LARGE-SCALE STRUCTURES IN THE EARLY SLOAN DIGITAL SKY SURVEY: COMPARISON OF THE NORTH AND SOUTH GALACTIC STRIPS

ENRIQUE GAZTAÑAGA1,2

Received 2001 October 9; accepted 2002 July 22

ABSTRACT

We compare the large-scale galaxy clustering between the north and south Sloan Digital Sky Survey (SDSS) early data release (EDR) and also with the clustering in the Automatic Plate Measuring Facility (APM) Galaxy Survey. The three samples are independent and cover an area of 150, 230, and 4300 deg$^2$, respectively. We combine SDSS data in different ways to approach the APM selection. Given the good photometric calibration of the SDSS data and the very good match of its north and south number counts, we combine them in a single sample. The joint clustering is compared with equivalent subsamples in the APM. The formal sampling errors are small enough to provide an independent test for some of the results in the APM. We find evidence for an inflection in the shape of the two-point function in the SDSS that is very similar to what is found in the APM. This feature has been interpreted as evidence for nonlinear gravitational growth. By studying higher-order correlations, we can also confirm good agreement with the hypothesis of Gaussian initial conditions (and small biasing) for the structure traced by the large-scale SDSS galaxy distribution.

Subject headings: cosmology: observations — galaxies: clusters: general — large-scale structure of universe

On-line material: color figures

1 INTRODUCTION

The Sloan Digital Sky Survey (SDSS) collaboration has recently made an early data release (EDR) publicly available. The EDR contains around a million galaxies distributed within a narrow strip of 2.5 arcmin across the equator. As the strip crosses the Galactic plane, the data is divided into two separate sets in the north and south Galactic caps.

The SDSS collaboration has presented a series of analyses (Zehavi et al. 2002; Scranton et al. 2002; Connolly et al. 2002; Dodelson et al. 2001; Tegmark et al. 2002) of large-scale angular clustering on the north Galactic strip, which contains data with the best seeing conditions in the EDR. Gaztañaga (2002, hereafter Ga02) presented a study of bright ($g < 20$) SDSS galaxies in the south Galactic EDR strip, centering the analysis on the comparison of clustering to the Automatic Plate Measuring (APM) Facility’s Galaxy Survey (Maddox et al. 1990).

In this paper, we combine the bright ($r' < 19$ or $g' < 20$) galaxies in the north and south strips to make a detailed comparison between north and south and also to the APM survey. Do the north and south strips have similar clustering? How do they compare to previous analyses? What does the EDR tell us about structure formation in the universe? Answering these questions will help us to understand the SDSS EDR data and, at the same time, will give us the opportunity to test the reliability of conclusions drawn from previous galaxy surveys, in particular, regarding the shape of the two-point function (Maddox et al. 1990; Gaztañaga & Juszkiewicz 2001) and higher-order correlations (e.g., Bernard et al. 2002, and references therein).

This paper is organized as follows. In § 2, we present the samples used and the galaxy selection and number counts. Section 3 shows the comparison of the two- and three-point correlation functions. We end with some discussion and a listing of conclusions.

2 SDSS SAMPLES AND PIXEL MAPS

We follow the steps described in Ga02. We download data from the SDSS public archives using the SDSS Science Archive Query Tool.3 We select objects from an equatorial SGC (south Galactic cap) strip 2.5$^\circ$ wide ($-1.2^\circ < \text{decl.} < 1.2^\circ$) and 66$^\circ$ long ($351^\circ < \text{RA.} < 56^\circ$), which will be called EDR/S, and also from a similar NGC (north Galactic cap) 2.5$^\circ$ wide and 91$^\circ$ long ($145^\circ < \text{RA.} < 236^\circ$), which will be called EDR/N. These strips (SDSS numbers 82N/82S and 10N/10S) correspond to some of the first runs of the early commissioning data (runs 94/125 and 752/756) and have variable seeing conditions. Runs 752 and 125 are the worst, with regions where the seeing fluctuates above 2$''$. Runs 756 and 94 are better but still have seeing fluctuations of a few tenths of an arcsecond within scales of a few degrees.4 These seeing conditions could introduce large-scale gradients because of the corresponding variations in the photometric reduction (e.g., star-galaxy separation) that could manifest as large-scale number density gradients (see Scranton et al. 2002 for a detailed account of these effects). We will test our results against the possible effects of seeing variations by restricting the analysis to runs 756 and 94 and by using a seeing mask (see § 3.3).

We will also consider a sample that includes both the north and south strips, which we call EDR/(N+S). Note that the clustering from this sample will not necessarily agree with the mean of EDR/N and EDR/S, e.g., EDR/(N+S) ≠ EDR/N + EDR/S (see below).

1 INAOE, Astrofisica, Tonantzintla, Apdo Postal 216 y 51, Puebla 7200, Mexico.
2 IEEC/CSIC, Edif. Nexus-104-c/Gran Capita 2-4, 08034 Barcelona, Spain.
3 Go to http://archive.stsci.edu/SDSS/software/.
4 See http://www-sdss.fnal.gov:8000/skent/seeingStatus.html or Fig. 4 in Scranton et al. 2001.
We first select all galaxies brighter than $u' = 22.3$, $g' = 23.3$, $r' = 23.1$, $i' = 22.3$, and $z' = 20.8$, which correspond to the SDSS limiting magnitudes for 5σ detection in point sources (York et al. 2000). Galaxies either are found from the $1 \times 1, 2 \times 2, or 4 \times 4$ binned CCD pixels, and they are deblended by the SDSS pipeline (Lupton et al. 2001). Only isolated objects, child objects (resulting from deblending), and objects on which the deblender gave up are used in constructing our galaxy catalog (see Yasuda et al. 2001).

There are about 375,000 objects in our sample classified as galaxies in the EDR/S and about 504,000 in the EDR/N. Figure 1 shows the number counts (surface density) for all these 879,000 galaxies as a function of the magnitude in each band, measured by the SDSS modified Petrosian magnitudes $m_u$, $m_g$, $m_r$, $m_i$, and $m_z$ (see Yasuda et al. 2001 for a discussion of the SDSS counts). Continuous diagonal lines show the 10^{0.6m} expected for a low-redshift homogeneous distribution with no k-correction, no evolution, and no extinction.

We next select galaxies with SDSS modified Petrosian magnitudes to match the APM selection $17 < B_1 < 20$, which corresponds to a mean depth of $D \sim 400$ h^{-1} Mpc. We try different prescriptions. We first apply the following transformation to mimic the APM filter $B_1$:

$$B_1 = g' + 0.193(g' - r') + 0.115 .$$

This results from combining the relation $B_1 = B - 0.28(B - V)$ (Maddox et al. 1990) with expressions (5) and (6) in Yasuda et al. (2001). As the mean color $g' - r' \approx 0.7$ the above relation gives a mean $B_1 \approx g' + 0.25$, which roughly agrees with the magnitude shift used in Ga02. For the $17 < B_1 < 20$ range (using the above transformation), we find $N \approx 123,000$ galaxies in the EDR/(N+S), with a galaxy surface which is very similar to the one in the APM (only 5% larger after subtraction of the 5% star-merger contribution in the APM). In any case, these types of color transformations between bands are not accurate, and they work only in some average statistical sense. The uncertainties are even larger when we recall that the APM uses a fixed isophotal aperture, while the SDSS uses Petrosian magnitudes—a difference that can introduce additional color terms and surface brightness dependence.

It is much cleaner to use a single SDSS band. We should use $g'$, which is the closest to the APM $\lambda_B = 4200$ Å ($\lambda_g \approx 3560, \lambda_r \approx 4680, \lambda_i \approx 6180$ Å). But how do we decide the range of $g'$ to match the APM $17 < B_1 < 20$? We try two approaches. One is to look for the magnitude interval that has the same counts, as done in Ga02. The resulting range is $16.8 < g' < 19.8$. This gives a reasonable match to the clustering amplitudes in the EDR/S and EDR/N. But there is no reason to expect a perfect match: the selection function and resulting depth is different for different colors. The other approach is to fix the magnitude range, i.e., $17 < g' < 20$, rather than the counts. This produces $N \approx 157,000$ galaxies, which corresponds to ±25% higher counts than the APM. This does not necessarily mean that this sample is deeper than the APM because of the intrinsic difference in color selection, k-correction, and possible color evolution.

Finally, we produce equal-area projection pixel maps of various resolutions similar to those made in Ga02 (Fig. 2). Except for a few tests, all the analyses presented here correspond to 0.75 resolution pixels. On making the pixel maps, we mask out about 1/3 of the EDR sample from the edges, which makes an integer number of pixels in our equatorial projection. This also avoids the potential problem of the galaxy photometry on the edges (although higher resolution maps show very similar results, indicating that this is not really a problem).
\[ A_b = C (\csc b - 1) \] with \( C = 0.1 \) was used for the APM photometry. This is a very small correction, \( A_b = 0 \) at the poles (\( b = 90^\circ \)) and the maximum \( A_b \approx 0.03 \) at the lowest Galactic declination (\( b \approx 50^\circ \)). This is in contrast to the Schlegel et al. (1998) extinction maps which have significant differential extinction \( E(B-V) \approx 0.02-0.03 \), even at the poles. The corresponding total absorption \( A_b \) for the \( B_J \) band according to Table 6 in Schlegel et al. (1998) is 4 times larger: \( A_b \approx 0.08-0.12 \). This increases up to \( A_b \approx 0.2-0.3 \) at Galactic declination \( b \approx 50^\circ \). Thus, the Schlegel et al. (1998) extinction correction has a large impact on the number counts for a fixed magnitude range. The change can be roughly accounted for by shifting the mean magnitude ranges by the mean extinction, e.g., \( \approx 0.2 \) mag in \( B_J \). It is therefore important to know what extinction correction has been applied when comparing different surveys or magnitude bands.

Despite the possible impact on the quoted magnitudes (and, therefore, on the counts), extinction has little impact on clustering, at least for \( r' < 21 \) (see Scranton et al. 2002; Tegmark et al. 1998). This is fortunate because of the uncertainties involved in making the extinction maps and its calibration. Moreover, the Schlegel et al. (1998) extinction map only has a 6'/1 FWHM, which is much larger than the individual galaxies we are interested in. Many dusty regions have a filamentary structure (with a fractal pattern) and large fluctuations in extinction from point to point. One would expect similar fluctuations on smaller (galaxy-size) scales, which introduces further uncertainties to individual corrections.

Here we decided as a default not to correct for extinction, because this will be closer to the APM analysis and has little effect on clustering at the depths and for the issues that will be explored here. This has been extensively checked for EDR/N by Scranton et al. (2002). See § 3.3 for checks in EDR/N, EDR/S, and EDR/(N+S).

To avoid confusion with other prescriptions by the SDSS collaboration, we will use \( z', r', u' \), and \( g' \) for raw, uncorrected magnitudes and \( z^*, r^*, g^*, \) and \( u^* \) for extinction-corrected magnitudes. For example, according to Schlegel et al. (1998) \( r' = 18 \) corresponds roughly to an average extinction-corrected \( r^* \approx 17.9 \) for a mean differential extinction \( E(B-V) \approx 0.03 \).

3. CLUSTERING COMPARISON

To study sampling and estimation-biasing effects on the SDSS clustering estimators, we have cut different SDSS-like strips out of the APM map (see Ga02). For the APM, we have considered a \( 17 < B_J < 20 \) magnitude slice in an equal-area projection pixel map with a resolution of 3.5 that covers over 4300 deg\(^2\) around the SGC. The APM sample can fit about 25 strips similar to EDR/S and 16 similar to EDR/N. The APM cannot cover the combined EDR/(N+S) as it extends across the whole equatorial circle. But we can select several subsamples consisting of sets of two strips, one like EDR/S and another one like EDR/N, well separated within the APM map by at least 10°. As correlations are negligible on angular scales greater than 10°, this simulates well the combined EDR/(N+S) analysis. To study sampling effects over individual scans, we also extract individual SDSS-like CCD scans out of the APM pixel maps. In all cases, we correct the clustering in the APM maps for a 5% contamination of randomly merged stars (Maddox et al. 1990); i.e., we scale fluctuations up by 5% (Gaztañaga 1994).

3.1. The Angular Two-Point Function

We first study the angular two-point function. Figure 3 shows the results from the EDR/S (long dashed lines), EDR/N (short dashed lines), and EDR/(N+S) (solid lines). As mentioned above, the clustering from the combined sample EDR/(N+S) will not necessarily agree with the mean of EDR/N and EDR/S (dotted lines) for several reasons: neither estimators nor sampling errors are linear, and local...

---

**Fig. 3.** Angular two-point function \( w_2(\theta) \) as a function of galaxy separation \( \theta \) for different SDSS magnitudes as labeled in the figures. Short and long dashed lines: SDSS EDR/N (north) and EDR/S (south) strips. Solid line: EDR/(N+S), a joint analysis of the SDSS south and north data. Dotted line: Mean of the EDR/N and EDR/S. Triangles with error bars: Mean and 1 \( \sigma \) confidence level in the values of 10 APM subsamples that simulate the different EDR samples: EDR/S (errors in the first panel), EDR/(N+S) (errors in the second and fourth panels), and EDR/S (errors in the third panel). [See the electronic edition of the Journal for a color version of this figure.]
galaxy fluctuations are estimated around the combined mean density (rather than the mean density in each subsample). As shown in Figure 3, the two estimators yield different results. In general, for a well-calibrated survey, the EDR/(N+S) should give better results than the sum of the parts, so we take the EDR(N+S) results as our best estimate.

In general, the results for the $w_2$ shape in EDR/N in Figure 3 agree well with the corresponding comparison in Figure 1 of Connolly et al. (2001), with a sharp break to zero around $2^\circ-3^\circ$. The results for EDR/S agree well with Ga02, showing a flattening at similar scales. Note how EDR/(N+S) $17 < B_J < 20$ (shown in the first panel) is about 15% higher in amplitude that the APM. This is not a very significant discrepancy for the EDR/S errors shown in the plot, but it is when compared to the EDR/(N+S) errors from the APM, shown in panels 2 and 3. As mentioned above, this is not totally surprising, as the magnitude conversion in equation (1) could work only in some average sense. Results for $16.8 < g' < 19.8$ are intermediate between $B_J$ and $17 < g' < 20$.

Scranton et al. (2002) studied the SDSS systematic effects with $r^*$ colors and found that systematic effects had negligible contributions to $w_2(\theta)$ for $r^* < 21$ (see their Fig. 15). The APM has a depth corresponding to $r^* \approx 18.5$, which is almost 3 mag brighter than the above limit. Nevertheless, for comparison, we also study the $w_2(\theta)$ shape in $r'$. The brighter sample of $r^* = 18-19$ in Connolly et al. (2001) is slightly deeper than the APM, with $z \approx 0.18$ rather than $z \approx 0.15$. We find that $r' = 17.8-18.8$ is the closest one-magnitude $r'$ bin in depth to the APM. Because of the average extinction, this corresponds roughly to extinction-corrected $r^* = 17.65-18.65$. This sample has about 40% fewer galaxies (per deg$^{-2}$) than the APM, presumably because of the color correction and differences in the photometric selection. Results for this $r^*$ sample (Fig. 3, third panel) agree quite well with the APM amplitude. Here the errors are from APM subsamples similar to EDR/N. Note how they are slightly smaller than the errors in the APM shown in the first panel of Figure 3, as expected from the smaller area of EDR/S.

Overall, we see how the shape of the two-point function in all EDR samples remains remarkably similar for the different magnitudes. This is despite the fact that the mean counts change by more than 60% from case to case. The amplitude of $w_2$ changes by about 20% from one SDSS sample to the other, but the shape remains quite similar. The best match to the APM amplitude is for the $17 < g' < 20$, which will be taken from now on as our reference sample.

3.2. $w_2(\theta)$ in Central Scans 756 and 94

As a test for systematics, we study $w_2(\theta)$ using only the central region of the CCD in scans 756 (north EDR) and 94 (south EDR). The seeing during run 756 is the best in the EDR, with only small fluctuations around 1.4. Regions in the SDSS where the seeing degraded to worse than 1.5 are marked for reobservation. As this includes most of the EDR data, one may worry that some of the results presented here could be affected by these seeing variations. As mentioned above, this has been shown not to be the case, at least for $r^* < 21$ (Scranton et al. 2002).

We estimate $w_2(\theta)$ using only galaxies in the central regions of the CCD in scans 756 (the best of EDR/N) and 94 (the best of EDR/S). Figures 4 and 5 show a piece of this new data set for the EDR/N. As can be seen in Figure 5, we consider only the central part of 756 to avoid any contamination from the CCD edges. This new data set contains only 30% of the area (and of the galaxies) from the whole strip.

Figure 6 (left panel) compares the results of $w_2(\theta)$ for the individual scans against the whole strip for all EDR/S, EDR/N, and EDR/(N+S). As can be seen in this figure, individual scans (dotted lines) agree very well with the corresponding overall strip values. Possible systematic errors seem quite small. In fact, the agreement is striking after a visual comparison of the heavy masking in the pixel maps of Figure 4 [which show the actual resolution use for the $w_2(\theta)$ estimation in Fig. 6]. One would naively expect some more significant sampling variations when we use only one third of the data. But nearby regions are strongly correlated, and we can get very similar results with only a fraction of the data (see also Figs. 13–15 of Scranton et al. 2002). This test shows the power of doing configuration space analysis as opposed to Fourier space analysis (Scoccimarro et al. 2001). It also illustrates that our estimator for $w_2(\theta)$ performs very well in dealing with masked data.

3.3. Seeing and Reddening Mask

Figure 7 shows the pixels EDR/N and EDR/S with a seeing better than 2$''$ and 0.2 maximum extinction. Pixels with larger seeing or larger extinction are masked out. As apparent in the figure, there is a significant reduction of...
the available area after the masking. Figure 6 (right panel) compares the results of $w_2(\theta)$ for the new masked maps with the results for the individual scans, for all EDR/S, EDR/N, and EDR/(N+S). There is now a much better agreement between the EDR/N and EDR/S, which suggests that the discrepancies between EDR/N and EDR/S apparent in the left panel of Figure 6 are due to these systematic effects. The number of available pixels (30% of the total) is comparable to the ones in individual scans (dotted line), which indicates that sampling errors cannot account for the observed differences between the dotted and solid lines. The biggest change is apparent in EDR/S, which is the smallest sample and the one subject to the worst seeing conditions. Most of the difference is due to the seeing rather than the extinction mask.

We find similar results for slightly lower cuts in seeing and extinction, but the number of pixels in EDR/S decrease very quickly as we lower the seeing, and sampling errors dominate over any possible systematics. Thus, such tests are not very conclusive.

Note that the APM absolute errors should be larger for the masked data, as there is less area available. Note also that the mean EDR/(N+S) is lower (Fig. 6, right panel, solid middle line). The absolute samplings error (as oppose to the relative errors) approach a constant on $\theta \sim 1^\circ$ scale (Fig. 8 and eq. [8]). All of this indicates that the final relative errors should be larger. Thus, taking into account these considerations, the discrepancies between EDR/N and EDR/S are not significant any more and certainly within $2\sigma$ errors from the APM.

3.4. Variance and Covariance

Bernstein (1994) has calculated the covariance in the angular two-point function $w_\theta \equiv w(\theta_\|)$, where $\theta_\|$ corresponds to the bins in angular separation (for a more general discussion on errors, see § 6 of Bernardeau et al. 2002). We will consider two main sources of errors. One is due to the finite number of particles $N$ in the distribution. This error is usually called the Poisson error and is expressed as $\sim 1/N$. The second is due to the finite size of the sample, which is characterized by

$$w_\Omega \equiv \frac{1}{\Omega^2} \iint \, d\Omega_1 \Omega_2 \, w(\theta_{12}) ,$$

Fig. 6.—Left panel: Zoom over a region of the second panel of Fig. 3. Here we compare the full strips of EDR/S, EDR/N, and EDR/(N+S) (solid lines from top to bottom at the largest angles) to the APM subsamples (triangle with error bars) of size similar to EDR/(N+S). Dotted lines: Central part of the CCD in scans 94 (top dotted line, next to EDR/S), 756 (bottom, next to EDR/N), and the joint 756 + 94 (middle dotted line, along EDR/[N+S]). Right panel, dotted lines: As in left panel. Right panel, solid lines: Seeing mask in Fig. 7. [See the electronic edition of the Journal for a color version of this figure.]

Fig. 7.—Pixel maps similar to Fig. 2, with the seeing and extinction mask.
the mean correlation function over the solid angle of the survey $\Omega$. This gives the uncertainty in the mean density on the scale of the sample, which is constrained to be zero in most estimators, as the mean density calculated from the same sample, i.e., estimators suffer from the integral constraint. In general, this integral is not zero, but we need a clustering model or a larger survey such as the APM to calculate its value. For the EDR size, $w_{12}$ is dominated by the value of $w(\theta)$ on the scale of the strip width $w(\theta = 5^\circ)$. From the APM $w(\theta)$ we find $\approx 10^{-2}$ for EDR/S. For the APM size itself, this integral should be significantly smaller but its value is quite uncertain. For both for EDR and APM subsamples, the Poisson errors $\approx 1/\sqrt{N} \approx 10^{-3}$ to $10^{-6}$ are typically smaller than the sampling errors. Neglecting Poisson errors and using $w_N \sim q_N w^{N-1}$ for higher order correlations, Bernstein (1994) found

$$\text{Cov} (w_i, w_j) \approx g(\gamma) w_i^2 + \beta w_{12}(w_i - w_{12})(w_j - w_{12}),$$  \hspace{1cm} (3)

where $g(\gamma)$ is a geometric term of order unity for power-law correlations $w(\theta) \approx A\theta^\gamma$. In the strict hierarchical model $w_N = q_N w^{N-1}$, we have $\beta = 4(1 - q_3 + q_4)$. As this model is a rough approximation,\(^5\) we will take $\beta$ to be a constant, which will be fitted using the simulations. The corresponding expression for the variance (diagonal of the covariance) is

$$\text{Var} (w_i) \approx g(\gamma) w_i^2 + \beta w_{12}(w_i - w_{12})^2.$$  \hspace{1cm} (4)

\(^5\)It neglects the configuration and the scale dependence of $q_3$ and $q_4$, which is only a good approximation on nonlinear scales (Bernardeau et al. 2002).

In Figure 8, we compare the square root of the above expression $\Delta w(\theta) \equiv [\text{Var} (w_i)]^{1/2}$ with the rms errors in $w_j$ from the dispersion in 10 APM subsamples that simulate the geometry of the EDR/S sample. We find that a value of $\beta \approx 4$ fits well the above theoretical model to the errors in the simulations. In principle, both $g$ and $\beta$ could be a function of scale, but the model seems to match well the simulations, at least in the range $\theta \approx 0.2' - 4.0'$. On smaller scales, we are approaching the map pixel resolution and should also include the variance due to the shot noise and finite cell size. On scales larger than $\theta \approx 4'$, we approach the EDR strip size and the integral constraint becomes important. As we have not corrected for the integral constraint, we do not expect our errors to follow the predictions on large scales. In the intermediate regime, the model seems to work quite well.

Bernstein (1994) has shown, using Monte Carlo simulations, that the model in equation (3) works well for the covariance matrix. His Figure 2 shows the covariance between adjacent bins $\text{Cov} (w_i, w_{i+1})$. These predictions should work well here if we compare alternate bins $\text{Cov} (w_i, w_{i+2})$ instead of adjacent bins, as we are using 12 bins decade$^{-1}$ as opposed to 6 bins decade$^{-1}$ in Bernstein (1994). The resulting covariance matrix is close to singular, and most of its principal components are degenerate. Thus, a significant test estimation is not straightforward.

With the help of Monte Carlo simulations Bernstein (1994) concluded that the effect of the off-diagonal errors is small when fitting parametric models, in particular, a power law to $w(\theta)$. He finds similar results for the amplitude and the slope when using the simple diagonal $\chi^2$ minimization or the principal components of the full covariance matrix. Both the level of clustering and the errors in his Monte
Carlo simulations are quite similar to the ones presented here (cf. our Fig. 8, left panel, to his Fig. 1). Thus, we can extend the conclusions of Bernstein (1994) to the present analysis and, for simplicity, ignore the off-diagonal errors in the covariance matrix. In order to make sure that the same conditions apply, we should use only every other bin in fitting models.

3.5. An Inflection Point in \( w_2(\theta) \)?

Figure 8 (right panel) shows the logarithmic slope

\[
\gamma(r) = \frac{d \log w_2(\theta)}{d \log \theta}
\]

of \( w_2(\theta) \) for the estimation in the second panel of Figure 3. The mean and errors in the top panel correspond to APM subsamples similar to the EDR/S. Within these errors, both the APM and SDSS data are compatible with a power law \( w_2(\theta) \propto \theta^{-\gamma} \) with \( \gamma \) between \( \gamma \approx -0.6 \) and \( \gamma \approx -0.8 \) (horizontal dotted lines), in good agreement with Table 1 in Connolly et al. (2001) and Maddox (1990). Even with these large errors, there is a hint of a systematic flattening of \( \gamma \) between 0.1 and 1.0 in all subsamples. This hint is clearer in the combined analysis EDR/(N+S), where the errors (according to the APM subsamples) are significantly smaller. This flattening of only \( \Delta \gamma \approx 0.1-0.2 \) as we move from 0.1 to 1.0, is apparent in all the APM and SDSS subsamples. It is reassuring that even at this detailed level, all data agree within the errors. It is also apparent from the top right panel of Figure 8 that the errors are too large to detect this effect separately in EDR/S or EDR/N, so it depends on the good calibration of EDR across the disjointed EDR/N and EDR/S samples.

The best fit to a power-law model gives \( \chi^2 \approx 20 \) for 10 degrees of freedom, which corresponds to a 3% confidence level for a power law to be a good fit. If we do not use adjacent bins (see § 3.4), we find \( \chi^2 \approx 19 \) for 5 degrees of freedom, which gives an even lower confidence level.

3.6. Smoothed One-Point Moments

We next compare the lower-order moments of counts in cells of variable size \( \theta \) (larger than the pixel map resolution). We follow closely the analysis of Ga02. Figure 9 shows the variance of fluctuations in density counts \( \delta \equiv \rho - \bar{\rho} \) smoothed over a scale \( \theta \), \( w_2 \equiv \langle \delta^2(\theta) \rangle \), which is plotted as a function of the smoothing radius \( \theta \). The errors show a 1 \( \sigma \) confidence interval for APM subsamples with EDR/(N+S) size. The individual results in each subsample are strongly correlated so that the whole curve for each subsample scales up and down within the errors, i.e., there is a strong covariance at all separations due to large-scale density fluctuations (Hui & Gaztanaga 1999 or eq. [2] above). The EDR/(N+S) results (solid and dotted lines) match perfectly well the APM results, in agreement with what we found for the two-point function above. The size of the error bars for EDR/N and EDR/S (not shown) are almost a factor of 2 larger than for EDR/(N+S), so that they are also in agreement with the APM within their respective sampling errors.

Figure 9 (right panel) shows the corresponding comparison for the normalized angular skewness,

\[
s_3(\theta) \equiv \frac{\langle \delta^3(\theta) \rangle}{\langle \delta^2(\theta) \rangle^{3/2}} = \frac{w_2(\theta)}{w_3(\theta)}.
\]

All SDSS \( g' \) subsamples (top panel) for \( s_3 \) show an excellent agreement with the APM at the smaller scales (contrary
to the Edinburgh/Durham Southern Galaxy Catalog [EDSGC] results in Szapudi & Gaztañaga 1998). On larger scales, the SDSS values are smaller, but the discrepancy is not significant given the strong covariance of individual APM subsamples. Note how the effect of the seeing mask (dotted line) is to increase the amplitude of $S_3$; this could be partially due to systematic errors, but it could also result from the smaller (one-third) sampling resulting from removing the pixels with bad seeing.

Figure 9 (bottom panel, left) shows the corresponding results in $r'$. At the smallest scale of about $2'$ or $240$ h$^{-1}$ Kpc, we find some slight discrepancies (at the 1 $\sigma$ level for a single point) with the APM. The $r'$ results seem to be a scaled-up version of the $g'$ results, which indicates that the apparent differences could be explained in terms of sampling effects (with strong covariance). Note also that the value of $S_3$ seems to peak at a slightly larger scale. This could indicate another explanation for this discrepancy. Szapudi & Gaztañaga (1998) argued that such a peak could be related to some systematic (or physical) effect related to the deblending of large galaxies. It is reasonable to expect that such an effect could be a strong function of color, as $g'$ and $r'$ trace different aspects of the galaxy morphology. We have also checked that results of individual scans 756 and 94 (and also 756 + 94) give slightly higher results, closer to the $r'$ results than to the mean $g'$. Higher results are also found for the results with the seeing mask (Fig. 9, top right panel, dotted line). This goes in the right direction if we think that deblending gets worse with bad seeing, but it could also be affected by sampling fluctuations because of the smaller area in the scans or masked data.

Similar results are found for higher-order moments. As we approach the scale of $2'$, the width of our strip, it becomes impossible to do counts for larger cells, and it is better to study the three-point function.

**3.7. Three-Point Correlation Function**

Following Ga02, we next explore the three-point function normalized as

$$q_3 = \frac{w_3(\theta_{12}, \theta_{13}, \theta_{23})}{w_2(\theta_{12})w_2(\theta_{13}) + w_2(\theta_{12})w_2(\theta_{23}) + w_2(\theta_{13})w_2(\theta_{23})},$$

where $\theta_{12}, \theta_{13},$ and $\theta_{23}$ correspond to the sides of the triangle formed by the three angular positions of $\delta_1 \delta_2 \delta_3$. Here we will consider isosceles triangles, i.e., $\theta_{12} = \theta_{13}$, so that $q_3 = q_3(\alpha)$ is given as a function of the interior angle $\alpha$, which determines the other side of the triangle $\theta_{23}$ (Frieman & Gaztañaga 1999).

We also consider the particular case of the collapsed configuration $\theta_{23} = 0$, which corresponds to $\langle \delta_1 \delta_2^2 \rangle$ and is normalized in a slightly different way (Szapudi & Szalay 1999).

$$c_{12} = \frac{\langle \delta_1 \delta_2^2 \rangle}{\langle \delta_1 \delta_2 \rangle} \approx 2q_3(\alpha = 0).$$

Figure 10 shows $c_{12}$ from the collapsed three-point function.

Note the strong covariance in comparing the EDR/N to EDR/S. The unmasked EDR/(N+S) results agree well with the APM within errors, but the results with the seeing mask (dotted line) show significant departures at small scales.

Figure 10 (right panel) shows the reduced three-point function $q_3$ for isosceles triangles of side $\theta_{12} = \theta_{13} = 0.5$ deg. Here there seem to be differences between APM and SDSS, but its significance is low because the error at a single point is only 1 $\sigma$ and there is a strong covariance between points; e.g., note how the EDR/S and EDR/N curves are shifted around the EDR/(N+S) curve. The APM seems closer to the EDR/N. This is a tendency that is apparent in previous fig-

---

**Fig. 10.—Left panel:** Collapsed three-point function $c_{12}$ as a function of galaxy separation $\theta$ for SDSS $17 < g' < 20$. Short and long dashed lines: SDSS EDR/N (north) and EDR/S (south) strips. Dotted and solid lines: EDR/(N+S) with and without the seeing mask shown in Fig. 7. Triangles with error bars: Mean and 1 $\sigma$ confidence level in the values of 10 APM subsamples ($17 < g' < 20$) with same size as the joint EDR/(N+S) with masked seeing. **Right panel:** Similar results for the three-point function $q_3(\alpha)$ in isosceles triangles of sides $\theta_{12} = \theta_{13} = 0.5$. [See the electronic edition of the Journal for a color version of this figure.]
ures, but it is only on $g_2$ that the discrepancies start to look significant. When estimation biases are present in the subsample mean, errors tend to be unrealistically small, i.e., errors are also biased down (Hui & Gaztañaga 1999).

4. DISCUSSION AND CONCLUSIONS

We have first explored the different uncertainties involved in the comparison of the SDSS with the APM, such as the band or magnitude range to use. After several tests, we conclude that clustering in both the north and south EDR strips (EDR/N and EDR/S) agree well in amplitude and shape with the APM on scales $\theta < 2'$. But we find inconsistencies with the APM $w_2(\theta)$ at the level of 90% significance on any individual scale at $\theta > 2'$. These inconsistencies are larger than 90% when we compare EDR/S to EDR/N at any given point at $\theta > 2'$ (cf. Fig. 3, short and long dashed lines, left panel). We have shown that this is mostly due to systematic photometric errors due to seeing variations across the SDSS EDR (Fig. 3, right panel).

We have pushed the comparison further by combining the north and south strips, which we call EDR/(N+S), and analyzing the EDR clustering as a whole. Combining samples in such a way is very risky because small systematic differences in the photometry tend to introduce large uncertainties in the overall mean surface density. This was overcome in the APM by a simultaneous match of many overlapping plates. For noncontiguous surveys, the task is almost impossible unless one has very well-calibrated photometric observations, as is the case for the SDSS, to a level of 0.03 mag (Lupton et al. 2001). The combined EDR/(N+S) sample shows very good agreement for the number counts (Fig. 1) and also with the APM $w_2(\theta)$, even at $\theta > 2'$. In this case, the agreement is, in fact, within the corresponding sampling errors in the APM.

Higher-order correlations show similar results. The mean SDSS skewness is in good agreement with the APM at all scales. The current SDSS sampling (1 $\sigma$) errors range from 10% at scales of arcminutes (less than 1 h$^{-1}$ Mpc) to about 50% on degree scales (~10 Mpc h$^{-1}$). At this level, both surveys are in perfect agreement. The collapsed three-point function, $c_{12}$, shows even smaller errors: this is because multipoint statistics are better sampled over narrow strips than counts in large smoothed cells. At degree scales that correspond to the weakly nonlinear regime $r \sim 8$ Mpc h$^{-1}$, we find $c_{12} \approx 2.4 \pm 0.6$. This amplitude and also the shape is remarkably similar to that found in simulations and what is theoretically expected from gravitational instability $c_{12} \approx 68/21 + 2/3\gamma$ (Bernardeau 1996; Gaztañaga, Fosalba, & Croft 2002). The three-point function for isosceles triangles of side $\theta_1$ = $\theta_2$ = $\theta_3$ = 0.5, shown in Figure 10. Figure 10 (left panel) seems lower than the APM values but within the 2 $\sigma$ confidence level at any single point.

Again, here we would need the covariance matrix to say more. In general, the north SDSS strip has higher amplitudes for the reduced skewness or three-point function than the southern strip.

We conclude that the SDSS is in good agreement with the previous galaxy surveys and, thus, with the idea that gravitational growth from Gaussian initial conditions is most probably responsible for the hierarchical structures we see in the sky (Bernardeau et al. 2002, and references therein).

The above agreement has encouraged us to look into the detailed shape of $w_2(\theta)$ on intermediate scales, where the uncertainties are smaller and errors from the APM are more reliable. On scales of 0.1–1 $', we find indications of slight (~20%) deviations from a simple power-law (note that these deviations are on larger scales than the power-law deviations found in Connolly et al. 2001). Figure 8 (right panel) shows that the different SDSS samples have very similar slopes to the APM survey, showing a characteristic inflection with a maximum slope. In hierarchical clustering models, the initial slope of $d \ln \xi / d \ln r$ is a smoothed decreasing function of the separation $r$. Projection effects can partially wash out this curve but cannot produce any inflection to the shape (at least if the selection function is also regular). Gaztañaga & Juszkiewicz (2001, and references therein) argued that weakly nonlinear evolution produces a characteristic shape in $d \ln \xi / d \ln r$. This shape, smoothed by projections, is evident in the APM data for $d \ln w(\theta)/d \ln \theta$. Here we also find evidence for such a shape in the combined EDR/(N+S) SDSS data. The maximum in the slope occurs around $\theta \approx 0.6'$, which corresponds to $r \approx 5$ h$^{-1}$ Mpc, as expected if biasing is small on those scales.

In summary, both the shape of the two-point function and the shape and amplitude of the three-point function and skewness in the SDSS EDR data confirm the idea that galaxies are tracing the large-scale matter distribution that started from Gaussian initial conditions (Bernardeau et al. 2002, and references therein).

I would like to thank R. Scranton and J. Frieman for useful discussions and comments on the first version of this paper. I acknowledge support by grants from IEEC/CSIC, DGI/MCT AyA2002-00850, and DGI/MCT BFM2000-0810 and from Coordinacion Astrofisica, INAOE. Funding for the creation and distribution of the SDSS Archive$^6$ has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society.

$^6$ The SDSS Web site is http://www.sdss.org/.

REFERENCES

Bernardeau, F. 1996, A&A, 312, 11 (B96)
Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Phys. Rpt., 367, 1
Bernstein, G. M. 1994, ApJ, 424, 569
Connolly, A., et al. 2002, ApJ, 579, 42
Dodelson, S., et al. 2001, ApJ, 572, 140
Frieman, J. A., & Gaztañaga, E. 1999, ApJ, 521, L83 (FG99)
Gaztañaga, E. 2002, 333, L21 (G02)
Hui, L., & Gaztañaga, E. 1999, ApJ, 519, 622
Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., Kent, S., & Yasuda, N. 2001, in ASP Conf. Ser. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Handen, F. A. Primini & H. E. Payne (San Francisco: ASP), 269
Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 43
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Scoccimarro, R., Feldman, H., Fry, J. N., & Frieman, J. A. 2001, ApJ, 546, 652

Gaztañaga, E., & Juszkiewicz, R. 2001, ApJ, 558, L1
Hui, L., & Gaztañaga, E. 1999, ApJ, 519, 622
Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., Kent, S., & Yasuda, N. 2001, in ASP Conf. Ser. 238, Astronomical Data Analysis Software and

$^{6}$ The SDSS Web site is http://www.sdss.org/.
Scranton, R., et al. 2002, ApJ, 579, 48
Szapudi, I., & Gaztañaga, E. 1998, MNRAS, 300, 493
Szapudi, I., & Szalay, A. S. 1999, ApJ, 515, L43
Tegmark, M., Hamilton, A. J. S., Strauss, M. A., Vogeley, M. S., & Szalay, A. S. 1998, ApJ, 499, 555

Tegmark, M., et al. 2002, ApJ, 571, 191
Yasuda, N., et al. 2001, AJ, 122, 1104
York, D. G., et al. 2000, AJ, 120, 1579
Zehavi, I., et al. 2002, ApJ, 571, 176