Angular Clustering in the SDSS

Andy Connolly and István Szapudi

1 Introduction

To allow some discussion of the work we have been doing on the angular clustering with the early SDSS data we have put together a few notes on some of the tests we have undertaken. The goal of these analyses is to understand the systematics within the SDSS data and to show that our current data sets are already reaching the accuracy of existing surveys. All results are based on the compilation of data put together in December by Ryan Scranton and David Johnson from Runs 752 and 756 (using Photo 5.0.3). The star galaxy separation was derived directly from the photo pipeline classification scheme (Ryan and Dave have already produced a comparison of photo's and their classification). In the first section we show the results of the angular 2 pt function and in subsequent sections the higher order statistics.

2 Angular Correlations

2.1 Code Comparisons

One of the primary goals of the Fermilab meeting in December was to derive a set of independent measures of the 2pt function in order to cross validate our codes. As part of this we took 5 independent codes realizing 3 different estimators for the two-point (and higher order) correlation functions. For reference, the three estimators are the Landy and Szalay estimator (LS)

$$w_{LS} = \frac{DD - 2DR + RR}{RR}, \tag{1}$$

it's modified version (SS) adapted to a grid for calculating window (W) correlation functions

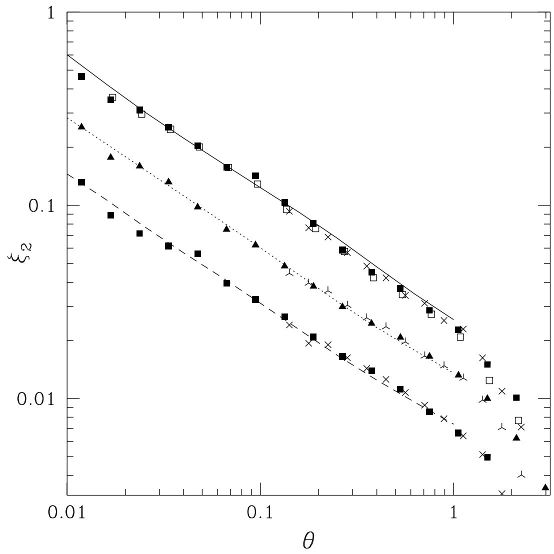
$$w_{SS} = \frac{DD - 2DW + WW}{WW}, \tag{2}$$

and finally counts-in-cells estimator (CIC)

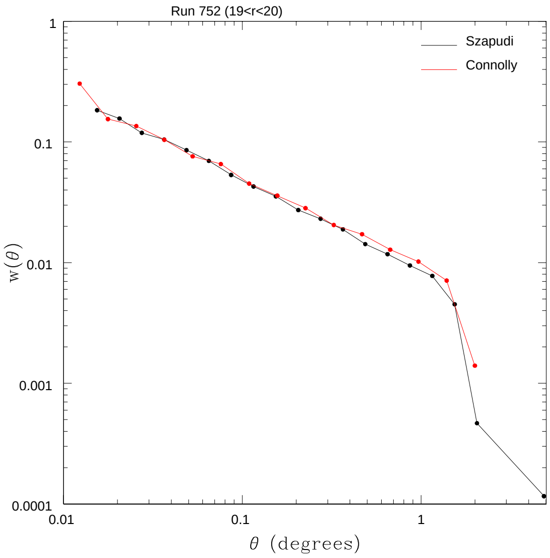
$$w_{CIC} = B\left(\frac{F_2}{\langle N\rangle^2} - 1\right), \tag{3}$$

where F_2 is the second factorial moment, $\langle N\rangle$ is the average count in a cell, and $B = 1.89$ from Monte Carlo integrating the power law correlation function on a unit square.

Two independent realizations of the LS estimator were provided by AJC, while one realization of each LS, SS, and CIC estimators were implemented by IS. Each of the AJC and IS sets of code were found to be internally consistent. In the following figure lines correspond to the CIC, filled symbols the LS (IS), open squares to the LS (AJC), and starred symbols to the SS estimator. The three different estimators are fully consistant with each other, three magnitude slices of width one are shown between $r = 17, 18, 19, 20$.



Below we show the comparision of the LS (AJC) and the SS (IS) estimators on scales of 0.01 degrees through 5 degrees. The data correspond to a single strip (Run 752) selected over the magnitude range $19 < r < 20$ after applying the extinction corrections derived from Schlegel et al. The two estimators are consistent within the errors (assuming Poisson statistics) across the full range of angular scales. For the rest of this write-up we use the SS estimator for speed considerations, and the CIC estimator, which in addition calculates the amplitudes of the higher order correlation function.



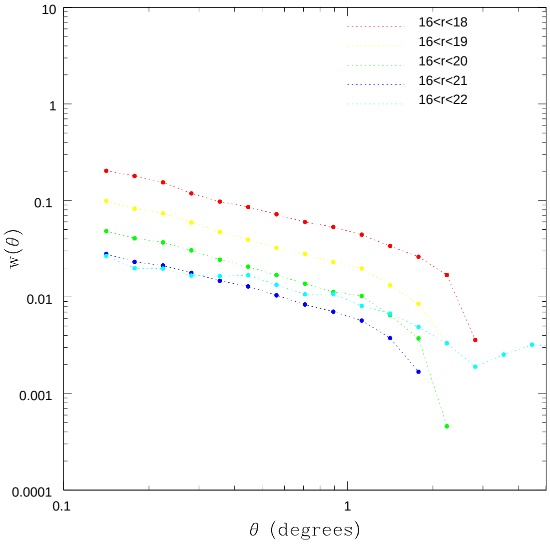
Besides estimating the two-point correlation function we have characterized clustering through moments of counts in cells statistics. The same infinitely oversampling CIC code calculated not only the two-point function, but the normalized amplitudes of the higher order correlation functions

$$S_N = \frac{\langle \xi_N \rangle}{\langle \xi \rangle^{N-1}}. \tag{4}$$

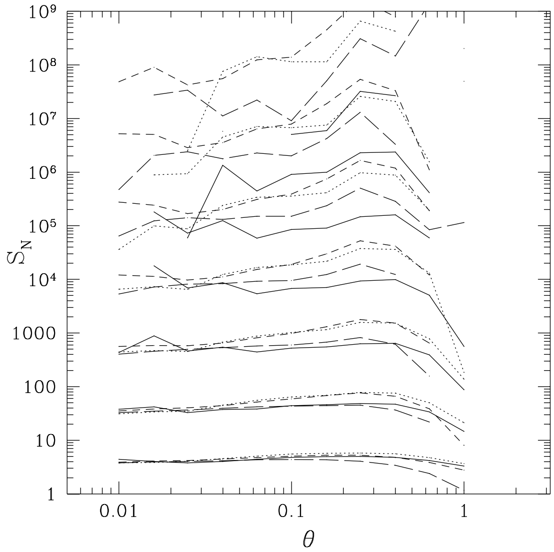
While the two-point component was tested as a part of this project, the higher order part was previously compared with at least two codes based on different principles by IS, as well as with two independent codes. This is undoubtedly the best existing CIC code for two dimensional problems.

2.2 Clustering as a Function of Magnitude

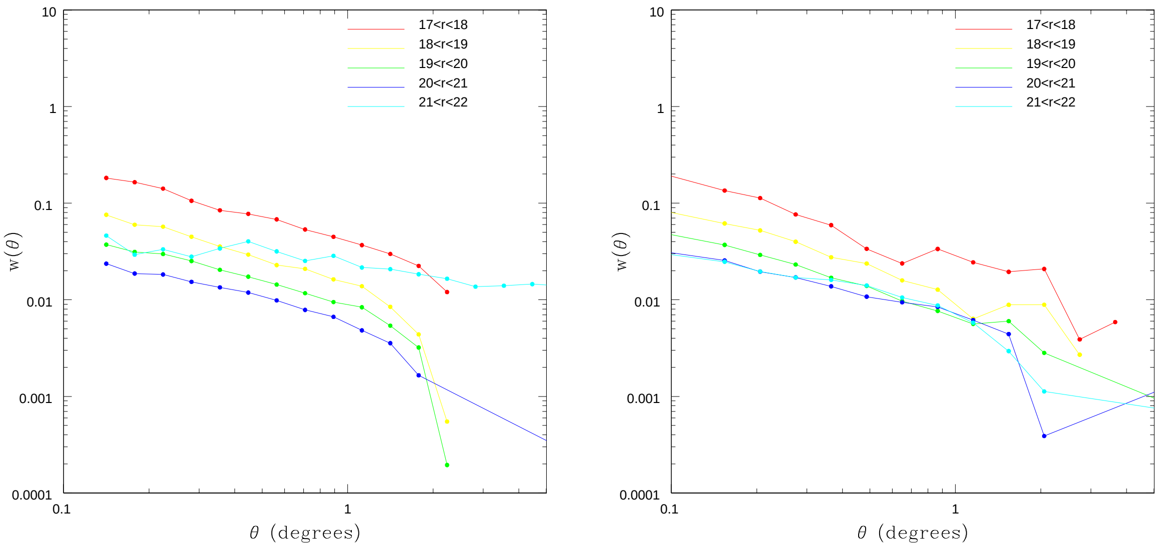
Given a consistent set of codes for measuring the 2pt function we consider the scaling of the correlation function with limiting magnitude. The figure below shows the correlation function for the combined 756 and 752 data sets as a function of limiting magnitude. The amplitude of the clustering (and the angular scales on which the break occurs) scale with limiting magnitude (consistent with previous surveys). At $r < 22$ we find more power on scales larger than 0.3 degrees than expected from the previous scaling. The shape of this excess power is consistent with that observed from the 2pt function of stars within the survey field.



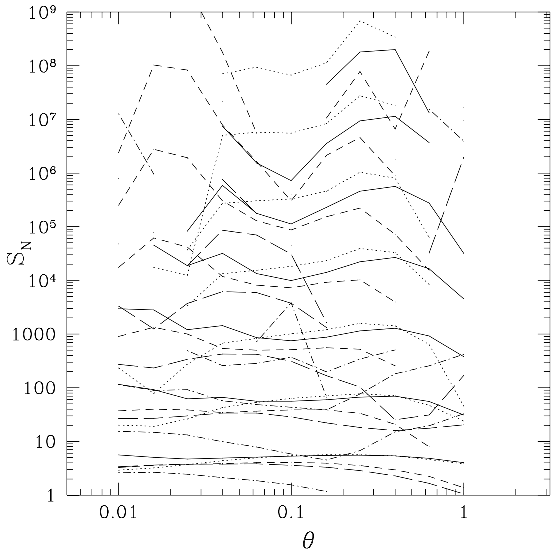
The corresponding higher order correlation functions up to $n = 9$ are shown on the next figure. Solid, dotted, dash, long dash lines correspond to upper limits of $r = 18, 19, 20, 21$. The measurements are comparable to that of the EDSGC; it appears that up to the sixth point correlation function the results are quite accurate. The approximate hierarchy is encouraging, although it would be too early to draw strong conclusions of the data, and no scientific analysis is attempted here.



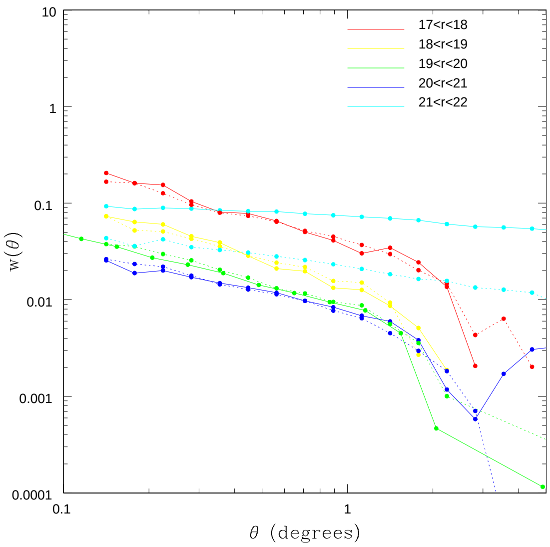
To isolate the effect on large scales we separate the data into intervals in magnitude (with a width of 1 magnitude). The scaling that we saw previously is repeated for these data and the excess power is again identified in the $21 < r < 22$ data set (see left figure below). Again the shape of the correlation function is consistent with the shape of the stellar 2pt function and the excess power appears consistent with stellar contamination within the complete dataset. The density of stars within the Runs 752 and 756 increases dramatically as the scan moves to low Galactic latitude. To remove this effect we isolate 1/4 of the survey region selected to be at high Galactic Latitude. The correlation function as a function of magnitude is given below on the right hand side of the figure. The excess power is no longer seen from this subset of the data (the correlation function is noticeably noisier). It is clear, therefore, that if we can better quantify the stellar contamination as a function of Galactic Latitude or if we work with high Galactic latitude data, the clustering signal can be traced to $r < 22$.



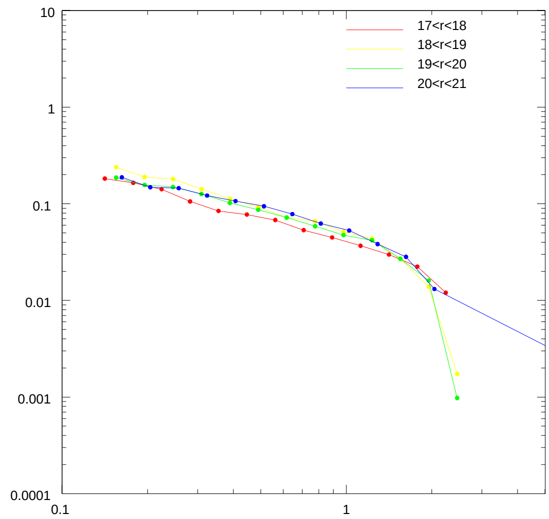
Again the same investigation can be repeated for higher order correlation functions. The scatter is larger compared to the previous plot, as expected. Linetypes correspond to the upper limit of the magnitude slices with width of 1 as on the previous figure concerning S_N's, and dot-dashes mean $21 < r < 22$. Note the strong depletion of the dot-dashed lines, which is likely to be indicative of strong stellar contamination. It is worth to note that higher order correlation functions tend to amplify systematic errors. At this point it is too early to state, whether the dip around 0.1 degrees is due to systematics or simply cosmic variance, nevertheless it appears to be present at all magnitude slice.



We consider the Runs 752 and 756 independently. The solid line shows the clustering for Run 752 and the dotted line for Run 756. For magnitudes $r < 21$ and the full catalog range the two runs give consistent clustering signal on scales out to 2 degrees. The excess power for $r > 21$ is present in both runs with a larger amplitude for the 752 data. This is consistent with the fact that Run 752 was observed in worse seeing conditions.

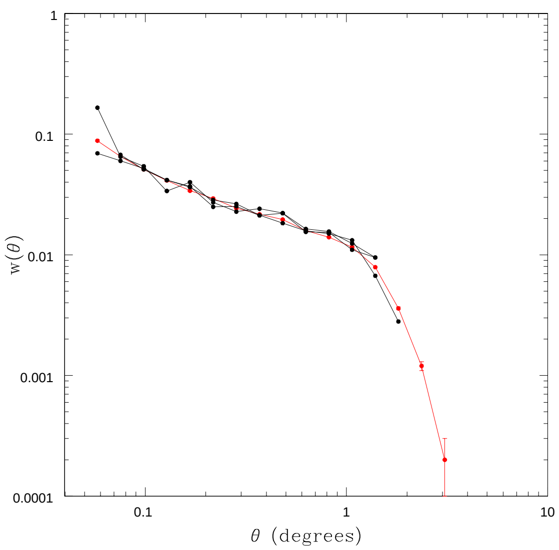


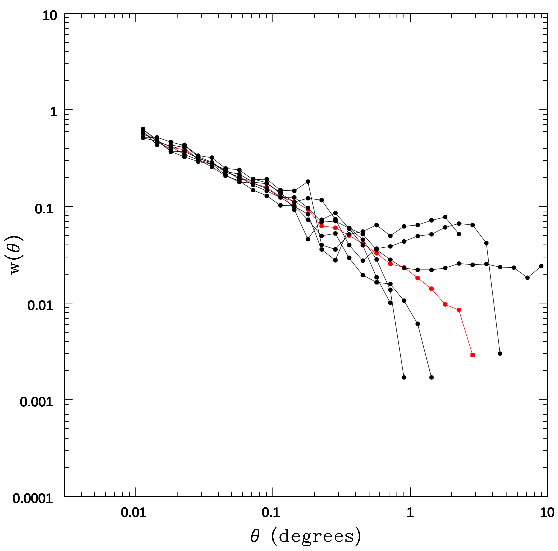
We can determine whether the scaling we see as a function of magnitude is consistent with the redshift distributions of each magnitude interval. By assuming an N(z) for each of the magnitude bins (taken from the parameterization given by Hudon and Lilly) we can use Limbers equation to rescale the amplitude and angular scale of the correlation function to a common magnitude range. The calculations were done assuming that the clustering remains static in proper coordinates and for an Omega=1 cosmology. The scaling is somewhat insensitive to these two parameters. Using the derived scalings we renormalize the angular scale and the amplitude of the correlation functions to that given by the $17 < r < 18$ data set. Given this somewhat crude approach the clustering signal is found to be consistent for all of the tested magnitude intervals (with a slight offset for the $17 < r < 19$ data set).



2.3 Angular Dependent Clustering

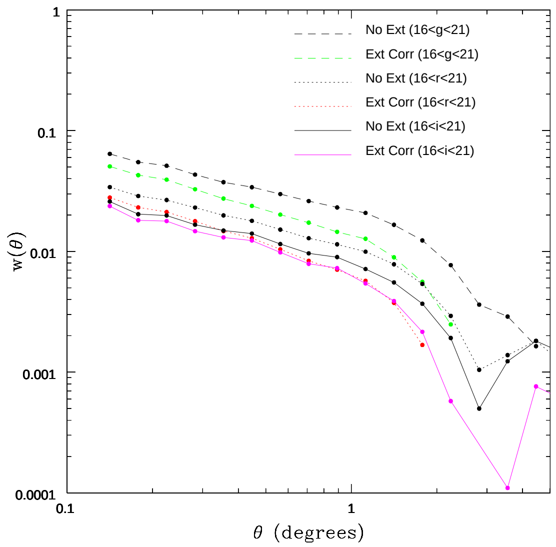
As shown at the Princeton and Fermilab meetings the azimuthal dependence (e.g. along or orthogonal to the scan direction) of the clustering signal is a sensitive diagnostic of systematics within the imaging stripes (and from stripe to stripe). The figures below show the 2pt function separated into azimuthal bins (each of width 30 degrees). The right hand figure shows the clustering as derived from the photo 4 reductions. There is good agreement between the azimuthal bins for scales less than 0.1 degrees. On scales greater than 0.1 the dispersion in the correlation function is marked and is indicative of the variation in the psf as a function of time and position on the focal plane of the camera. The inclusion of a variable psf into the photo 5 reductions (left figure) produces a marked improvement in the dispersion between the azimuthal bins (out to 1 degree).



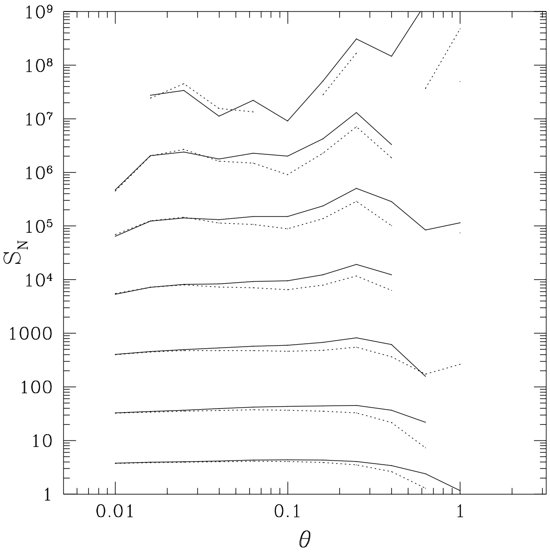


2.4 Extinction Dependent Clustering

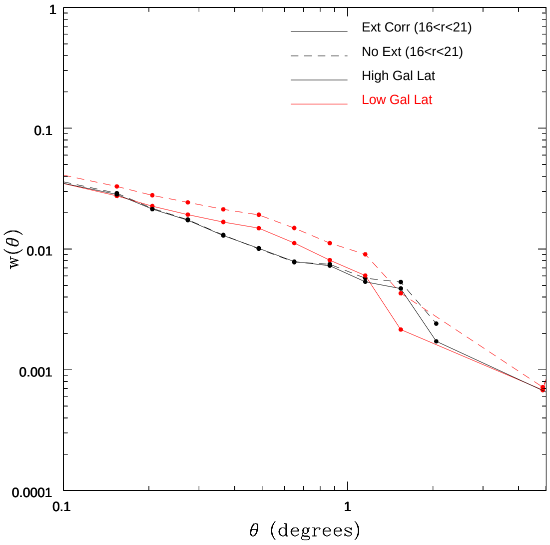
We consider the effect of extinction on the clustering signal using the full data set (where we cover regions of high extinction). The figure below shows the clustering in the 756 and 752 data using magnitudes with and without extinction corrections. These plots are shown for the g, r and i passbands. For each filter the extinction correction reduces the observed clustering signal on all scales (with the amplitude of the correction increasing with angular scale). The amplitude of this correction is comparable for the g and r filters and about a factor of 2 smaller in the i band (on the log plot). Given the g, r and i number counts and the extinction curve of our Galaxy we should be able to model this effect and thereby estimate the spectrum of the extinction due to dust.



Extinction has a clearly detectable effect on the higher order correlation functions as well, as the next figure shows. Solid lines display the higher order correlation functions for $16 < r < 21$ with extinction correction included, while the dotted line shows the raw results without correction. The difference can be as large as a factor of 2. Nevertheless, the feature at around 0.1 previously mentioned, is certainly not due to extinction, as the dotted lines follow the solid ones fairly closely, opening up towards larger scales. At first sight perhaps it is counter-intuitive, that the S_N's are depleted if no correction is applied, while the two-point correlation function has excess power. This is due to the normalization of the S_N's with $\langle\xi\rangle^{N-1}$, and perhaps indicates the extinction acts most strongly on the two-point properties.



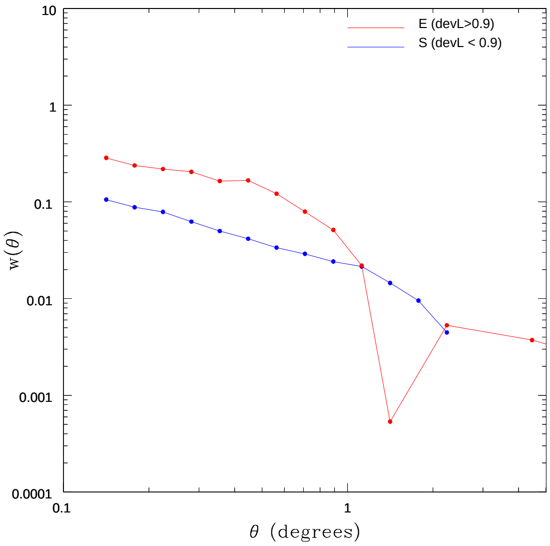
By excluding low Galactic latitude regions we can reduce the effect of extinction (and stellar contamination). The plot below shows 2 quartiles selected from the 752 and 756 data. The black line is the high Galactic Latitude slice and the red line the low Galactic latitude slice (solid lines correspond to extinction corrected data and dashed lines to the raw data). At high Galactic latitude the extinction and no extinction corrections are consistent (as we would expect from the relative absence of dust). At low Galactic latitude the extinction correction reduces the clustering signal to a value comparable to that found for the high Galactic regions. While the overall effect of the correction clearly goes in the right sense (ie we move closer to the correlation function found at high latitudes) it does not appear to completely remove the effects of the modulation of the galaxy density due to dust. It is clear that to extend to lower latitudes we will have to address this issue is some detail (and that the SDSS data provide a great resource for studying the spectrum and clustering of dust).



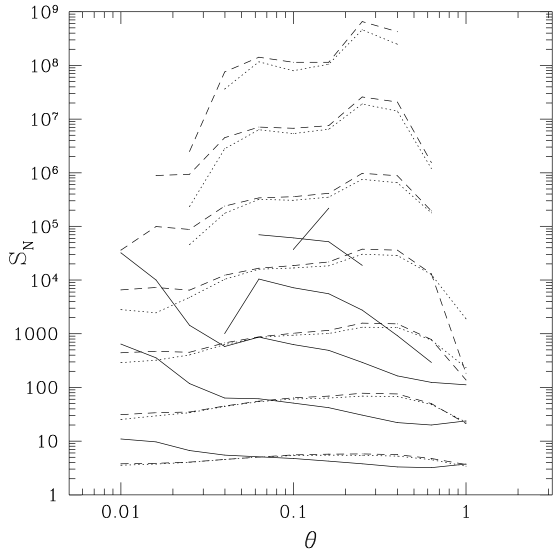
2.5 Clustering as a Function of Morphological Type

Finally we consider the type dependent clustering (to determine if the model fits to the galaxy profiles correspond to meaningful classifications). We use the likelihood values derived from the photo pipeline for the star, exponential and deVaucouleurs profiles. We initially take a conservative cut, selecting those galaxies with $16 < r < 19$ and separating into an elliptical subset based on the r band devL parameter

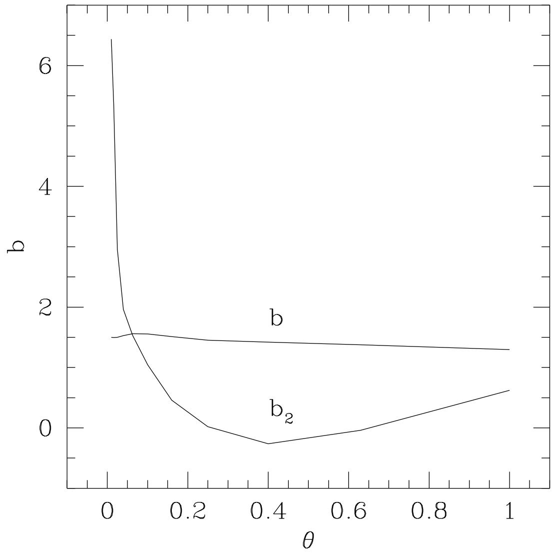
($devL > 0.9$). The correlation function for the "elliptical" galaxies is shown below as a red line and for the "spiral" galaxies by a blue line. On scales smaller than 1 degree the clustering amplitude of the ellipticals is approximately 3 times larger than for spirals (consistent with local observations). On scales greater than 1 degree the clustering signal drops rapidly.



Morphological dependence of higher order clustering is especially interesting, since the SDSS could deliver the first reliable data on this subject. The next figure shows the S_N's for "spiral" and "elliptical" galaxies, in the sense of the previous paragraph, with dotted and solid lines respectively in our reference $16 < r < 19$ slice. The total clustering apparently is dominated by spirals, while "ellipticals" have stroner clustering on small scales (as perhaps intuitively expected), break the clustring hierarchy with an increasingly rising scale dependence towards small scales, and they perhaps display a suppression of correlation our scales between $0.1 < \theta < 1$.



The previous figure is clearly indicative of non-linear scale-dependent bias. While our main focus in this draft was an extensive examination of the systematics, we could not resist to add the following figure. If the density of ellipicals is Taylor expanded as a function of the density of spirals, $\delta_E = b\delta_S + b_2^2\delta_S^2/2 + \ldots$, the nonlinear bias coefficents b, and b_2 arise. The next figure shows that while the (usual) linear bias is almost constant, the first non-trivial non-linear coefficient exhibits a sharp rise on small scales. While the rise of b_2 on larger scales after it "decayed" to zero might be suspicious and should be subject of further confirmation, it is hard to imagine that the rise on small scales could disappear in these conservative magnitude limits, even if many parameters, among others the definition of "ellipticals" might be altered in future version of the survey and the corresponding analyses software.



3 Conclusion

By paying careful attention to the systematics (mainly star galaxy separation and the effect of dust) the early SDSS data clearly provide a beautiful data set from which to measure the 2pt and higher order angular clustering statistics. Our current results are consistent with existing surveys (e.g. APM and EDGSC) but with a much higher photometric accuracy. We would welcome a discussion at the LSS working group regarding these results and the next steps we need to take.