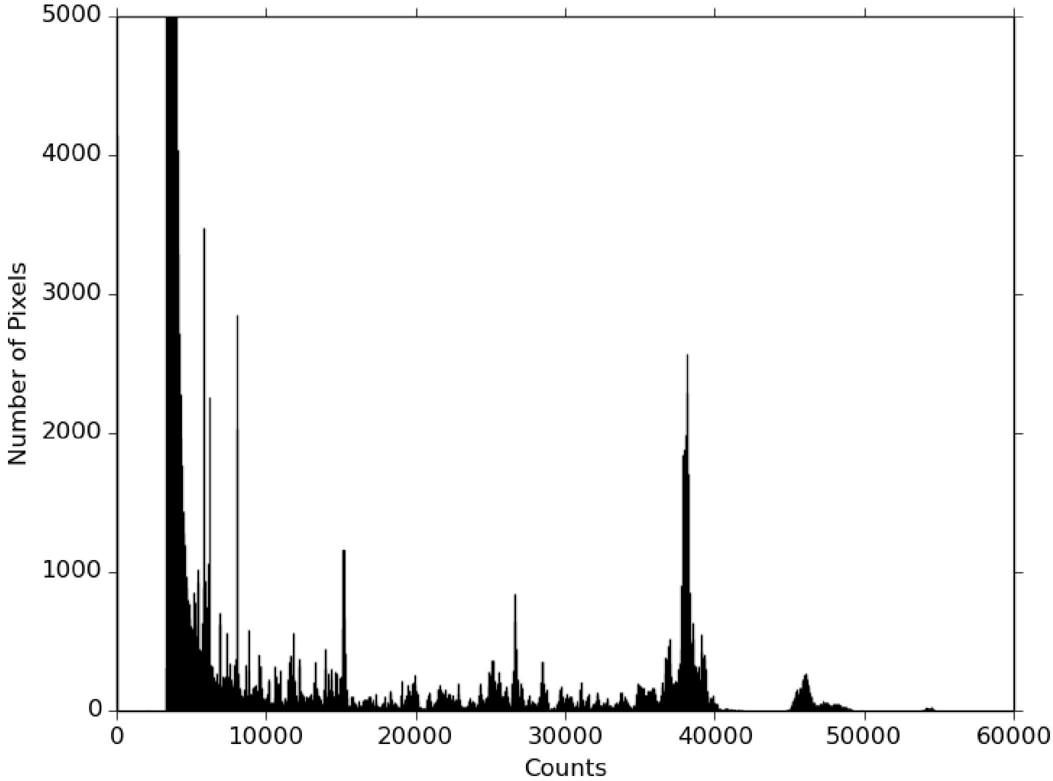

Fig. 1: Histogram of pixels per count – whole count range

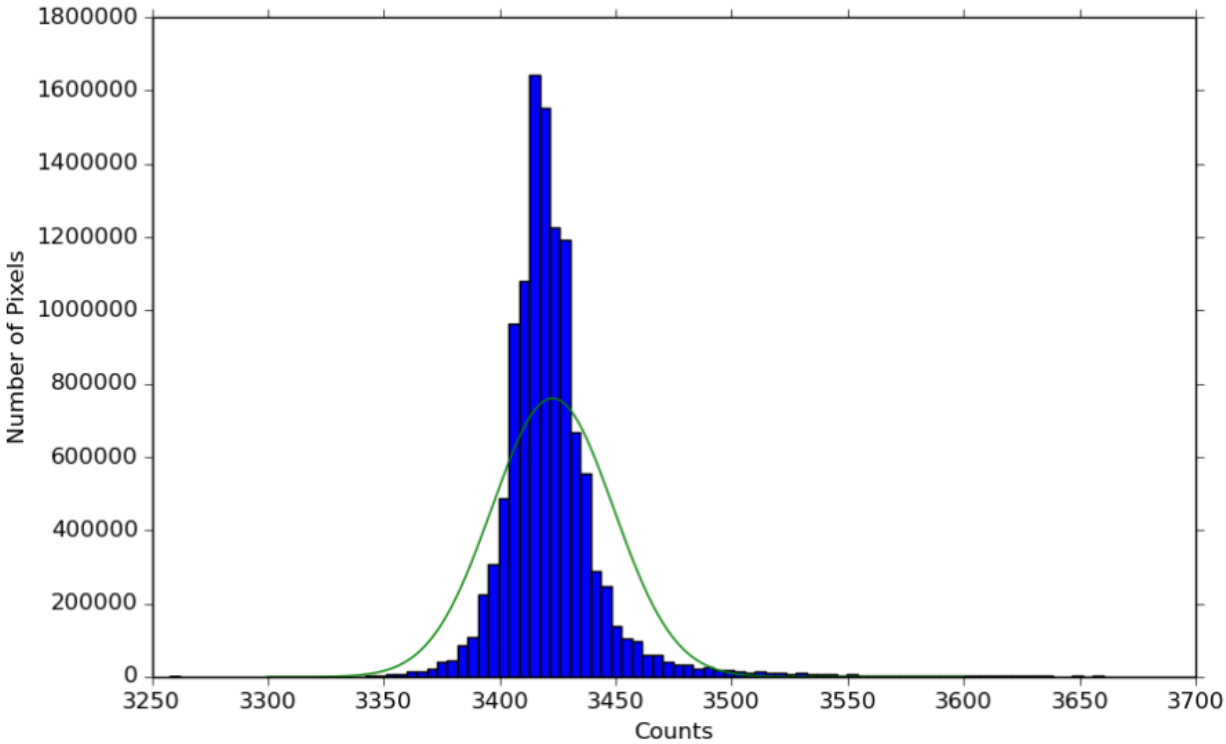

Fig. 2: Histogram of pixels per count - 3200-3700 counts

After a few more histograms were produced to narrow down the area of interest, a final one was produced ranging from 3200 to 3700 (figure 2). The mean and standard deviation were calculated off this range (3423 and 26 respectively), and a normal-distribution-like curve with the same parameters is overlaid.

Based on these results, the background count was taken to be 3423. A decision then had to be made on what confidence level to detect galaxies at. A 4-sigma level was used, but led to far too many spurious results. Figure 3 shows a surfeit of insignificant results around the central star (the black masked out area). A 5-sigma confidence level was used instead, meaning any value below 3553 counts was discounted.

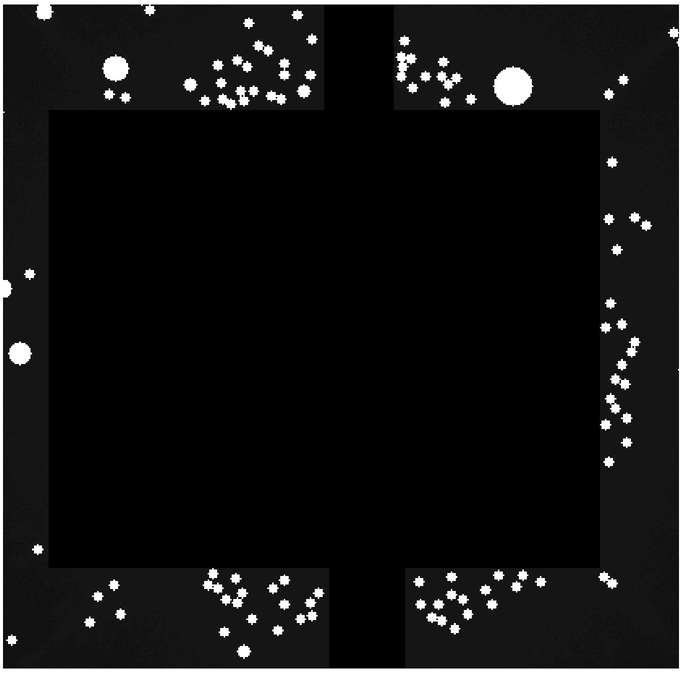

Fig. 3: 4-sigma results around the central star

The first processing step consisted of smoothing out any dead pixels (those with a count of 0), due to defective detector elements. This was done simply by replacing any such pixels with the average of their neighbours.

A mask array was used to signify which areas of the image had been dealt with. It was simply an array filled with ones the same dimensions as the astronomical image. A masked area on the image had its corresponding area on mask replaced with zeros.

Before any image segmentation could be performed as the subsequent step, all the artifacts in the image such as the large central star and other stars, had to be masked out. This was for two reasons: they were far brighter than galaxies in the image and therefore would have skewed the detection and categorisation; but more importantly, their CCD elements had 'bloomed' – the electrons on the elements had overflowed to neighbouring ones, producing bright mostly vertical lines (fig. 4). Thus the bloomed areas had to be masked out as well.

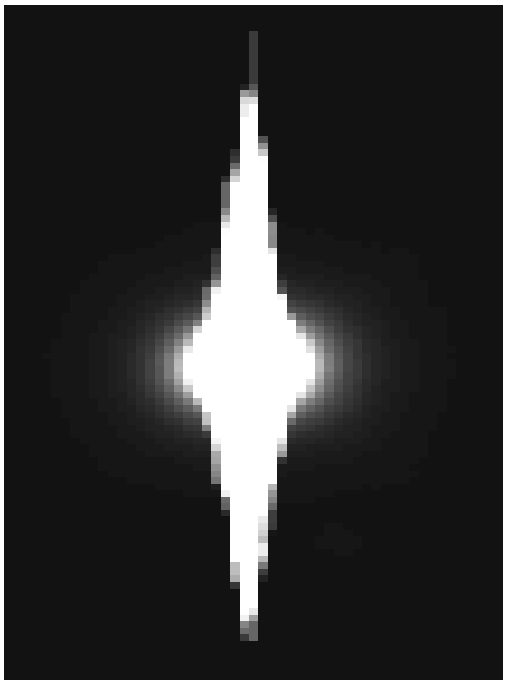

Fig. 4: Star around (2130, 3760), showing bloom

This process was done by visual inspection. Even though a rhomboid would have perhaps been a better shape for most masks, rectangles were chosen due to the ease of notation – only two opposite corner coordinates were needed. The masking was performed by simply iterating over rows determined by the coordinates. Additionally, a mask was applied around the very edge of the image, to avoid possible index errors when processing galaxies.

Three main arrays were used to process the file: `originaldata`, which was not modified at all beyond the initial smoothing; `workdata`, in which any areas of the image catalogued were set to 0; and the mask array as described previously.

Image segmentation was performed quite simply – a list of pixels with the current maximum value in the `workdata` image was found. If the pixel lay in a masked area, it was set to zero and discarded immediately. If however it lay in an available area, it was assumed to be a galaxy and additional processing was applied to it. This processing, whether it produced a cataloguable galaxy or not, set the pixel to 0. In this way, the next execution of the routine to get a list of maxima always returns a new list of pixels.

The processing of galaxies presented problems. At first, a fixed circular aperture with a radius of 6 pixels was used. However, this had the significant problem of being altogether too small for certain large galaxies. Figure 6 shows a 6 pixel aperture as being far too small for the galaxy it represents. Another method was thus developed, which counted out from the centre until the value fell below the 5 sigma level, in the 8 cardinal directions. The longest of these lines was taken to be the radius of the aperture. Figure 6 shows this method applied to the same galaxy as fig. 5 – the results are much improved.

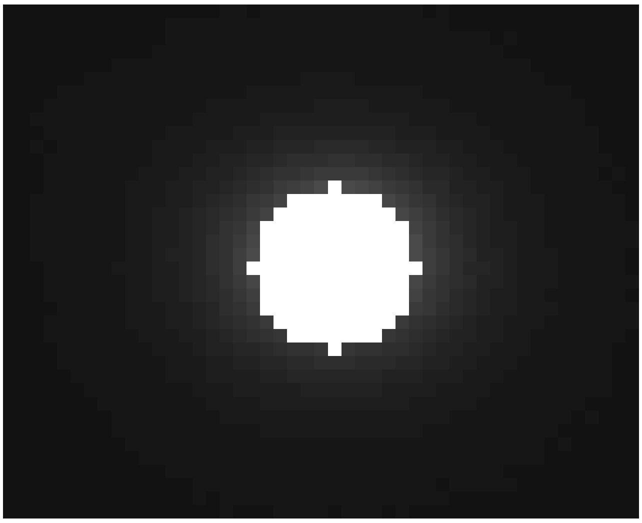

Fig. 5: 6 pixel aperture around (2260, 3300)

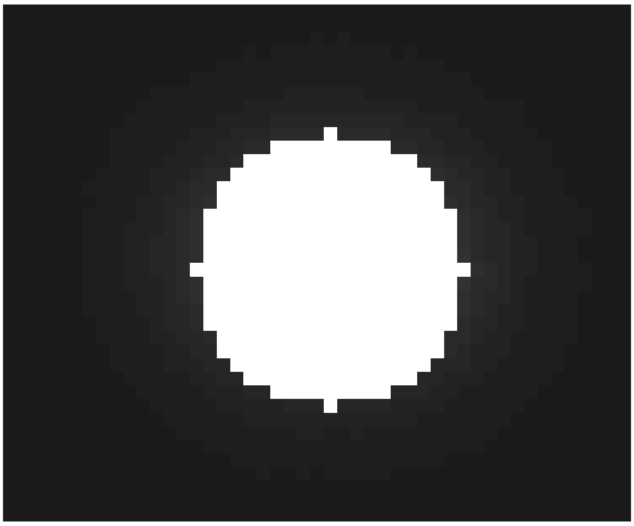

Fig. 6: 8-direction method around (2260, 3300)

However, this method suffered serious drawbacks. Figure 7 illustrates the problem – one galaxy is detected well, but the second galaxy is completely skipped. The line going south-west from the second galaxy soon hits the mask and stops, but the lines in the other directions keep going in order to detect the whole galaxy. Then, when it comes to processing, the program takes the aperture to be the longer directions, detects that some pixels then lie in the mask (i.e. they have already been counted), and so discounts the whole galaxy. Obviously this is not desirable, so another method was needed.

The final solution used was to consider the average counts of successive 1-pixel wide circular rings going outwards. If the ring hit a masked area, or if the average was greater than or within a 1-sigma tolerance of the average of the previous ring, it stopped going outwards. Figure 8 shows this ring method applied to the same object as fig. 7. It can be seen that the detection is much better – both galaxies are captured about as well as they could be with circular apertures.

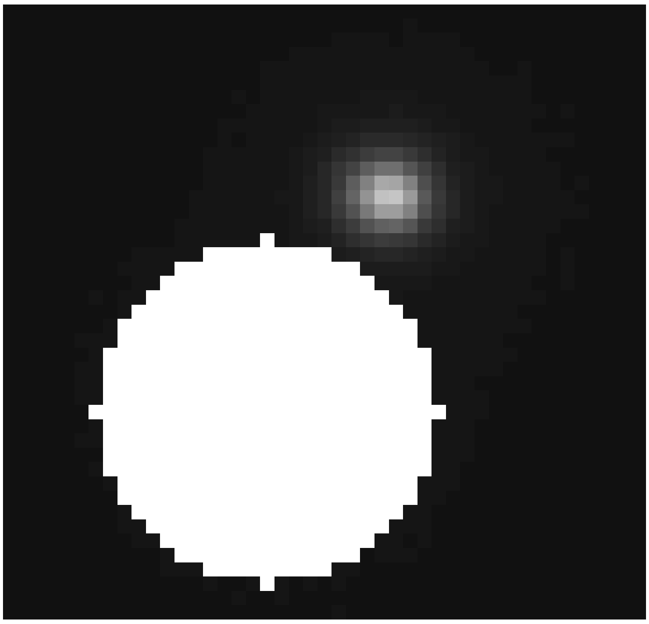

Fig. 7: 8-direction method around (30, 3600)

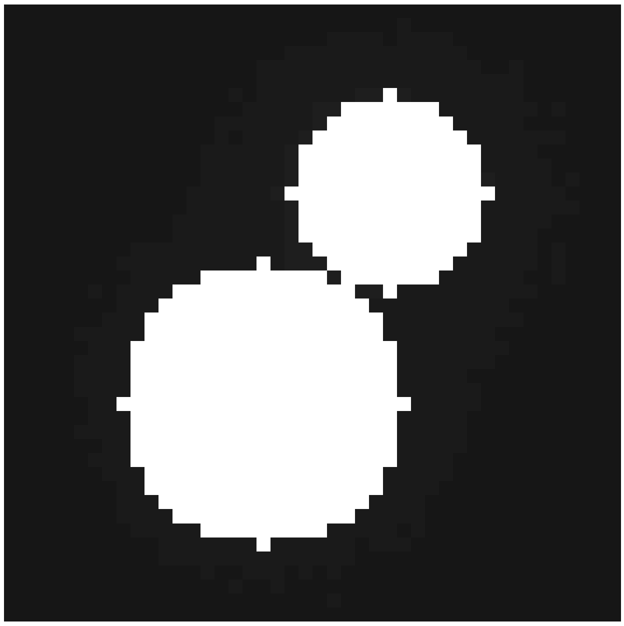

Fig. 8: Ring method around (30, 3600)

Once the radius is established, the processing is simple. The total count is found, and after subtracting the background count from the area of the circle, the count is converted into a magnitude via

$$m = ZP_{inst} - 2.5\log_{10}(counts) \quad (2)$$

where ZP_{inst} is the instrumental zero point. Together with the location, radius, total count, peak count, and error on the magnitude, this information is then stored into a catalogue. When complete, this is pickled into a file for later reference. The catalogue can be unpickled and displayed at any time thereafter – a sample is shown in figure 9.

Location	Radius	Tot Count	Max Count	Magnitude	Mag. Error
(657, 2530)	2	61813.0	65535.0	13.3223004452	0.0204712187202
(3443, 2562)	2	62028.0	65535.0	13.3185305545	0.0204696041277
(3600, 2535)	2	61939.0	65535.0	13.3200895264	0.0204702711497
(4548, 1054)	2	62043.0	65535.0	13.3182680267	0.0204694918945
(30, 1692)	16	3891928.0	42298.0	8.82458800676	0.020007570798
(1495, 634)	17	2228691.0	40045.0	9.42987535179	0.0200132188992
(1938, 675)	17	1866187.0	39374.0	9.62261210056	0.0200157856377
(2300, 446)	18	2407783.0	39199.0	9.34595664045	0.020012235972
(578, 1772)	17	2400895.0	37927.0	9.34906708207	0.0200122710654
(4273, 55)	19	1879742.0	37694.0	9.61475438703	0.0200156718506
(3923, 188)	16	1300921.0	37461.0	10.0143726892	0.0200226408107
(2263, 704)	16	1563948.0	37293.0	9.81444422744	0.0200188348381
(662, 65)	43	6508644.0	37237.0	8.26627370548	0.0200045274012

Fig. 9: Part of the catalogue printout

Results and discussion

From equation 1, a plot of $\log N$ against apparent magnitude should reveal a straight line, with a gradient of 0.6. The graph produced from our results is shown in figure 10. A straight line is visible, with deviations at higher and lower apparent magnitudes.

Error bars are obtained by propagating the error in counts per pixel, as follows:

Counts per pixel are a Poisson process, so $\sigma_{counts} = \sqrt{(counts)}$.

The error in the log of the counts is $\sigma_{\log(counts)} = \dfrac{1}{\ln 10}\dfrac{\sigma_{counts}}{counts}$.

So, from equation 2, $\sigma_{\log(N)} = \dfrac{\partial \log(N)}{\partial \log(counts)}\sigma_{\log(counts)} = 1.5\log(counts)$.

The error on the instrumental zero point was neglected as insignificant.

The central blue section has a straight line fit applied to it (the red line), with a gradient of 0.654±0.002, which is in reasonable agreement with the theoretical value of 0.6, especially considering our survey only had 1455 objects to work on, whereas other surveys have catalogued tens of thousands[2].

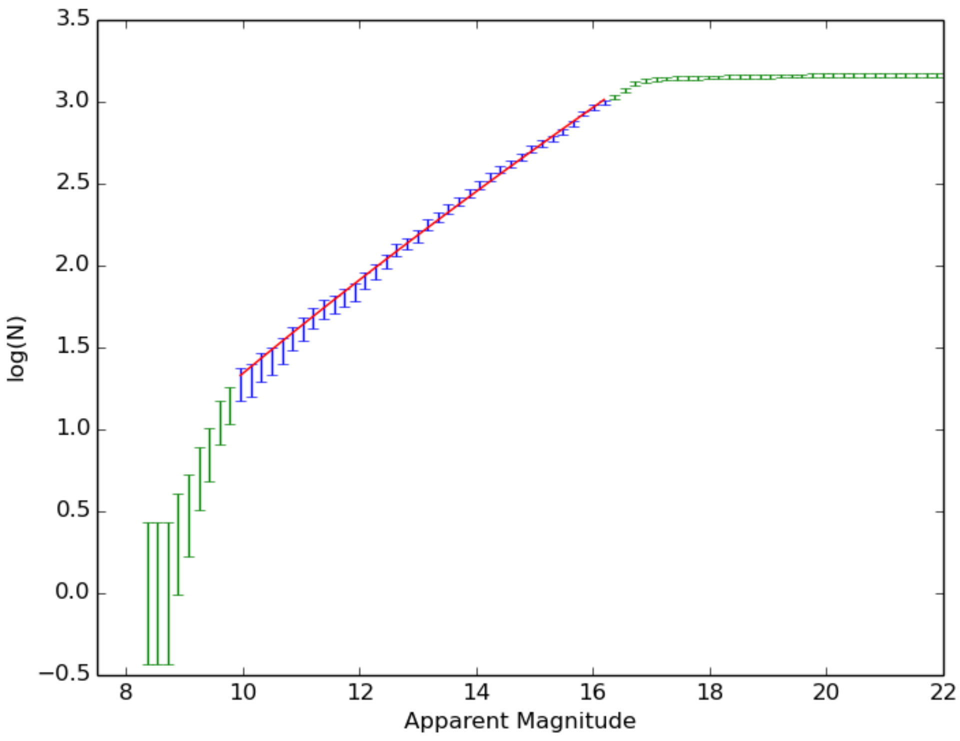

Fig. 10: Apparent magnitude of a galaxy against log of number of galaxies brighter than it

The deviation at higher apparent magnitudes is due to 'incompleteness' in our survey: that is, galaxies become much harder to detect as they become fainter. The first point of the roll-off (the second green section) occurs at an apparent magnitude of around 16.4, which following formula 2 corresponds to a count of around 3630. This is roughly the tolerance level of our image (3553 counts per pixel).

The deviation at lower apparent magnitudes however, is due to several factors. Firstly, foreground contamination. The survey was directed away from the galactic plane to avoid as many stars as possible, but some are still present, most obviously the large central one. Efforts were made to mask out all the stars, but there are likely still some remaining that contaminate the data. Another possible reason is our ring method for calculating the aperture misses some flux. A careful inspection of figure 8 reveals that the aperture is smaller than that of figure 7. Although this method is preferable to the 8-point one because it is able to detect both galaxies, it misses flux from each of them, which causes them to be fainter in the catalogue.

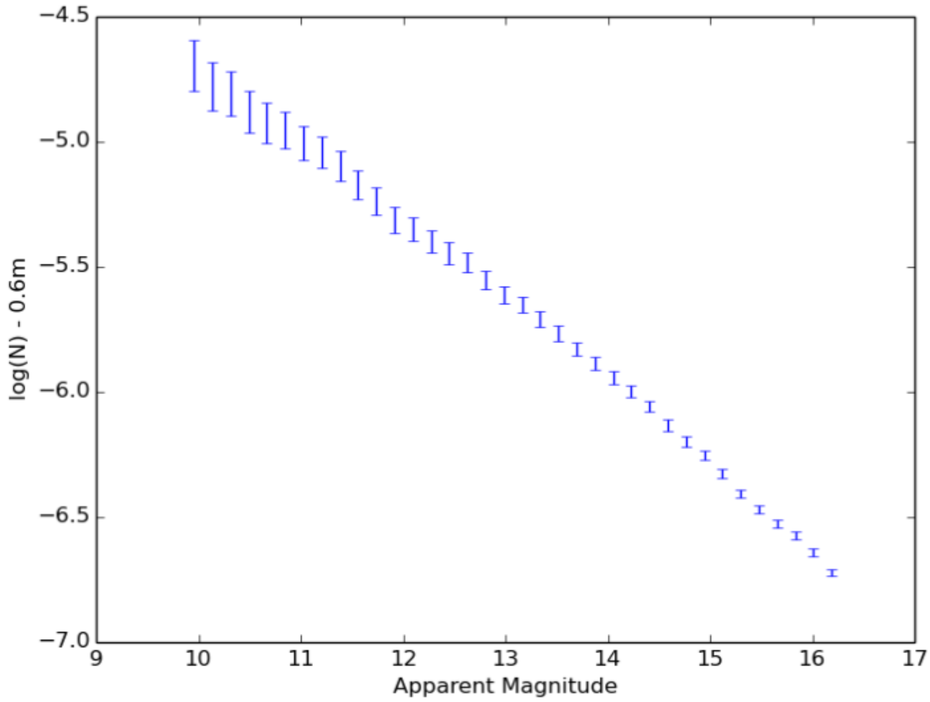

Fig. 10: Normalised plot of straight line in fig. 10

Our catalogue was also analysed for signs of galaxy evolution. The data was normalised by plotting $\log(N/N_0)$, where $\log N_0 \propto 0.6\,m$. This factors out the Euclidean dependence.

Figure 10 shows our normalised data. Figure 11 shows Tyson and Jarvis' normalised data against various models of evolution[3]. The first sign of evolution appears at an apparent magnitude of around 19. Hence, because our data only presented a straight line up until an apparent magnitude of 16 before reaching the detection limit, the survey was not sensitive enough to detect galaxy evolution in any case.

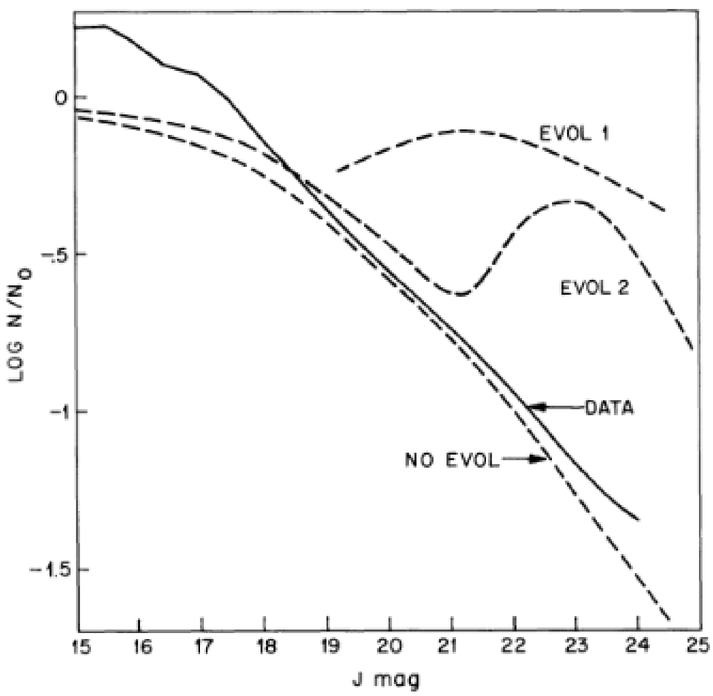

Fig. 11: Normalised data with evolution models

Conclusion

In this project, an image of the extragalactic sky was processed to produce a catalogue of galaxies, recording their brightness and other pertinent information. By comparing each galaxy to the whole population, it was shown that the apparent magnitudes closely followed the relationship $\log N(m) = 0.6\, m + const.$

Deviations were found at high and low apparent magnitudes – the former being due to the sensitivity of the survey, and the latter due to star contamination and missed flux. Once these deviations were ignored, a straight line was found with a gradient of 0.654 ± 0.002, a very good agreement to the expected value given the few galaxies identified compared to other surveys.

The deviations could have been eliminated, or at least improved, by a survey with a much higher sensitivity and a better method of picking out and masking stars. In addition, apertures of varying shapes could have been used to capture more flux from each galaxy.

Finally, our survey did not show any signs of galaxy evolution – however, according to certain models, evolution is expected to take place in younger galaxies. These are so far away and faint that they fall beneath the sensitivity level of our survey, and therefore it was not possible to draw any firm conclusions on this aspect.

Bibliography

[1]: D.L. Clements, N. Skrzypek (2013), *Deep Galaxy Survey : A 3rd Year Lab Module*, Astrophysics Group, Imperial College London

[2]: N. Yasuda, M. Fukugita, V.K. Narayanan et al. (2001), *Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data*, DOI: 10.1086/322093

[3]: J.A. Tyson, J.F. Jarvis (1979), *Evolution of galaxies - Automated faint object counts to 24th magnitude*, DOI: 10.1086/182982