Mon. Not. R. Astron. Soc. **356,** 456–474 (2005) doi:10.1111/j.1365-2966.2004.08446.x

The 2dF Galaxy Redshift Survey: the nature of the relative bias between galaxies of different spectral type

Edward Conway,[1⋆] Steve Maddox,[1⋆] Vivienne Wild,[3] John A. Peacock,[6] Ed Hawkins,[1] Peder Norberg,[5] Darren S. Madgwick,[4] Ivan K. Baldry,[7] Carlton M. Baugh,[2] Joss Bland-Hawthorn,[8] Terry Bridges,[8] Russell Cannon,[8] Shaun Cole,[2] Matthew Colless,[9] Chris Collins,[10] Warrick Couch,[11] Gavin Dalton,[12,13] Roberto De Propris,[11,17] Simon P. Driver,[9] George Efstathiou,[3] Richard S. Ellis,[14] Carlos S. Frenk,[2] Karl Glazebrook,[7] Carole Jackson,[9] Bryn Jones,[1] Ofer Lahav,[16] Ian Lewis,[12] Stuart Lumsden,[15] Will Percival,[6] Bruce A. Peterson,[9] Will Sutherland[3] and Keith Taylor[14] (The 2dFGRS Team)

[1]*School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD*
[2]*Department of Physics, University of Durham, South Road, Durham DH1 3LE*
[3]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*
[4]*Hubble Fellow, Lawrence Berkeley National Laboratory, MS50R-5032, Berkeley, CA 94720, USA*
[5]*Institut für Astronomie, ETH Hönggerberg, CH-8093 Zürich, Switzerland*
[6]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[7]*Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218-2686, USA*
[8]*Anglo-Australian Observatory, PO Box 296, Epping, NSW 2121, Australia*
[9]*Research School of Astronomy & Astrophysics, The Australian National University, Weston Creek, ACT 2611, Australia*
[10]*Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Birkenhead, L14 1LD*
[11]*Department of Astrophysics, University of New South Wales, Sydney, NSW 2052, Australia*
[12]*Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH*
[13]*Space Science and Technology Division, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX*
[14]*Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*
[15]*Department of Physics, University of Leeds, Woodhouse Lane, Leeds LS2 9JT*
[16]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*
[17]*Astrophysics Group, Department of Physics, Bristol University, Tyndall Avenue, Bristol BS8 1TL*

Accepted 2004 September 23. Received 2004 July 19; in original form 2004 April 15

ABSTRACT

We present an analysis of the relative bias between early- and late-type galaxies in the Two-degree Field Galaxy Redshift Survey (2dFGRS) – as defined by the η parameter of Madgwick et al., which quantifies the spectral type of galaxies in the survey. We calculate counts in cells for flux-limited samples of early- and late-type galaxies, using approximately cubical cells with sides ranging from 7 to 42 h^{-1} Mpc. We measure the variance of the counts in cells using the method of Efstathiou et al., which we find requires a correction for a finite volume effect equivalent to the integral constraint bias of the autocorrelation function. Using a maximum-likelihood technique we fit lognormal models to the one-point density distribution, and develop methods of dealing with biases in the recovered variances resulting from this technique. We then examine the joint density distribution function, $f(\delta_{\rm E}, \delta_{\rm L})$, and directly fit deterministic bias models to the joint counts in cells. We measure a linear relative bias of $\approx$1.3, which does not vary significantly with ℓ. A deterministic linear bias model is, however, a poor approximation to the data, especially on small scales ($\ell \leqslant 28\,h^{-1}$ Mpc) where deterministic linear bias is excluded at high significance. A power-law bias model with index $b_1 \approx 0.75$ is a significantly better fit to the data on all scales, although linear bias becomes consistent with the data for $\ell \gtrsim 40\,h^{-1}$ Mpc.

⋆E-mail: edward.conway@astro.nottingham.ac.uk (EC); steve.maddox@nottingham.ac.uk (SM)

Key words: surveys – galaxies: distances and redshifts – galaxies: statistics – large-scale structure of Universe.

1 INTRODUCTION

Measurements of large-scale structure from galaxy redshift surveys obviously measure the distribution of luminous matter only; the total mass distribution will be dominated by dark matter. The question of how the galaxies trace the total matter density field is therefore extremely pertinent, both to the estimation of cosmological parameters, and also as a probe of the physics of galaxy formation. A common assumption is 'linear biasing', which can be expressed as $\delta_{\rm g} = b\delta_{\rm m}$, where $\delta_{\rm g}, \delta_{\rm m}$ are the fractional overdensities relative to the mean in galaxies and mass, respectively. This assumption becomes unphysical when $b > 1$ since, by definition, $\delta_{\rm g} \geqslant -1$, but we can still define a bias parameter, $b(r)$, by e.g. $\xi_{\rm gg}(r) = b(r)^2 \xi_{\rm mm}(r)$. Many of the constraints on cosmological parameters derived from large-scale structure measurements rely on an understanding of galaxy bias. Both the 2dFGRS power spectrum analysis (Percival et al. 2001) and the constraints obtained for the neutrino mass (Elgarøy & Lahav 2003) assume scale-independent bias. Joint constraints obtained by combining the 2dFGRS results with measurements of the CMB power spectrum (Efstathiou et al. 2002; Percival et al. 2002; Verde et al. 2003) also require a model for galaxy bias. Dekel & Lahav (1999) show that non-linearity and stochasticity in the bias relation can explain discrepancies between different methods of measuring parameters which assume a linear bias factor, such as measurements of $\beta = \Omega_{\rm m}^{0.6}/b$ (Peacock et al. 2001; Hawkins et al. 2003).

In fact both theoretical approaches (Mo & White 1996) and simulations predict that bias may be non-linear and scale dependent, at least on some (small) scales. Kauffmann, Nusser & Steinmetz (1997) find only weak scale dependence on large scales and a bias relation consistent with linear bias. Benson et al. (2000) find that semi-analytic galaxies in a ΛCDM model could reproduce the APM survey correlation function given a scale-dependent bias taking the form of an antibias of galaxies relative to matter on small scales. Somerville et al. (2001) also use semi-analytic modelling to demonstrate that the physics of galaxy formation introduces a small scatter in the galaxy–mass relation; they find the mean bias to have only a weak dependence on scale for $r \lesssim 12\,h^{-1}$ Mpc (where the Hubble constant is $H_0 = 100\,h\,{\rm km\,s^{-1}}$).

In principle the true mass distribution can be directly measured from galaxy peculiar velocities using, e.g. POTENT reconstruction (Dekel, Bertschinger & Faber 1990). In practice accuracy is hard to achieve by such methods; the technique requires heavy smoothing since the error bars per galaxy are large and the volumes surveyed up to the present are relatively local. An alternative probe is to compare the clustering of different types of galaxy : if these cluster differently, at least one type cannot exactly follow the mass distribution.

It has been known for some considerable time that galaxies of different morphological type have different clustering properties. Early-type galaxies, such as ellipticals or S0s, are highly clustered, accounting for almost 90 per cent of galaxies in the cores of rich clusters. This fraction drops off steeply with distance from the cluster cores and in the field 70 per cent of galaxies are late-type galaxies: spirals and irregulars (Dressler 1980; Postman & Geller 1984). The level of fluctuations in each of the early- and late-type density fields can also be compared using the correlation functions or power spectra for the two subpopulations. This kind of study is optimized for small separations ($\lesssim 10\,h^{-1}$ Mpc) and has generally revealed that the clustering amplitude of ellipticals is greater than that of spirals by a factor of 1.3–1.5 (e.g. Loveday et al. 1995; Norberg et al. 2002a; Zehavi et al. 2002; Madgwick et al. 2003b). If both density fields were linearly proportional to the matter density field this factor would be equivalent to the ratio between linear bias parameters $(b_{\rm E}(r)/b_{\rm L}(r))^2$. There is also evidence that the relative bias between subpopulations of galaxies is more complex than the global galaxy bias. Measurements of the 2dFGRS bispectrum (Verde et al. 2002) found no evidence for non-linearity in the bias for 2dFGRS galaxies. More recently, however, Kayo et al. (2004) find evidence for relative bias being complex on weakly non-linear to non-linear scales from a measurement of the redshift-space three-point correlation function, as a function of galaxy colour and morphology, in the Sloan Digital Sky Survey. Wild et al. (2004) have carried out a counts-in-cells analysis using volume-limited samples from the 2dFGRS, and find evidence for non-linearity and stochastic effects.

A detailed framework for dealing with possible non-linearities and stochasticity in the bias relation is given by Dekel & Lahav (1999), based on the joint probability distribution of the galaxy and mass densities $f(\delta_{\rm g}, \delta_{\rm m})$. In an analogous manner we will consider the joint probability distribution of the early- and late-type galaxy density fields for magnitude-limited samples in the 2dFGRS. This approach in large part follows the methods described in Blanton (2000) for the Las Campanas Redshift Survey (LCRS), although the geometry of the 2dFGRS is considerably more amenable to this kind of study than that of the LCRS and allows us, for example, to examine a large range of scales.

This paper aims to explore quantitatively the nature of the relative bias on several levels. First we measure the value of the linear relative bias factor between galaxies of different spectral types over a large range of scales. We then discriminate between a range of possible models for the relative bias to determine if any of the aspects of complex bias (e.g. scale dependence, non-linearity or stochasticity) are present in the data. Finally, we attempt to quantify the nature of these aspects of biasing. Our approach complements that of the related study of Wild et al. (2004) who analyse volume-limited samples of the 2dFGRS segregated by both spectral type and rest-frame colour. The approach adopted by Wild et al. (2004) is more focused on measuring the stochasticity of the relative bias, which is poorly constrained by the results of this paper. We place greater emphasis on alternative measurements of the linear relative bias factor, and on quantifying the degree of non-linearity present in the bias relation.

The paper is organized into sections as follows. In Section 2 we summarize details of the 2dFGRS, the PCA-η parameter and the division into cells. We present a measurement of the variances of the counts in cells using the method of Efstathiou et al. (1990), which we have corrected for integral constraint bias, in Section 3. In Section 4 we present an analysis of the one-point distribution of the counts in cells based on fits to a lognormal distribution function. In Section 5 we discuss the relative bias. We present the results of applying the 'modified χ^2' statistic of Tegmark (1999) to the joint counts in cells and then move on to describe the application of the maximum-likelihood technique of Blanton (2000) to constrain the relative bias between spectral types. We discuss in detail some of the main strands of the paper in Section 6 and present a summary of our conclusions in Section 7.

2 THE 2DF GALAXY REDSHIFT SURVEY

The 2dFGRS observations were carried out between 1997 May and 2002 April using the 2dF instrument: a multi-object spectrograph on the Anglo-Australian Telescope. The main survey region consists of two broad strips, one in the South Galactic Pole region (SGP) covering approximately $-37.^\circ5 < \delta < -22.^\circ5$, $21^{\rm h}40^{\rm m} < \alpha < 3^{\rm h}40^{\rm m}$ and the other in the direction of the North Galactic Pole (NGP), spanning $-7.^\circ5 < \delta < 2.^\circ5$, $9^{\rm h}50^{\rm m} < \alpha < 14^{\rm h}50^{\rm m}$. In addition there are a number of circular two-degree fields scattered randomly over the full extent of the low-extinction regions of the southern APM galaxy survey.

The parent catalogue for the survey was selected in the photometric $b_{\rm J}$ band from a revised and extended version of the APM galaxy survey (Maddox et al. 1990a,b,c, 1996). The magnitude limit at the start of the survey was set at $b_{\rm J} = 19.45$ but both the photometry of the input catalogue and the dust extinction map have since been revised and so there are small variations in magnitude limit as a function of position over the sky. The effective median magnitude limit, over the area of the survey, is $b_{\rm J} \approx 19.3$ (Colless et al. 2001, 2003).

The completeness of the survey data varies according to the position on the sky because of unobserved fields (mostly around the survey edges), untargeted objects in observed fields (due to collision constraints or broken fibres) and observed objects with poor spectra; also there are drill-holes around bright stars. The variation in completeness with angular position, $\boldsymbol{\theta}$, is fully described by the completeness mask (Colless et al. 2001; Norberg et al. 2002b; Colless et al. 2003). Note that since we use exclusively those galaxies for which a principal component spectral type has been derived, we require a slightly modified completeness mask from that describing the completeness of the full survey, one which reflects the completeness of galaxies with measured η-type (Norberg et al. 2002a).

We use the completed 2dFGRS data set which was released publicly at the end of June 2003 (Colless et al. 2003). This includes 221 414 unique, reliable galaxy redshifts (quality flag $\geqslant 3$, Colless et al. 2001, 2003). The random fields, which contain nearly 25 000 reliable redshifts, are not included in this analysis. Throughout the paper we treat the NGP and SGP regions as independent data sets, which means we have two estimates for each of the statistics we derive. This approach functions both as a 'reality check' for our error estimates and also gives an idea of the variation due to cosmic variance.

2.1 PCA classification of galaxy spectra

The spectral properties of 2dFGRS galaxies have been analysed and the galaxies split into spectral type classes using a principal component analysis (PCA) described by Madgwick et al. (2002). This technique splits the galaxies on the basis of the characteristics of their spectra which show the most variation across the sample, without using any prior assumptions or template spectra. Madgwick et al. (2002) define a scalar parameter, η, which is a linear combination of the first two principal components chosen to minimize instrumental effects which make the determination of the continuum uncertain. In effect, η quantifies the relative strengths of emission and absorption lines, and can be shown to be tightly correlated to the equivalent width of Hα in particular, so a simple physical interpretation of η is as a measure of the relative star formation rate in a galaxy (i.e. the ratio of present to past star formation; Madgwick et al. 2003a).

The PCA classification makes use of the spectral information in the rest-frame wavelength range 3700–6650 Å, which includes all the major optical diagnostics between O II and Hα. The spectral coverage imposes a limit on the maximum redshift at which this analysis can be used of $z = 0.2$. For galaxies with $z > 0.15$, however, sky absorption bands contaminate the Hα line, which affects the stability of the classification. For this reason we restrict our analysis to galaxies with $z < 0.15$ following Madgwick et al. (2002).

The distribution of η for the 2dFGRS spectra is clearly bimodal (see fig. 4 of Madgwick et al. 2002), with a local minimum at $\eta = -1.4$. Madgwick et al. (2002) divide galaxies into four spectral type bins based on the shape of this distribution; the local minimum at $\eta = -1.4$ is used to separate early and late types while the late-type 'shoulder' is divided in two and also separated from the tail, which will be dominated by particularly active galaxies such as starbursts and active galactic nuclei (AGN). Because of the effects of possible evolution in the last two spectral type bins, discussed further in the next section, we use only the spectral classes 1 and 2 of Madgwick et al. (2002) in this paper, which we refer to as early and late type, respectively, and exclude the bluer classes 3 and 4.

A comprehensive review of the relationship between η and the traditional morphological type is presented in Madgwick et al. (2003a), along with an analysis of the effectiveness of the PCA method in recovering the morphological type. The median value of η for each morphological type does correlate rather well with the classification bins described above, although there is considerable scatter in the η-classification of spirals. More recently it has been possible to determine accurate rest-frame colours for the 2dFGRS galaxies. A similar bimodality exists in the colour distribution as that present in the distribution of η although the relationship is not exactly one-to-one between colour and η-type (Wild et al. 2004). The important aspect of the PCA classification used here is that it represents a coherent method for dividing a galaxy sample into classes based on a diagnostic with a relatively clear physical interpretation, i.e. relative star formation rate. Note that this specific interpretation will not necessarily be the case for all galaxy samples classified using a PCA method; comparing classifications based on a PCA analysis between samples is in general non-trivial. In our case, however, the fact that the PCA classification of the 2dFGRS is dominated by Hα means that a classification of galaxies based on η is virtually the same as a classification based on the equivalent width of Hα.

2.2 Counts in cells

The analysis presented in this paper is complementary to approaches to the study of the relative bias based on correlation functions, in that we compare the two density fields on a point-by-point basis, rather than measuring overall clustering amplitudes. To estimate the local galaxy density contrast for each type we use the method of counts in cells. Analysis of the 2dFGRS counts in cells has already been used by Croton et al. (2004a,b) and Baugh et al. (2004) to constrain the higher-order correlation functions and the void probability function. Another respect in which this analysis is complementary to the correlation function approach is that it is optimized for much larger scales.

The method by which we divide the survey region into cells is identical to that first used by Efstathiou et al. (1990) for measuring the variance of the counts in cells for a sparse-sampled redshift survey of IRAS galaxies (Rowan-Robinson et al. 1990). First the surveyed region of space is divided into redshift shells of thickness ℓ centred on the observer. Each shell is then subdivided into approximately cubical cells using lines of constant right ascension and declination, chosen for each radial distance and declination to ensure that the sides of the cell are of length $\approx \ell$, the shell separation.

In this paper we analyse cell divisions over a range in ℓ from $\ell = 7\,h^{-1}$ Mpc to $\ell = 42\,h^{-1}$ Mpc. We assume a concordance cosmology of $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$. Effective scales corresponding to a given value of ℓ can be computed using the approximate relationship between the radius of a Gaussian sphere window, $R_{\rm G}$, the radius of a spherical top-hat window, $R_{\rm T}$, and ℓ, given in equation (1) (Peacock 1999):

$$R_{\rm G} \sim \frac{R_{\rm T}}{\sqrt{5}} \sim \frac{\ell}{\sqrt{12}}. \tag{1}$$

Note that for this analysis we have translated redshifts directly into distances, ignoring the effects of peculiar velocity. Since peculiar velocities for early-type galaxies are known to be larger than for late types, it is important to consider whether redshift-space distortions will affect the conclusions drawn in this paper. On non-linear scales, the dominant effect of peculiar velocities is to produce the Fingers of God, but on the larger scales relevant here, the linear β-effect will be dominant (Kaiser 1987). From the measurements of the value of β for galaxies of different spectral class obtained by Madgwick et al. (2003b) we obtain

$$\frac{P_{\rm s,E}}{P_{\rm s,L}} = 0.99\frac{P_{\rm r,E}}{P_{\rm r,L}}, \tag{2}$$

where $P_{\rm s,E}$, $P_{\rm s,L}$ are the redshift-space power spectra for early and late types, respectively, and the subscript 'r' denotes their real-space counterparts. From this it is clear that the effect of peculiar velocities is not significant within our current errors, although any effect will be in the sense of reducing the difference in the measured clustering between spectral types.

Any analysis of the galaxy counts in cells for a flux-limited survey must take into account the selection function, which quantifies the probability that a galaxy with a given redshift, z, is included in the survey. We define $M_{\rm min}(z, \boldsymbol{\theta})$ and $M_{\rm max}(z, \boldsymbol{\theta})$ to be the minimum and maximum absolute magnitudes visible at redshift z given the magnitude limit of the survey which, in the case of the 2dFGRS, varies with angular position $\boldsymbol{\theta}$ as described in Section 2, and we take the luminosity function $\Phi(M)$ to be normalized in that range. Then the selection function can be written

$$\phi(z, \boldsymbol{\theta}) = \int_{M_{\rm min}(z,\boldsymbol{\theta})}^{M_{\rm max}(z,\boldsymbol{\theta})} {\rm d}M\ \Phi(M) c_z(\boldsymbol{\theta}). \tag{3}$$

The $c_z(\boldsymbol{\theta})$ term describes the variation in completeness of the survey over the sampled area, for which we use the survey mask for η-typed galaxies described in Section 2. The completeness also depends slightly on apparent magnitude and, for the full survey, one can use the relation given in Colless et al. (2001) to parametrize this variation:

$$c_z(\boldsymbol{\theta}, b_{\rm J}) = \gamma\{1 - \exp[b_{\rm J} - \mu(\boldsymbol{\theta})]\}, \tag{4}$$

where γ is set at 0.99 and $\mu(\boldsymbol{\theta})$ can be set by the requirement that

$$\langle c_z(\boldsymbol{\theta}, b_{\rm J})\rangle = c_z(\boldsymbol{\theta}), \tag{5}$$

where the average is made over the expected distribution of apparent magnitudes, $b_{\rm J}$.

Unfortunately, the $\mu(\boldsymbol{\theta})$ mask is undefined for η-typed galaxies. To generate such a mask one would need to assume a form for the number counts as a function of η, as well as making an assumption for the variation of completeness with limiting redshift, $z_{\rm max}$, since the η parameter is only defined for $z \leqslant 0.15$. Given the large number of assumptions which would be necessary we have not included a correction for the effect of apparent magnitude on completeness in this study; the effect is in any case a small one, particularly in a flux-limited catalogue.

Once we have a knowledge of the selection function we can define the expected number of galaxies in each cell i by

$$N_{{\rm exp},i} = \int_{V_i} {\rm d}V_i\ \phi(z, \boldsymbol{\theta}) \tag{6}$$

where the integral is over the volume of the cell i.

We use expected counts in cells ($N_{{\rm E,exp},i}$ and $N_{{\rm L,exp},i}$ for early and late types, respectively), which we obtain by integrating the η-type dependent luminosity functions of Madgwick et al. (2002) using the average magnitude limit over the surface of each cell (Colless et al. 2001, 2003) and corrected for $c_z(\boldsymbol{\theta})$ as described above. We reject from the analysis cells for which the average completeness over the cell is less than 70 per cent. We also renormalize the expected counts to ensure that $\langle N/N_{\rm exp}\rangle = 1$ in order to correct for possible errors in the normalization of the luminosity functions. We find that this choice of renormalization gave the most stable results although the exclusion of empty cells from the renormalization step was necessary to ensure stability (see Section 4.2). In practice the details of this renormalization step do not significantly affect the best-fitting parameters for the variance or relative bias, but they do change the Kolmogorov–Smirnov test probabilities for our models.

This approach to dealing with the selection function for subsets of the data, such as the division into early and late-type galaxies, is vulnerable to systematic errors in the selection function. An example of such an effect would be a surface-brightness term in the selection function for η-typed galaxies, which one would expect would affect early- and late-type galaxies differently. Any systematic error in calculating the expected number of counts in cells for the early- or late-type galaxies will bias results for the non-linearity and stochasticity of the bias function. Such an effect was noted in Blanton (2000), who found that his results were sensitive to excluding the low-redshift part of his sample. Madgwick et al. (2002) detect a large overabundance of spectral types 3 and 4 beyond $z = 0.11$, relative to the predicted $n(z)$ based on the luminosity functions for these spectral types. Although such an observation could be due to aperture effects, a more plausible explanation is the presence of evolution for these spectral type bins. In principle, such evolution could be modelled in our analysis, and we could derive accurate $N_{\rm exp}$ for all late-type galaxies. If there is evolution in these spectral types, however, we may expect that the relative bias could also be evolving over the redshifts used in this analysis. For this reason we have used only η-types 1 and 2 in the analysis.

Fig. 1 shows the division into cells for $\ell = 14\,h^{-1}$ Mpc. The galaxies in each cell of the respective spectral type are shown overlaid on a grey-scale indicating the estimated galaxy density contrast in that cell. The intermediate density contrasts are much less prevalent in the early-type density field, implying that the contrast between clusters and voids is increased, as one would expect. Table 1 shows the number of cells in each cell division together with the median and 16 and 84 per cent percentiles of the distribution of expected counts.

Wild et al. (2004) have also analysed colour-selected volume-limited samples from the 2dFGRS. Unfortunately, luminosity functions for different colour selections from the 2dFGRS have not yet been measured, and so we cannot define the selection functions required to carry out our analyses on colour-selected magnitude-limited samples.

3 THE VARIANCE OF THE COUNTS IN CELLS

One of the most fundamental statistics accessible from the counts in cells is the variance of the counts. This is directly related to the

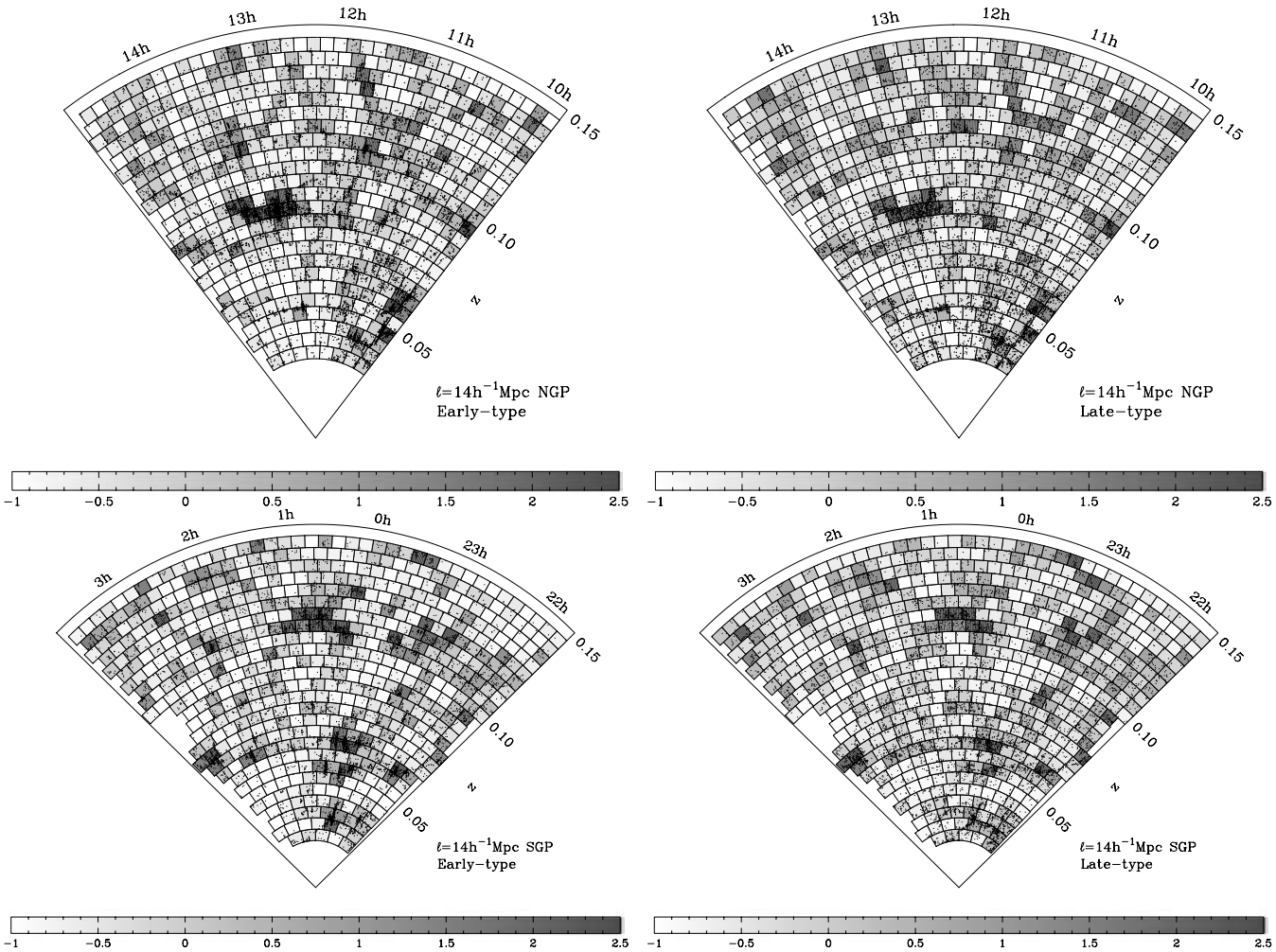

Figure 1. An example of the division into cells for the NGP (top) and SGP (bottom) regions, with cells of side $\ell = 14\,h^{-1}$ Mpc; the cells spanning the central declination of each slice are plotted along with the galaxies of each type (early types on the left, late types on the right) in the cell. The grey-scale indicates the estimated galaxy density contrast as defined in the text.

Table 1. Total number of cells and expected counts (presented as the 16–50–84 per cent percentiles of the distribution).

	NGP			SGP		
ℓ (h^{-1} Mpc)	$N_{\rm cells}$	$N_{\rm exp,E}$	$N_{\rm exp,L}$	$N_{\rm cells}$	$N_{\rm exp,E}$	$N_{\rm exp,L}$
7.0	11056	2.0–3.2–5.7	1.2–2.2–4.7	14593	2.0–3.0–5.0	1.3–2.2–4.5
8.75	5689	2.8–4.5–8.3	1.7–3.0–6.7	7543	2.7–4.1–6.9	1.8–3.0–6.2
10.5	3170	4.0–6.5–11.7	2.5–4.4–9.5	4114	3.8–5.8–9.6	2.6–4.3–8.7
12.25	2026	5.4–8.9–16.4	3.4–6.0–13.3	2567	5.3–7.8–13.1	3.6–5.9–12.0
14.0	1198	8.4–13.1–23.3	5.3–8.8–18.9	1484	7.7–11.4–18.7	5.5–8.8–17.4
17.5	620	13.7–22.8–40.5	8.7–15.7–32.6	729	12.7–18.3–30.0	9.2–14.4–28.3
21.0	336	24.5–39.3–69.4	15.7–26.8–56.0	372	22.0–32.8–52.4	15.8–25.5–49.8
24.5	187	41.3–64.6–117	25.8–43.3–95.6	205	36.0–49.2–80.3	26.6–38.6–75.9
28.0	113	63.1–95.9–175	39.1–63.1–136	117	51.8–77.0–114	38.1–61.0–110
31.5	80	98.1–162–252	61.8–109–197	125	86.3–122–179	62.6–99.5–171
35.0	57	140–226–321	85.2–152–250	101	120–160–260	85.9–125–255
38.5	45	183–278–390	112–184–301	77	143–175–305	104–138–296
42.0	34	248–351–476	156–237–349	53	168–227–388	125–176–369

galaxy autocorrelation function, being equal to the volume average of the correlation function once the contribution of discreteness noise is removed. In later sections we will fit a parametric model to the one-point distribution of the counts in cells as the first stage in a maximum-likelihood approach to fitting for the relative bias; we would expect that an accurate model for the one-point distribution will reproduce the variance of the counts as measured in this section. Furthermore, by comparing the variance between spectral types, we can obtain an estimate of the linear relative bias parameter.

3.1 Predictions for the variance

The real-space and redshift-space correlation functions for the full set of 2dFGRS galaxies have been accurately measured by Hawkins et al. (2003). We have used these results to obtain predictions for the variance of the counts in cells using a number of approaches outlined below.

If we assume that the real-space correlation function is well described by a power-law of the form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \qquad 3.1$$

then we can use the following form for the power spectrum expressed as the variance per ln k:

$$\Delta^2(k) = \frac{2}{\pi}(kr_0)^{\gamma}\Gamma(2-\gamma)\sin\frac{(2-\gamma)\pi}{2}. \qquad (7)$$

For the case of Gaussian spheres of radius $R_{\rm G}$, the variance in spheres, σ^2, is equal to the value of $\Delta^2(k)$ at

$$k = \left[\frac{1}{2}\left(\frac{\gamma-2}{2}\right)\right]^{1/\gamma} R_{\rm G}^{-1}, \qquad (8)$$

but this is also a good approximation to the variance in cubical cells if we take $R_{\rm G} = \ell/\sqrt{12}$, as described in Section 2.2.

More accurate predictions can be obtained by directly integrating the correlation function over a cubical volume of side ℓ to calculate the variance in cells,

$$\sigma^2 = \frac{1}{V^2}\int_V {\rm d}V_1\,{\rm d}V_2\,\xi(r). \qquad (9)$$

We have used this method to calculate predictions based on both the real-space and redshift-space correlation functions of Hawkins et al. (2003). The variance obtained from a volume average of the best-fitting power-law form for the real-space correlation function (Hawkins et al. 2003) is almost identical to that using the scaling relation (equations 7 and 8) as one might expect. We have also calculated redshift-space variance estimates using both a power-law approximation for $\xi(s)$ and a direct interpolation from the data of Hawkins et al. (2003), since $\xi(s)$ is not well approximated by a power law. The estimated variance from the interpolated $\xi(s)$ data (solid lines in Figs 2, 4 and 7) is likely to give the most accurate prediction for the variance of the counts in cells for the combination of spectral types 1 and 2.

3.2 Measuring the variance from the counts in cells

A maximum-likelihood technique for calculating the variance of counts in cells is presented by Efstathiou et al. (1990). The key advantage of this method is that by calculating the variance in individual redshift shells, it does not require a weighting scheme to compensate for the radial selection function. In each redshift shell in our cell division we can compute the statistic

$$S = \frac{1}{M-1}\sum_i (N_i - \bar{N})^2 - \bar{N}, \qquad (10)$$

where the sum extends over the M cells in the shell, and $\bar{N}$ is the mean cell count; N_i are the observed counts in cell i. Note that this technique is based only on the measured counts in cells and does not use the expected counts, $N_{\rm exp}$, calculated in Section 2.2. The expectation value for S is

$$\langle S\rangle = n^2V^2\sigma^2 = \bar{N}^2\sigma^2, \qquad (11)$$

where n is the mean number density and V is the volume of the cells. Note that we have implicitly assumed Poissonian noise in the counts, which we have removed in equation (10) by subtracting $\bar{N}$. The variance of S for the case where the underlying density fluctuations are Gaussian is given by

$${\rm Var}(S) = \frac{2n^2V^2(1+\sigma^2) + 4n^3V^3\sigma^2 + 2n^4V^4\sigma^4}{M}. \qquad (12)$$

Clearly this variance will be underestimated since we have made two key assumptions which are not strictly correct, namely that the underlying fluctuations are Gaussian and that the cells are independent. The effect of correlations between cells was addressed by Broadhurst, Taylor & Peacock (1995). They show that for all but adjacent cells the covariance in the cell counts will be negligible and, within the accuracy to which σ^2 can be calculated, the error in treating even adjacent cells as independent is unimportant. A far more serious concern is with the assumption of Gaussian fluctuations; even on the relatively large scales of this analysis this assumption is far from valid. By using the variance estimator of equation (10) on Monte Carlo realizations of lognormal fields we find that the variance of S is many times larger than expected from equation (12).

This method must be modified to deal with completeness variations in the 2dFGRS, as quantified by the survey mask. Efstathiou et al. (1990) gives the following modified estimator for S for the case where cells are incomplete due to the survey mask:

$$S = A/B,$$

where

$$A = \left(\frac{\ell^3}{M}\sum_i \frac{N_i}{V_i}\right)^2\left[\sum_i\left(N_i - V_i\frac{n_1}{v_1}\right)^2 - n_1\left(1-\frac{v_2}{v_1^2}\right)\right]$$

$$B = \left(\frac{n_1}{v_1}\right)^2\left[v_2 - 2\frac{v_3}{v_1} + \frac{v_2^2}{v_1^2}\right], \qquad (13)$$

where for convenience we have used the quantities n_1, v_1, v_2, v_3, defined as:

$$n_1 = \sum_i N_i,\ v_1 = \sum_i V_i,\ v_2 = \sum_i V_i^2,\ v_3 = \sum_i V_i^3, \qquad (14)$$

where V_i is the usable volume of cell i. This correction is only valid if only a small fraction of a cell is excluded by the mask, so for this test we reject all cells where the fraction of the cell of low completeness (<70 per cent) is less than 30 per cent. We also upweight the counts in cells to compensate for incompleteness, based on the survey mask.

For each shell, j, we define a likelihood,

$$L_j(\sigma) = \frac{1}{\sqrt{2\pi{\rm Var}(S_j)}}\exp\left[\frac{-\left(S_j - n_j^2V^2\sigma^2\right)^2}{2{\rm Var}(S_j)}\right], \qquad (15)$$

which is calculated using equation (13) as an estimator for $n_j^2V^2\sigma^2$. We then minimize with respect to σ the quantity

$$\mathcal{L} \equiv -2\sum_j \ln L_j, \qquad (16)$$

where the sum is over all redshift shells. Note that this differs by a factor 2 compared to $\mathcal{L}$ defined by Wild et al. (2004).

As we discussed above, the variance on the estimator, Var(S), is in fact badly underestimated by equation (12) for realistic non-Gaussian density fluctuations. Although in practice the procedure adopted by Efstathiou et al. (1990) of deriving errors in σ from the likelihood function will not underestimate errors as dramatically as this, since the variation between shells will contribute to the error

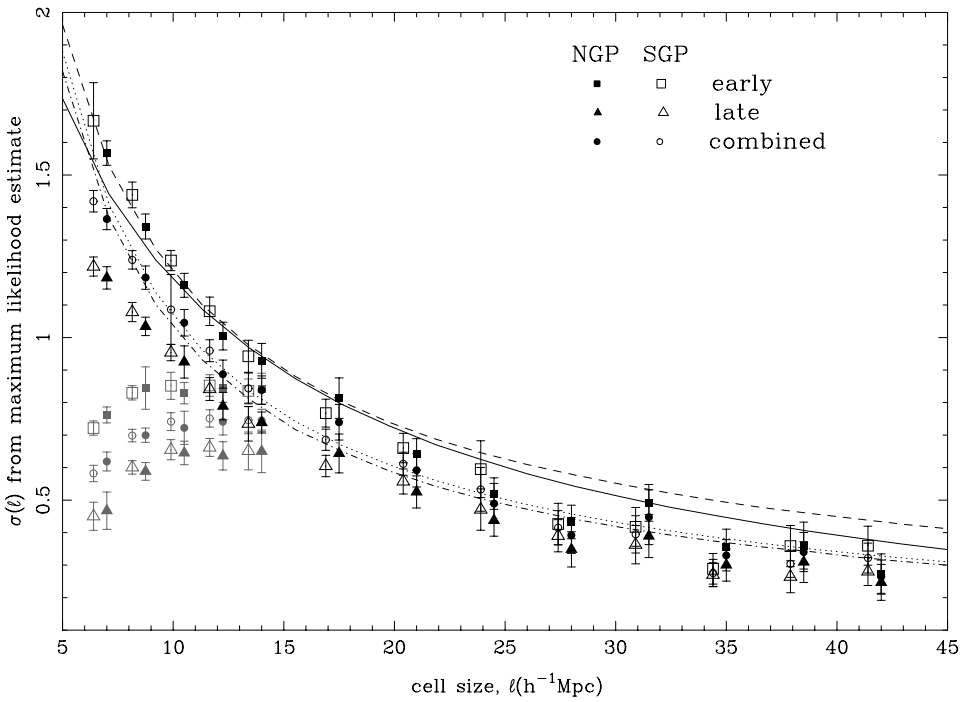

Figure 2. σ as a function of cell size, ℓ, as measured by the maximum-likelihood technique of Efstathiou et al. (1990). Filled symbols are for the NGP region, open symbols are SGP. The results shown are fits to the early-type galaxies (squares), late types (triangles), and to both types combined (circles). Predicted values are overlaid, calculated using a power-law form for $\Delta^2(k)$ (equation 7; dotted line); from the integral over the power-law fit to $\xi(r)$ (dash-dot line), from the integral over the power-law fit to the redshift-space correlation function, $\xi(s)$ (dashed line), and from the integral over the interpolated data table for $\xi(s)$ (solid line). (The points for the SGP are offset by a small amount for clarity.) Points in grey are the measured values when empty cells are removed from the analysis; for $\ell > 14\,h^{-1}$ Mpc this has no effect on the variance measurements. The error bars are derived from Monte Carlo realizations of lognormal models as described in the text.

estimate for σ, we have instead used Monte Carlo realizations of lognormal fields at the appropriate variance to derive more realistic error bars. Even though these errors are model dependent, the density fluctuations are much more closely approximated by a lognormal model than the Gaussian assumption of equation (12).

3.2.1 Cell variances and estimation bias

The variances of the counts in cells calculated using the maximum-likelihood approach are shown in Fig. 2. The figure shows that the measured variances for the combined sample (η-types 1 and 2) lie below the variance predicted from the redshift-space correlation function. The fact that the variances appear to be consistent with the real-space predictions is fortuitous. The variances are underestimated due to a finite volume effect exactly equivalent to the integral constraint, which is discussed in detail in Hui & Gaztañaga (1999). It becomes relevant here because we are using redshift shells which have a rather small volume because of the geometry of the 2dFGRS slabs. Hence the individual shell contributions, S_j in equation (15), can be significantly biased.

Hui & Gaztañaga give a useful analytical approximation for the magnitude of the integral constraint bias in the variance. We express the expected value of our variance estimator as

$$\langle\hat{\sigma}^2\rangle \equiv \sigma^2\left(1+\frac{\Delta_{\sigma^2}}{\sigma^2}\right), \tag{17}$$

where $\Delta_{\sigma^2}/\sigma^2$ is the fractional bias in σ^2. The fractional bias can be approximated by the expression:

$$\frac{\Delta_{\sigma^2}}{\sigma^2} = -\frac{\sigma_V^2}{\sigma^2} + (3-2c_{12})\sigma_V^2, \tag{18}$$

where σ_V^2 is the two-point correlation function averaged over the whole volume in question, which in this case is the volume of a shell, and c_{12} is a coefficient derived from the hierarchical relation. We have used the perturbative value for c_{12} from Bernardeau (1994):

$$c_{12} = 68/21 + \gamma/3. \tag{19}$$

Using this approximation we can correct the S_j and $\mathrm{Var}(S_j)$ in equation (15) to obtain a bias-corrected estimate for σ. In Fig. 3 we show an example of the estimated variances in redshift shells for both the original Efstathiou estimator (equation 13; grey points and horizontal lines) and the bias-corrected version (black points and horizontal lines), for a single scale $\ell = 14\,h^{-1}$ Mpc. The correction for integral constraint bias shifts the maximum-likelihood variance estimator such that our results for the variance of the counts in cells for the combined sample are now consistent with the predicted values from the redshift-space correlation function as shown in Fig. 4. The bias correction also increases the errors on the individual shell variance measurements.

The effect of large-scale structure can clearly be seen in Fig. 3; in particular the noticeable spike around $r = 250\,h^{-1}$ Mpc corresponds to the prominent group of large clusters in the NGP region at around $z = 0.09$. This is the NGP 'hotspot' observed by Baugh et al. (2004). Removing the shells around this value of r does not alter our results appreciably relative to the magnitude of the estimated errors.

4 LOGNORMAL MODELS FOR THE ONE-POINT DISTRIBUTION FUNCTION

Given a model for the one-point density distribution function, we can model the one-point distribution of galaxy counts as being

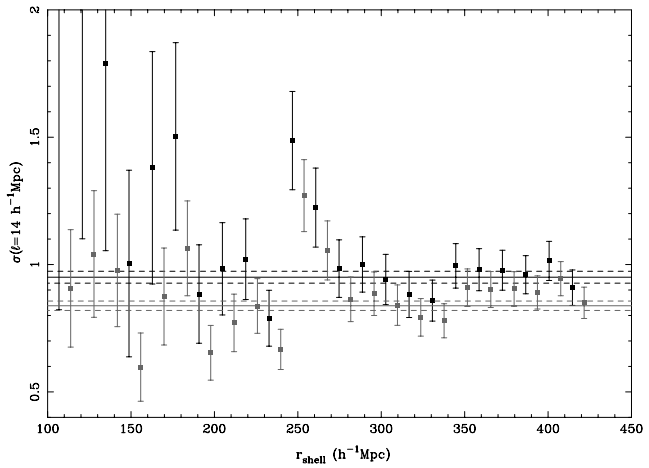

Figure 3. An example of the estimated cell variance in redshift shells compared to the maximum-likelihood value and its associated 1σ error shown in the horizontal black solid and dashed lines. This example is for the $\ell = 14\,h^{-1}$ Mpc cell division of the combined sample (η-types 1 and 2) in the NGP region. The points in grey (offset for clarity) show the same plot before the effects of the integral constraint bias have been corrected for and the grey horizontal solid and dashed lines show the uncorrected maximum-likelihood estimate with its associated 1σ error.

equivalent to a convolution of the density field with Poisson fluctuations of intensity $\lambda = N_{\rm exp}(1+\delta)$, i.e.

$$P(N) = \int {\rm d}\delta\, \frac{N_{\rm exp}^N (1+\delta)^N}{N!} {\rm e}^{-N_{\rm exp}(1+\delta)} f(\delta). \quad (20)$$

We can then use equation (20) as the basis for a maximum-likelihood method to determine the parameters of the best-fitting model for $f(\delta)$. A number of models have been proposed for $f(\delta)$ including the lognormal distribution (Coles & Jones 1991), negative binomial model (Fry 1986; Carruthers 1991; Bouchet et al. 1993; Gaztañaga & Yokoyama 1993), Edgeworth expansion around the Gaussian distribution (Juszkiewicz et al. 1995) and Edgeworth expansion around the lognormal model – the skewed lognormal approximation of Colombi (1994). Ueda & Yokoyama (1996) consider fits of all of the above models to the counts in cells of a low-density CDM N-body simulation. They find that the most satisfactory fit is given by the skewed lognormal model but unfortunately it is not positive definite, making it unsuitable for the maximum-likelihood fitting procedure outlined below. The lognormal model is the most mathematically convenient since the version given below is already normalized in the interval $-1 \leqslant \delta < \infty$ and ensures that $\langle\delta\rangle = 0$.

4.1 Lognormal model fitting

The lognormal model, which we use to fit the one-point distribution function, is given in equation (21), where $x = \ln(1+\delta) + \sigma_{\rm LN}^2/2$,

$$f(\delta)\,{\rm d}\delta = \frac{{\rm d}\delta}{\sigma_{\rm LN}\sqrt{2\pi}(1+\delta)} \exp\left(-\frac{x^2}{2\sigma_{\rm LN}^2}\right). \quad (21)$$

It should be noted that we have here followed the notation of Coles & Jones (1991) in that $\sigma_{\rm LN}^2$ is the variance of the Gaussian model from which the lognormal is derived by transformation – note that Wild et al. (2004) use ω^2 for the same parameter. The variance for the lognormal model is given by

$$\langle\delta^2\rangle = \exp\left(\sigma_{\rm LN}^2\right) - 1. \quad (22)$$

To ensure that the results of this section can be easily compared to those measuring the variance in the previous section we have used the above relation to transform our variances from $\sigma_{\rm LN}$ to $\sigma(\ell) \equiv \sqrt{\langle\delta^2\rangle}$.

Using this model, we define a likelihood for each cell i, as

$$L_i \equiv P(N_i|\sigma_{\rm LN}),$$

in which the probability of observing N_i galaxies in cell i is determined by equation (20). We then find $\sigma_{\rm LN}$ for the best-fitting lognormal model by minimizing with respect to $\sigma_{\rm LN}$ the quantity

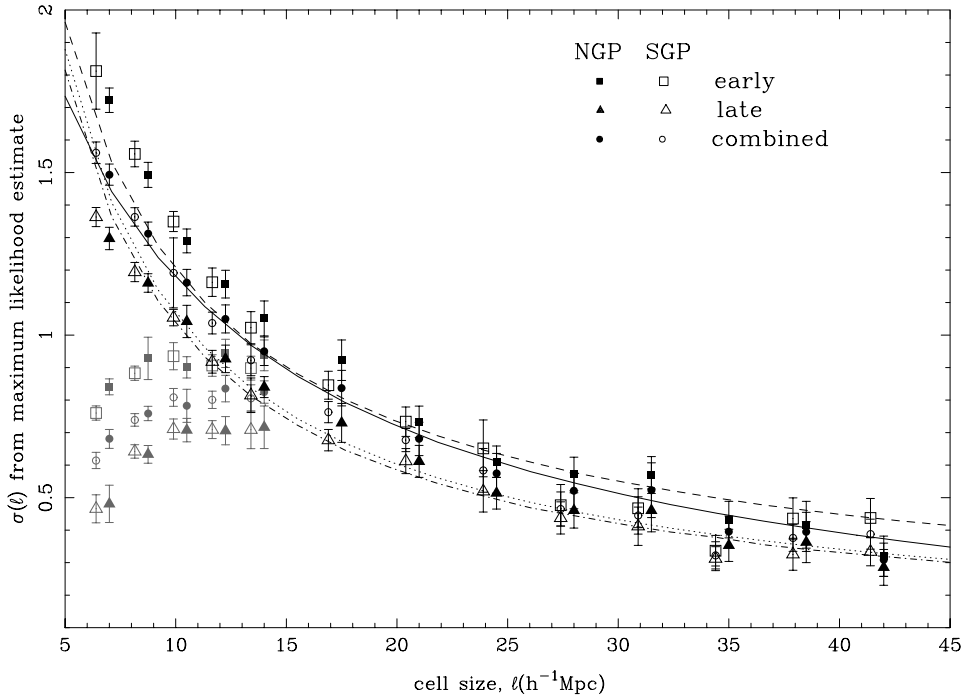

Figure 4. The same as Fig. 2 but using the approximation of Hui & Gaztañaga to correct for the integral constraint bias.

$$\mathcal{L} \equiv -2 \sum_i \ln L_i, \tag{23}$$

where the sum is over all cells, as defined previously in equation (16). Again, this differs by a factor 2 compared to $\mathcal{L}$ defined by Wild et al. (2004)

4.2 Lognormal model variances and empty cells

We have fitted lognormal models for the one-point density distribution to both the early- and late-type galaxy populations, and to the whole sample. Examples of the actual one-point distributions for a cell size of $\ell = 17.5\,h^{-1}$ Mpc for early and late types, as well as for the full data set (η-types 1 and 2), are shown in Fig. 5, and one-point distributions as a function of cell size ℓ for the SGP region are shown in Fig. 6, where we have used $N/N_{\rm exp}$ as an estimator for the overdensity in the same manner as equation (27), with the $N_{\rm exp}$ calculated using equation (6). When we fit models to all cells in the data, including empty cells, we recover variances which are many sigma above both the variance predicted from $\xi(s)$ and the measurements of the variance in cells of Section 3. For example, the best-fitting lognormal model for $\ell = 10.5\,h^{-1}$ Mpc gives a variance in early types of $\sigma_{\rm E} \sim 1.9$, whereas our previous measurements give $\sigma_{\rm E} \sim 1.1$.

Some insight into why empty cells have such a dramatic effect can be gained by examining the number of empty cells in the data on scales where they become significant ($\ell \lesssim 15\,h^{-1}$ Mpc). Table 2 shows the number of empty cells in the full survey data for each type and for η-types 1 and 2 combined, along with the 10 and 90 per cent percentiles of the distribution of empty cells in a large number of Monte Carlo realizations of lognormal models with variances matching our previous measurements of the variance in cells. A Poisson-sampled lognormal model with a realistic variance cannot explain the number of empty cells in the data; the large excess of empty cells will increase the variance of the best-fitting lognormal model. Table 2 also shows the number of empty cells which are found when we analyse the Hubble volume mock catalogues appropriately sampled to match the survey data (Cole et al. 1998; Norberg et al. 2002b) using the same method. The number of empty cells in the real and mock data on small scales exceeds the number predicted by the lognormal model, which suggests that a Poisson-convolved lognormal model is not a good fit to the actual density distribution function. In the real data this discrepancy is more pronounced and extends to smaller scales.

A possible explanation for this effect is that the luminosity function of the 2dFGRS galaxies is dependent on the local density. Croton et al. (2004c) investigate this possibility and find that L_* increases as a function of the overdensity in $8\,h^{-1}$ Mpc spheres. This effect would mean that the appropriate intensity for the Poisson process in each cell, rather than being given by $\lambda = N_{\rm exp}(1+\delta)$, would change to $\lambda = N_{\rm exp}(1+\delta)\psi(\delta)$, where $\psi(\delta)$ is a correction factor which also depends on the redshift of the cell. Hence if ψ is significantly less than unity for very low overdensities, the observed number of empty cells will be significantly greater than expected.

A simple solution to this problem is to fit lognormal models to the counts in cells excluding empty cells. Clearly this will cause variance estimates to be biased on scales where empty cells are significant ($\ell \lesssim 15\,h^{-1}$ Mpc). We have estimated the magnitude of this bias both by measuring the effect of excluding empty cells from Monte Carlo realizations of lognormal models, and by considering the bias introduced into the variance measurements of Section 3 on small scales when we excluded empty cells (grey points shown in Fig. 4); both methods give similar results.

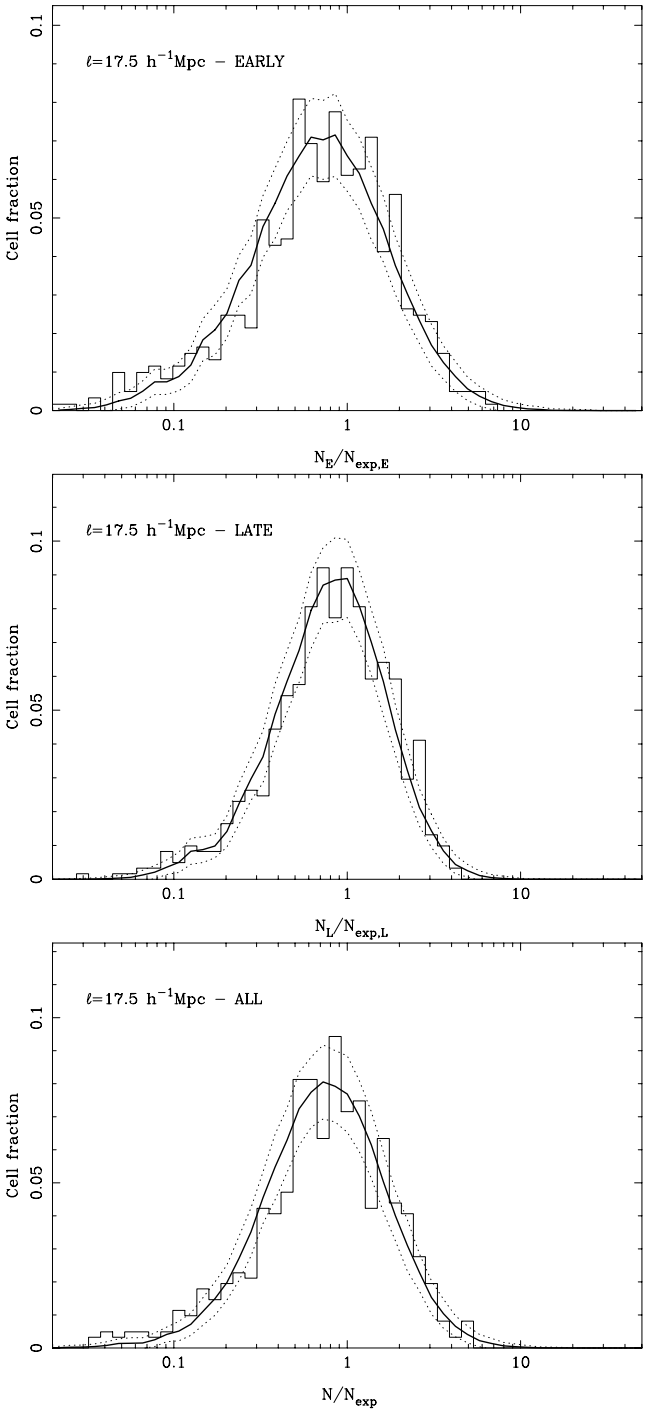

Figure 5. An example of the one-point distribution function for counts in cells of size $\ell = 17.5\,h^{-1}$ Mpc in the NGP region for early-type galaxies (top), late types (middle), and η-types 1 and 2 (bottom). The average of a large number of realizations of the best-fitting lognormal models, convolved with the same $N_{\rm exp,i}$ as the data, are also shown, together with their 1σ spread. The higher variance of the early-type distribution is very clear.

The values of σ from the lognormal fits to the early- and late-type subsets, as well as the full catalogue, are shown in Fig. 7. The black points show the measurements excluding empty cells and corrected for the bias; points in grey for small ℓ show the original measurements illustrating the bias introduced by excluding empty cells in a similar manner to Fig. 2. Since the likelihood function is

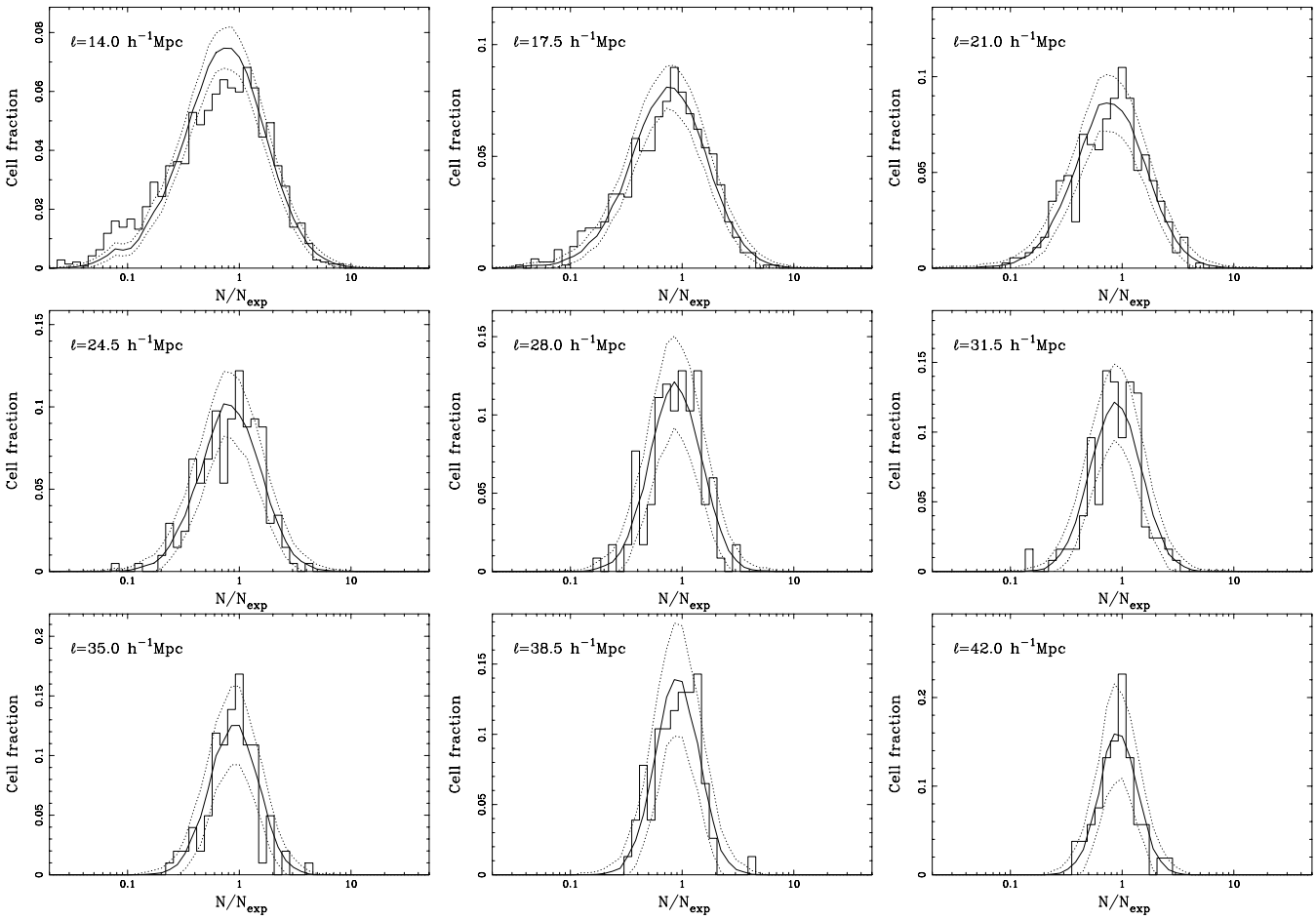

Figure 6. The one-point distribution function for counts in cells of the combined sample for the SGP region over a range of cell sizes, from left–right and top–bottom, $\ell = 14$, 17.5, 21, 24.5, 28, 31.5, 35, 38.5, 42 h^{-1} Mpc. The average of a large number of realizations of the best-fitting lognormal models, convolved with the same $N_{\mathrm{exp},i}$ as the data, are also shown, together with their 1σ spread. (Note that the y-axis changes between plots in this figure.)

well approximated by a χ^2 distribution with one degree of freedom we have derived 1σ errors by considering the values for σ_{LN} at which $\mathcal{L} = \mathcal{L}_{\mathrm{min}} + 1$. We have also obtained Monte Carlo estimates of the errors by using the procedure outlined above to fit a large number of models generated by randomly drawing the galaxy density contrast, δ, from a lognormal distribution with the best-fitting value of σ and then generating model counts from a Poisson distribution with intensity $\lambda = (1 + \delta)N_{\mathrm{exp}}$, using the same expected counts in cells as were calculated for the data. The magnitudes of errors using both methods are identical.

As another measure of the goodness of fit for these models we show in Table 3 probabilities obtained by the application of a Kolmogorov–Smirnov (KS) test to the distribution of N/N_{exp}. The KS test is not ideal for a number of reasons; it is rather insensitive to variation in the tails of distributions, which is where we would expect the lognormal model will have the most difficulty in matching the data. Strictly speaking the estimate of P_{KS} which we use here is no longer valid once the data have been used to fix any free parameters of the model, although any effects should be small since the number of data points we use is very much larger than the number of free parameters.

The KS test probabilities indicate that the lognormal model is an acceptable fit to the data on large scales for both early and late types, as well as for the combined sample. None of the values for P_{KS} for $\ell \geqslant 17.5\,h^{-1}$ Mpc are sufficiently low to exclude the model at a high level of confidence. In general the early-type distribution is well-fitted by lognormal models to smaller scales than the distribution of late types or the combined galaxy distribution. On scales smaller than $\ell = 10.5\,h^{-1}$ Mpc a Poisson-convolved lognormal model is not a satisfactory fit to any of the distributions. We expect, based on the findings of Ueda & Yokoyama (1996), that the lognormal model will not adequately describe the data in the non-linear regime. Our results are consistent with this expectation since the variance is significantly higher than unity on the scales at which the lognormal model becomes unsatisfactory.

Although the values of σ for types 1 and 2 combined, shown in Fig. 7, are broadly consistent with the cell variance derived from $\xi(s)$ in Section 3.1, there is a systematic trend for the fitted values of $\sigma(\ell)$ to be higher than predicted. The magnitude of this effect is between 10 and 15 per cent, as shown in the ratio plot of Fig. 8; this corresponds to around a 2σ effect. Maximum-likelihood fits to lognormal models giving variances consistent with predictions are obtained with the introduction of an additional weighting factor to the likelihood defined in equation (23):

$$\mathcal{L}' = -2 \sum_i \frac{N_i}{N_{\mathrm{exp},i}} \ln L_i. \qquad (24)$$

Table 2. Number of empty cells in full survey data compared to the 10 and 90 per cent percentiles of the empty cells in lognormal models matching the measured variances in cells. The rightmost column shows the number of empty cells when the Hubble volume mock catalogues are analysed in the same way.

	early-type galaxies		late-type galaxies		types 1 & 2		HV mocks
ℓ (h^{-1} Mpc)	$N_{\rm empty}$	Models	$N_{\rm empty}$	Models	$N_{\rm empty}$	Models	$\langle N_{\rm empty}\rangle$
7.0	14103	7482–7670	13495	8050–8237	10527	4243–4406	8029
8.75	5238	2633–2743	4757	2803–2909	3405	1276–1355	2063
10.5	1863	884–961	1590	962–1040	1005	354–405	478
12.2	709	321–370	601	339–387	333	99–128	109
14.0	211	89–114	163	88–113	83	19–31	17

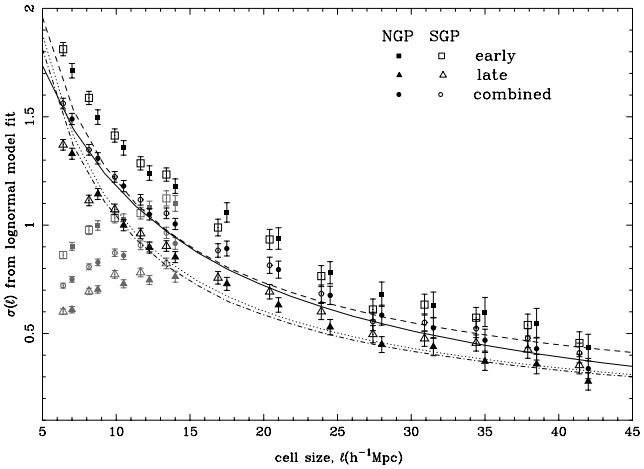

Figure 7. σ from the best-fitting lognormal model as a function of cell size, ℓ. Filled symbols are for the NGP region, open symbols are SGP (offset as previously). The results shown are fits to the early-type galaxies (squares), late types (triangles), and to both types combined (circles). Predictions are overlaid as in Fig. 2. The results are based on counts in cells with empty cells removed from the analysis. The small-scale results (for $\ell \leqslant 14\,h^{-1}$ Mpc) are corrected for the bias resulting from excluding empty cells which causes the variance to be underestimated, as illustrated by the grey points (see text).

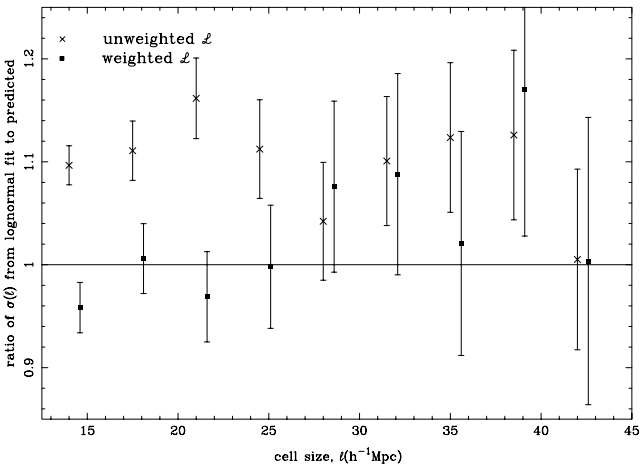

Figure 8. Ratio plot of $\sigma(\ell)$ from the best-fitting lognormal model compared to the predicted value integrating $\xi(s)$ over cells. The crosses show the unweighted maximum-likelihood results for η-types 1 and 2, with NGP and SGP regions combined. Filled squares (offset) show the results when an additional weighting is applied to giving more weight to more overdense cells.

Table 3. Best-fitting lognormal model parameters and KS test probabilities for the lognormal model fit. The parameters for low ℓ cell divisions are derived from the data excluding empty cells and corrected for the resultant bias as described in the text.

ℓ (h^{-1} Mpc)	$\sigma_{\rm LN,E}$	$P_{\rm KS}$	$\sigma_{\rm LN,L}$	$P_{\rm KS}$	$\sigma_{\rm LN,all}$	$P_{\rm KS}$
7.0	1.20	6e-14	1.05	7e-55	1.12	2e-99
8.75	1.12	0.001	0.94	3e-18	1.03	4e-47
10.5	1.05	0.229	0.88	0.002	0.97	5e-15
12.2	0.99	0.268	0.81	0.134	0.90	0.001
14.0	0.95	0.105	0.77	0.183	0.87	0.015
17.5	0.84	0.292	0.67	0.442	0.77	0.212
21.0	0.78	0.670	0.61	0.193	0.70	0.332
24.5	0.68	0.291	0.53	0.710	0.61	0.414
28.0	0.59	0.992	0.45	0.695	0.53	0.987
31.5	0.57	0.829	0.44	0.988	0.51	0.962
35.0	0.53	0.553	0.41	0.625	0.48	0.817
38.5	0.48	0.195	0.39	0.308	0.44	0.965
42.0	0.42	0.850	0.32	0.925	0.37	0.825

The results of applying the above weighting factor are shown by the filled squares in Fig. 8. In effect this modified likelihood gives more weight to the most dense regions, and suggests that the lognormal model is more appropriate to describe the density distribution of high-density regions. This would be consistent with our observation that in general the early type distribution is better fitted by a lognormal model than the distribution of late types, since early types are more prevalent in dense regions. We have tested the weighting scheme on model data based on a lognormal model and verified that it does not underestimate σ in this case; the fact that a discrepancy such as this exists indicates that the lognormal model is not a completely satisfactory model for the one-point distribution functions.

5 THE RELATIVE BIAS

We are now in a position to consider the characteristics of the joint distribution of the counts in cells. We postulate a smoothed density contrast field for galaxies $\delta_{\rm g}$, which can be related to the density field of dark matter, δ, using the general biasing framework of Dekel & Lahav (1999):

$$\delta_{\rm g} = b(\delta)\delta + \epsilon, \tag{25}$$

which in principle is able to deal with both non-linearity and stochasticity, parametrized by the function $b(\delta)$ and ϵ, respectively. We further assume that a similar relationship holds independently for the separate spectral types; in other words we consider early and late types with their own separate smoothed density fields denoted by $\delta_{\rm E}$ and $\delta_{\rm L}$, respectively. Then we can specify the relative bias between the density fields analogously to equation (25):

$$\delta_{\rm L} = b(\delta_{\rm E})\delta_{\rm E} + \epsilon_{\rm L}. \tag{26}$$

We have taken two approaches to quantifying the relative bias. Our first method considers an estimate of the galaxy density contrast for each spectral type in each cell, i, which we denote as

$$g_{\mathrm{E},i} = N_{\mathrm{E},i}/N_{\mathrm{E,exp},i} - 1, \tag{27}$$

for the early-type galaxies, and analogously for late types.

Under the assumption that the density fields of both spectral types are related to the underlying dark matter field by a linear bias factor and that the only scatter is due to the Poissonian scatter caused by galaxy discreteness, we have for the early-type galaxies

$$g_{\mathrm{E},i} = b_{\mathrm{E}}\delta_i + \epsilon_{\mathrm{E},i}, \tag{28}$$

where $\epsilon_{\mathrm{E},i}$ is the Poisson noise for the early types in cell i. $g_{\mathrm{L},i}$ can be defined similarly. This is the basis for the null test described in Section 5.2.

In the second approach we attempt to fit to the joint distribution of the underlying smoothed density fields,

$$f(\delta_{\mathrm{E}}, \delta_{\mathrm{L}}) = f(\delta_{\mathrm{L}}|\delta_{\mathrm{E}})f(\delta_{\mathrm{E}}). \tag{29}$$

We follow the example of Dekel & Lahav (1999) and adopt a general description for relative bias:

$$b(\delta_{\mathrm{E}})\delta_{\mathrm{E}} \equiv \langle\delta_{\mathrm{L}}|\delta_{\mathrm{E}}\rangle = \int d\delta_{\mathrm{L}}\ f(\delta_{\mathrm{L}}|\delta_{\mathrm{E}})\delta_{\mathrm{L}}. \tag{30}$$

The function $b(\delta_{\mathrm{E}})$ is characterized by Dekel & Lahav (1999) by defining the following moments:

$$\hat{b} \equiv \frac{\langle b(\delta_{\mathrm{E}})\delta_{\mathrm{E}}^2\rangle}{\sigma_{\mathrm{E}}^2}, \quad \tilde{b}^2 \equiv \frac{\langle b^2(\delta_{\mathrm{E}})\delta_{\mathrm{E}}^2\rangle}{\sigma_{\mathrm{E}}^2}, \tag{31}$$

where $\sigma_{\mathrm{E}} \equiv \sqrt{\langle\delta_{\mathrm{E}}^2\rangle}$, as we have used previously throughout this paper.

A random biasing field, $\boldsymbol{\epsilon}$, is defined in our case as

$$\boldsymbol{\epsilon} \equiv \delta_{\mathrm{L}} - \langle\delta_{\mathrm{L}}|\delta_{\mathrm{E}}\rangle, \tag{32}$$

and the average biasing scatter, σ_{b},

$$\sigma_{\mathrm{b}}^2 \equiv \frac{\langle\boldsymbol{\epsilon}^2\rangle}{\sigma_{\mathrm{E}}^2}. \tag{33}$$

These moments separate the effects of non-linearity and stochasticity of the bias relation. Linear bias is often described by the ratio of variances of the density field. In fact this bias parameter, b_{var}, is a mixture of non-linear and stochastic effects and can be expressed in terms of the above moments as:

$$b_{\mathrm{var}}^2 = \tilde{b}^2 + \sigma_{\mathrm{b}}^2, \quad \text{where} \quad b_{\mathrm{var}} = \sigma_{\mathrm{L}}/\sigma_{\mathrm{E}}. \tag{34}$$

The bivariate lognormal model considered by Wild et al. (2004) explicitly includes a stochastic term in the relative bias, which also effectively introduces a non-linear term.

5.1 Direct estimates of b_{var}

We estimated the relative bias from the cell variances calculated using the Efstathiou estimator, after correcting the estimator for the integral constraint bias, shown in Fig. 4. We have used $1/b_{\mathrm{var}}$ to facilitate comparison with other papers where the relative bias is generally defined as a ratio $b_{\mathrm{rel}} = b_{\mathrm{E}}/b_{\mathrm{L}}$. The results are plotted in Fig. 9 which shows that the relative bias is consistent with a constant $1/b_{\mathrm{var}} = 1.25 \pm 0.05$ for both the NGP and SGP regions and for all cell sizes. In calculating the error bars we have used the measured error bars at each value of ℓ and made the assumption that points adjacent in ℓ in Fig. 4 are perfectly correlated. For comparison we have plotted the relative bias, $b_{\mathrm{rel}} = \sqrt{\xi_{\mathrm{E}}(r)/\xi_{\mathrm{L}}(r)}$, from the real-space correlation functions per η-type of Madgwick et al. (2003b), where we have converted the separation, r, to ℓ by assuming $r \equiv R_{\mathrm{T}}$ and using equation (1). This is intended mainly for illustration since, following the bias framework of Dekel & Lahav (1999), the bias parameter formed from the ratio of correlation functions is not, in general, equivalent to b_{var}. Clearly, though, they are consistent within the rather large errors from the correlation function estimate.

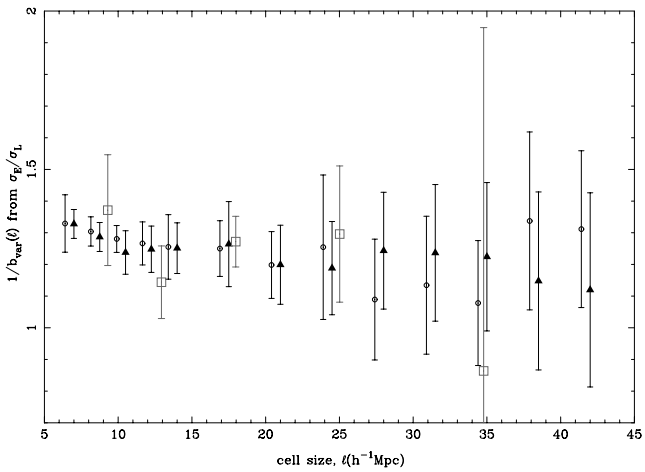

Figure 9. The relative bias estimated from the ratio of $\sigma_{\mathrm{E}}/\sigma_{\mathrm{L}}$ from the variance estimator of Efstathiou et al. (1990) for the NGP (filled triangles) and SGP (open circles) regions. The relative bias predictions from the real-space correlation functions per η-type of Madgwick et al. (2003) are shown in grey.

We have also calculated the relative bias from the variances of the lognormal model fits to the early- and late-type one-point distributions as shown in Fig. 7,

$$b_{\mathrm{var}} = \sqrt{\frac{\exp\left(\sigma_{\mathrm{LN,E}}^2\right) - 1}{\exp\left(\sigma_{\mathrm{LN,L}}^2\right) - 1}}. \tag{35}$$

The results are shown in Fig. 10, again compared to the results of Madgwick et al. (2003b). The relative bias factor is again consistent with a scale-invariant bias and we derive a value of $1/b_{\mathrm{var}} = 1.28 \pm 0.05$ from the $\ell > 14\,h^{-1}$ Mpc data again assuming correlation of adjacent bins in ℓ from Fig. 7.

5.2 The Tegmark 'null-buster' test

Tegmark (1999) describes a simple 'null buster' test, based on a generalized χ^2 statistic, to test the assumption that the density fields traced by galaxies of two different spectral types can be related by a simple deterministic linear biasing prescription. This test has been used by Seaborne et al. (1999) to compare the PSCz and Stromlo–APM redshift surveys.

If we first assume that the estimated galaxy density contrasts for each spectral type, g_{E} and g_{L}, are related to an underlying dark matter density field by the prescription of equation (28), then we can construct the difference map:

$$\Delta\boldsymbol{g} \equiv \boldsymbol{g}_{\mathrm{E}} - f\boldsymbol{g}_{\mathrm{L}}, \tag{36}$$

for different values of the relative bias factor $f = b_{\mathrm{E}}/b_{\mathrm{L}}$.

If the deterministic linear bias model is valid, then for the correct value of f, the relative bias factor, $\Delta\boldsymbol{g}$ will consist merely of Poisson noise, which will have a covariance matrix given by Tegmark &

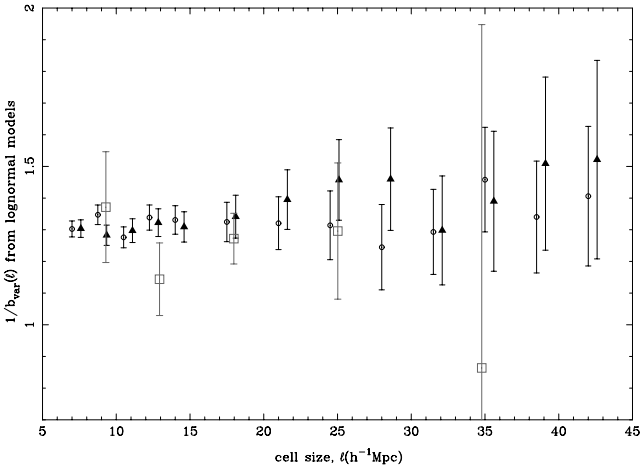

Figure 10. The relative bias estimated from the parameters of the best-fitting lognormal models for the early- and late-type one-point distributions, for the NGP (filled triangles) and SGP (open circles) regions. The relative bias predictions from the real-space correlation functions per η-type of Madgwick et al. (2003) are shown in grey.

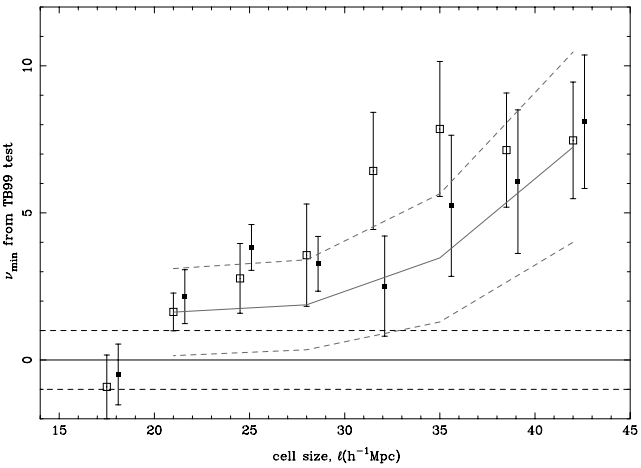

Figure 11. Variation of ($\nu_{\rm min}$) from the Tegmark test with scale. The solid and dashed lines show the expected value and 1σ variation for results consistent with a linear bias relation. The solid and open squares show the average $\nu_{\rm min}$ and its 1σ scatter measured over a number of separate cell divisions obtained by shifting the original divisions, for the NGP and SGP, respectively. The grey solid and dashed curves show, respectively, the average and 1σ spread of $\nu_{\rm min}$ for models including the effects of the selection function variations on scales $< \ell$.

Bromley (1999) as

$$\boldsymbol{N} \equiv \langle \Delta \boldsymbol{g} \Delta \boldsymbol{g}^t \rangle = \delta_{ij} \left[1/N_{\mathrm{E,exp},i} + f^2 (1/N_{\mathrm{L,exp},i}) \right]. \quad (37)$$

This equation for the covariance matrix is valid for density fluctuations $\lesssim 1$; following Seaborne et al. (1999) we therefore exclude cells with $g_{\rm E} > 1$ or $g_{\rm L} > 1$.

Since we are testing the null hypothesis that $\langle \Delta \boldsymbol{g} \Delta \boldsymbol{g}^t \rangle = \boldsymbol{N}$, we can define $\chi^2 = \Delta \boldsymbol{g}^t \boldsymbol{N}^{-1} \Delta \boldsymbol{g}$. If the null hypothesis is correct, the quantity $\nu = (\chi^2 - N_{\rm c})/\sqrt{2N_{\rm c}}$, where $N_{\rm c}$ is the number of cells, has an expectation of zero and standard deviation of one. We can therefore interpret ν as a measure of the significance with which the null hypothesis is ruled out.

In the case where there is extra signal, $\boldsymbol{S}$, in the covariance matrix of the difference map, so that $\langle \Delta \boldsymbol{g} \Delta \boldsymbol{g}^t \rangle = \boldsymbol{N} + \boldsymbol{S}$, the generalized χ^2 statistic (Tegmark 1999) is a more powerful way to rule out the null hypothesis:

$$\nu \equiv \frac{\Delta \boldsymbol{g}^t \boldsymbol{N}^{-1} \boldsymbol{S} \boldsymbol{N}^{-1} \Delta \boldsymbol{g} - \mathrm{Tr}\,(\boldsymbol{N}^{-1}\boldsymbol{S})}{[2\ \mathrm{Tr}\,(\boldsymbol{N}^{-1}\boldsymbol{S}\boldsymbol{N}^{-1}\boldsymbol{S})]^{1/2}}. \quad (38)$$

If there are any deviations from deterministic linear bias we would expect these to be correlated with large-scale structure. We therefore choose the matrix $\boldsymbol{S}$ to be the covariance between cell overdensities calculated using the redshift-space correlation function calculated by Hawkins et al. (2003), i.e. the volume average of $\xi(s_{ij})$ over cells i and j. The value of ν depends only on the shape of $\boldsymbol{S}$, not on its amplitude.

Note that this test uses a Gaussian approximation for the Poisson fluctuations described by the covariance matrix $\boldsymbol{N}$, so we apply an additional cut on cells where this will be particularly inaccurate, and reject cells with $N_{\rm exp} \leqslant 10$. This value for the cut in $N_{\rm exp}$ is a compromise between a value which renders the error term from using the Gaussian approximation negligible and the necessity of not removing too many cells. For this reason we do not apply this test on scales with $\ell \leqslant 17.5\,h^{-1}$ Mpc. We have reduced the redshift range to $0.03 < z \leqslant 0.12$ in order that our cut in $N_{\rm exp}$ does not introduce any systematic effects.

The minimum values of ν for a range of values for the cell size ℓ are shown in Fig. 11. We also plot error bars which are derived from applying the test to cell divisions which are shifted by up to $\ell/\sqrt{2}$ parallel to the RA –z plane, relative to the original division. Clearly the error bars resulting from such an approach will underestimate the true errors. In general the values of $\nu_{\rm min}$ are not consistent with a linear and deterministic relative bias; indeed the significance of this detection increases as ℓ increases.

5.2.1 *Interpretation of* $\nu_{\rm min}$

The values of $\nu_{\rm min}$ presented in Fig. 11 rule out the simple deterministic linear bias model of equation (28) at high significance. The level of significance suggested by $\nu_{\rm min}$ should, however, be treated with some caution, in particular since $\nu_{\rm min}$ increases with increasing ℓ, which would be particularly counterintuitive if interpreted as indicating non-linearity and/or stochasticity which was increasing with scale. The largest cells contain a large number of galaxies so the Poisson noise is small, but as there are so few cells for these values of ℓ, the results will be very sensitive to any systematic effects in the counts, or in the determination of $N_{\rm exp}$. The extremely high values for $\nu_{\rm min}$ in Tegmark & Bromley (1999) when this technique was applied to the LCRS are also an indication that $\nu_{\rm exp}$ cannot be naively interpreted as the number of 'sigma' at which linear deterministic bias is rule out by the data.

A possible instrumental effect which may exaggerate the signal from this test on large scales is an interaction between complex relative bias on small scales and a selection function varying on scales less than the cell division. Differences in the distribution of early- and late-type galaxies on small scales, such as the morphology–density relation, could lead to a significant fraction of the galaxies of one type preferentially residing in a region of the cell in which the selection function differs significantly from the cell average, due, for example, to one of the drill holes in the survey mask. This effect would cause the observed covariance of $\Delta \boldsymbol{g}$ to be enhanced relative to what would be expected from simple Poisson noise as described by equation (37). Note that in this example the observed non-linearity/stochasticity at large ℓ is produced as an effect of non-linearity/stochasticity on smaller scales.

To test to what extent an effect such as the above could be responsible for the results for $\nu_{\rm min}$, we consider a simple model including small-scale stochasticity and the small-scale variations of the selection function. We first generate linear bias models matching the data by drawing $\delta_{\rm E}$ from the best-fitting lognormal model and applying a linear bias function to obtain $\delta_{\rm L}$. We then generate a parent number of galaxies in each cell by Poisson sampling the density field with a constant sampling rate assuming all the cells are 100 per cent complete and with $n(z)$ set to be a constant equal to the maximum $n(z)$ at the mean redshift of the survey. The parent galaxies are distributed within each cell using a modified Rayleigh–Lévy flight model (see Peebles 1980, section 62) matching the correlation function. We have modified the original Rayleigh–Lévy flight model so that the lacuniarity of the process is more realistic – in effect the voids in our clustered point process are less empty. We then select or reject the parent galaxies based on the selection function at the location of each galaxy. The point processes for early- and late-type models are independent.

The grey solid and dashed lines in Fig. 11 show the average and 1σ spread of $\nu_{\rm min}$ when the Tegmark test is applied to our Rayleigh–Lévy flight models. Qualitatively, the models appear to reproduce the effect seen in our results. However, our simple model is in some sense an extreme case: since the two galaxy distributions are completely independent on scales $< \ell$, our models contain a maximum degree of stochasticity on all scales less than the model cell division. A realistic model for deterministic linear bias would be very unlikely to give values for $\nu_{\rm min}$ as high as this, even if instrumental effects such as the above were included. Hence, even though the actual value of $\nu_{\rm min}$ cannot be interpreted as an accurate reflection of the degree of non-linearity or stochasticity present, we conclude that complex bias must to some extent be present even on large scales.

5.2.2 Linear relative bias parameters

We plot f from the best-fitting linear bias models in Fig. 12, again with errors derived from cell shifts. Note that the values of f from this test are not strictly comparable to the other values quoted in this paper except in the case where a deterministic linear bias model is an exact representation of the data, since we have of necessity imposed a cut on $\delta_{\rm E}$. However we expect that on large scales this approximation will be close enough for comparison to be instructive. If we use the slightly larger error bars derived from our Monte Carlo realizations of Rayleigh–Lévy flight models, and assume measurements of f in adjacent bins are correlated, we obtain $f = 1.28 \pm 0.03$ for the NGP and $f = 1.16 \pm 0.03$ for the SGP. It is notable that the best-fitting linear bias factors for the NGP and SGP regions do not seem to be consistent. Averaging the two regions we find a value for f which is consistent with our measurements of $b_{\rm var}$ presented in the previous section. A more relevant comparison is with the maximum-likelihood measurements of the linear bias parameter, which we present in the following section, where we find a similar discrepancy between NGP and SGP regions which we discuss more fully in Section 6.2.

5.3 Fitting the joint counts in cells

The results of the Tegmark test clearly indicate that a deterministic linear bias model is not a satisfactory model for the relative bias between early- and late-type galaxies, although we have indicated some shortcomings of the test which may cause it to exaggerate the significance with which such a model is ruled out on large scales.

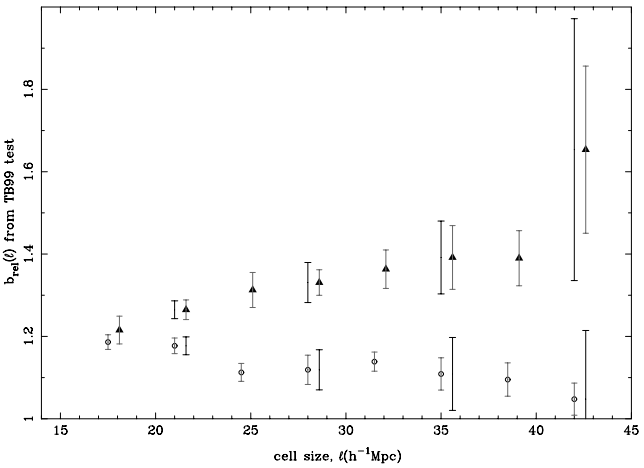

Figure 12. The value of the relative bias factor, $b_{\rm rel}$, giving the minimum value for ν in the Tegmark test. The solid triangles and open circles show the average $b_{\rm rel}$ and its 1σ scatter measured over a number of separate cell divisions obtained by shifting the original divisions, for the NGP and SGP, respectively. The black error bars are a more realistic estimate for the errors based on models including the effects of selection function variations on scales $< \ell$.

However, the test does not reveal any further details of the nature of the relative bias between galaxies of different spectral type. Blanton (2000) describes a more direct approach to measuring the relative bias and applies it to the Las Campanas Redshift Survey (LCRS). The basis of this approach is a maximum-likelihood fit to the joint counts in cells, $P(N_{\rm E}, N_{\rm L})$, which is simply the joint probability of the density fields convolved with Poisson distributions.

If we convolve equation (29) with the expected Poissonian scatter we derive the following joint probability for the counts:

$$P(N_{\rm E}, N_{\rm L}) = \int {\rm d}\delta_{\rm E}\, \frac{\lambda_{\rm E}^{N_{\rm E}}}{N_{\rm E}!} {\rm e}^{-\lambda_{\rm E}} f(\delta_{\rm E}) \times \int {\rm d}\delta_{\rm L}\, \frac{\lambda_{\rm L}^{N_{\rm L}}}{N_{\rm L}!} {\rm e}^{-\lambda_{\rm L}} f(\delta_{\rm L}|\delta_{\rm E}), \quad (39)$$

where

$$\lambda_{\rm E} = N_{\rm E,exp}(1 + \delta_{\rm E}),$$

and similarly for the late types.

We use equation (39) to define the likelihood as a function of $\sigma_{\rm LN,E}$ and a model for the relative bias $f(\delta_{\rm L}|\delta_{\rm E})$. We then find the maximum likelihood using a downhill simplex method (Press et al. 1992) and hence estimate the best-fitting relative bias.

5.3.1 Deterministic bias models

The conditional density distribution function, $f(\delta_{\rm L}|\delta_{\rm E})$, can in principle describe completely the relationship between the galaxy density fields, including any non-linearity or stochasticity. In this paper we have concentrated on bias models in which the density field of the late types is related to that of the early types in a deterministic manner, i.e. $f(\delta_{\rm L}|\delta_{\rm E})$ can be expressed in the form

$$f(\delta_{\rm L}|\delta_{\rm E}) = \delta_{\rm D}[\delta_{\rm L} - b(\delta_{\rm E})\delta_{\rm E}], \quad (40)$$

where $\delta_{\rm D}$ is the Dirac delta function. The possibility that the bias relation may exhibit additional scatter above the Poisson fluctuations is considered by Wild et al. (2004).

The simplest form for the bias relation is

$$b(\delta_{\rm E})\delta_{\rm E} = b_0 + b_1\delta_{\rm E}, \tag{41}$$

corresponding to linear bias. Of course, this model gives unphysical values for $\delta_{\rm L}$ when $b_1 > 1$ and $-1 \leqslant \delta_{\rm E} < 0$; in this case we set $b(\delta_{\rm E}) = 0$, following the example of Blanton (2000) although it is worth noting that in practice, since we have defined the late-type density field in terms of the density field of early-type galaxies, the best-fitting value for b_1 is always significantly less than 1.

A simple generalization of the bias relation which includes non-linear effects is a power-law bias model,

$$b(\delta_{\rm E})\delta_{\rm E} = b_0(1+\delta_{\rm E})^{b_1} - 1. \tag{42}$$

For both these models there is only one free parameter (b_1), since b_0 is set by the requirement $\langle\delta_{\rm L}\rangle = 0$.

We have fitted deterministic bias models to the counts in cells only for $\ell \geqslant 14\,h^{-1}$ Mpc since below this scale the problem of empty cells becomes significant. The parameters for the best-fitting linear bias model over a range of cell sizes are shown in Fig. 13 (where we have used $1/b_1$ since this corresponds to what is normally understood as the relative bias, namely $b_{\rm rel} = b_{\rm E}/b_{\rm L}$, where $b_{\rm E}$, $b_{\rm L}$ correspond to the linear bias factors for early and late types).

There is a systematic trend for the NGP early types to be more strongly biased than in the SGP. Such a discrepancy, also seen in the results from the modified χ^2 test, was not observed in the variance measurements so it is important to consider whether the effect is indeed as significant as it would appear. Examination of the two-dimensional likelihood contours reveals that in fact the bias parameter and $\sigma_{\rm E}$ are not significantly correlated, so neglecting $\sigma_{\rm E}$ will not lead to an underestimation of the errors in b_1. We have also calculated errors by fitting Monte Carlo realizations of the bias models including our measured errors in $\sigma_{\rm E}$. These are identical in magnitude to the errors obtained from the likelihood function, again suggesting that any correlation between b_1 and $\sigma_{\rm E}$ is not significant.

If we repeat our Monte Carlo error analysis using the Rayleigh–Lévy flight models with a selection function that varies on scales less than ℓ, as in Section 5.2, the actual errors become rather larger.

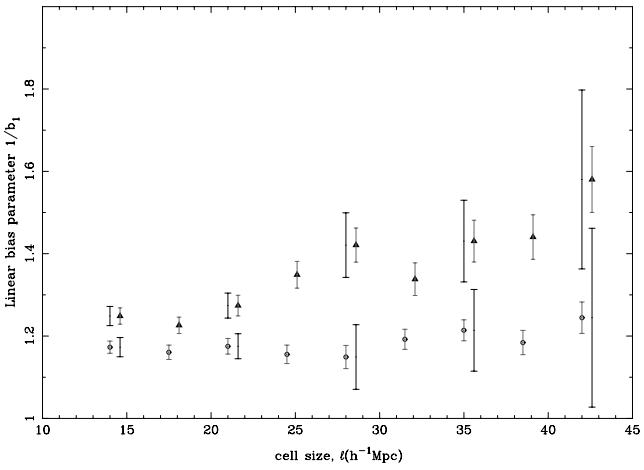

Figure 13. The best-fitting linear bias parameter, $b_{\rm E}/b_{\rm L} = 1/b_1$, as a function of cell size ℓ for the NGP (filled triangles) and SGP (open circles) regions. The grey error bars are derived by considering the value of b_1 for which $\mathcal{L} = \mathcal{L}_{\rm min} + 1$. Black error bars adjacent to selected points are a more realistic error estimate showing the effect on the errors of variation of the selection function on scales less than ℓ.

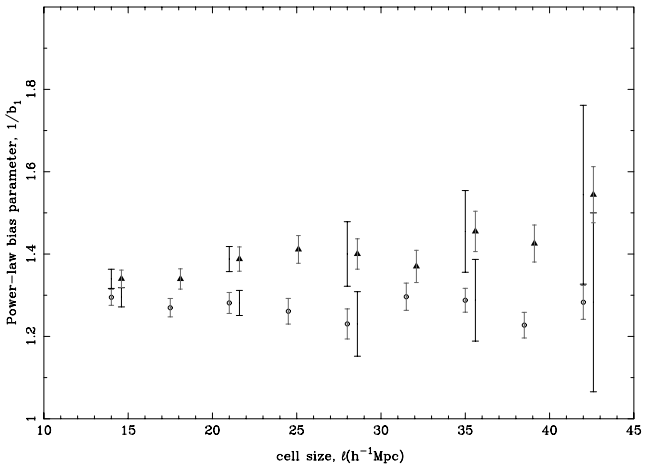

Figure 14. The best-fitting power-law bias parameter, $1/b_1$, as a function of cell size ℓ for the NGP (filled triangles) and SGP (open circles) regions. The grey error bars are derived by considering the value of b_1 for which $\mathcal{L} = \mathcal{L}_{\rm min} + 1$. Black error bars adjacent to selected points are a more realistic error estimate showing the effect on the errors of variation of the selection function on scales less than ℓ.

If we assume that measurements of b_1 in adjacent bins of ℓ are correlated we obtain $b_{1,\rm lin} = 1.27 \pm 0.04$ for the NGP and $b_{1,\rm lin} = 1.17 \pm 0.04$ for the SGP, which corresponds to just less than a 2σ discrepancy.

The variation of the best-fitting power-law bias parameter (again using $1/b_1$ for consistency) with cell size ℓ is shown in Fig. 14. If we again assume that measurements of b_1 in adjacent bins of ℓ are correlated, we obtain $b_{1,\rm PL} = 1.36 \pm 0.05$ for the NGP and $b_{1,\rm PL} = 1.29 \pm 0.04$ for the SGP. These results are noticeably higher than the linear bias parameters, showing that the assumption of linear bias pushes estimates of the bias parameter closer to unity to compensate for non-linearities in the data. Once we account properly for non-linear biasing the bias parameters approach consistency between regions.

Examples of the joint counts in cells for $\ell = 21\,h^{-1}$ Mpc compared to the best-fitting linear and power-law bias models are shown in Fig. 15, and illustrations of the power-law bias fits at a range of scales are shown in Fig. 16. The points show the actual counts in cells. The grey-scale and contour levels show the expected distribution of cell counts, which we have generated using a large number of Monte Carlo realizations of the bias model using the same expected counts as the data. Poisson effects are responsible for the uneven contours on smaller scales; these effects can also be seen in the data.

To test if the models are acceptable fits to the data we have applied a KS test to the 1D distributions of the late-type galaxies for these bias models, i.e. to the projection on the late-type axis of the two-dimensional distributions shown in Figs 15 and 16. The values of $P_{\rm KS}$ for the linear and power-law bias models are shown in the final two columns of Table 4; linear bias is excluded for $\ell < 28\,h^{-1}$ Mpc cells whereas a power-law bias model is not ruled out for any of the scales considered. Fig. 17 shows the relative likelihood $\mathcal{L}_{\rm lin} - \mathcal{L}_{\rm pl}$ of the two models; the power-law bias model is clearly a better fit on smaller scales, although the difference between the goodness of fit of the models decreases with scale, as one would expect from the theoretical prejudice that linear bias should be a good approximation on large scales.

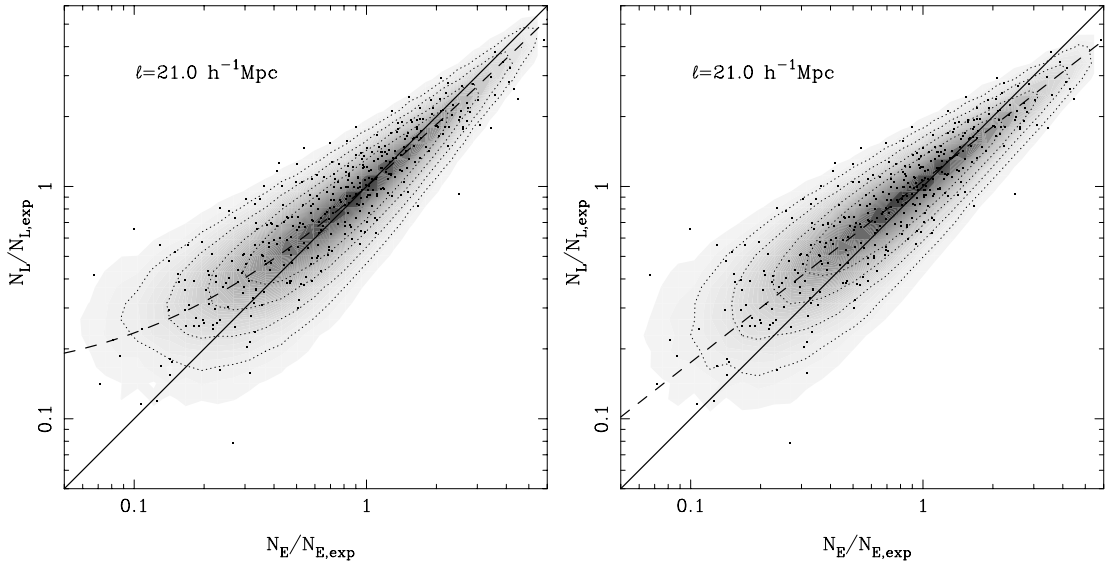

Figure 15. An example of the joint counts in cells for the SGP region with $\ell = 21\,h^{-1}$ Mpc. The grey-scale and contour levels are derived from Monte Carlo realizations of the best-fitting linear bias model (left) and the best-fitting power-law bias model (right), using the same expected counts as the data, and the dotted lines indicate the 50, 70, 85 and 93 per cent significance levels. The solid line indicates a mean relative bias of 1 and the dashed line shows the mean relative bias for the model.

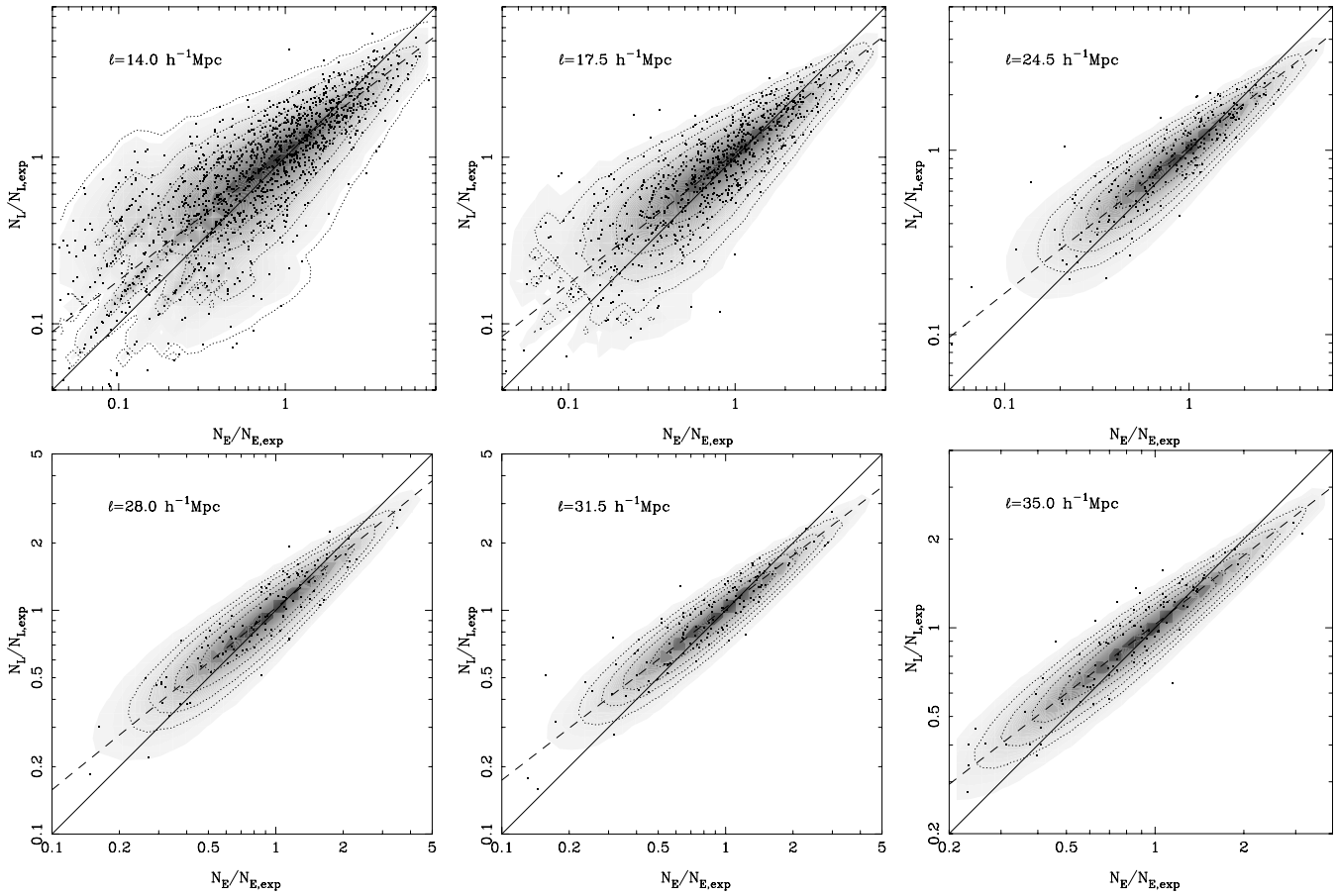

Figure 16. Contour plots of the joint counts in cells compared to the best-fitting power-law bias models for the SGP region over a range of cell sizes, from left–right and top–bottom: $\ell = 14, 17.5, 24.5, 28, 31.5, 35\ h^{-1}$ Mpc. The grey-scale and contour levels are derived from Monte Carlo realizations as previously and the dotted contours are at 50, 70, 85 and 93 per cent significance levels. The solid line indicates a mean relative bias of 1 and the dashed line shows the mean relative bias for the model.

Table 4. Average biasing parameters for the best-fitting power-law bias models assuming $b_{\rm var}$ from the individual lognormal model fits to early and late types. $b_{1,\rm lin}$ and $b_{1,\rm PL}$ are the maximum-likelihood results of Section 5.3.1 using realistic error bars and , and $\sigma_{\rm b}$ are calculated from equations (31) and (33). The KS test probabilities for the linear and power-law bias models are also shown.

ℓ (h^{-1} Mpc)	$b_{1,\rm lin}$	$b_{1,\rm PL}$	$\hat{b}$	$\tilde{b}$	$b_{\rm var}$	$\sigma_{\rm b}$	$P_{\rm KS}$(linear bias)	$P_{\rm KS}$(PL bias)
14	0.83±0.02	0.76±0.02	0.68±0.06	0.7±0.1	0.73±0.02	0.2±0.3	1.3e-5	0.269
21	0.82±0.03	0.75±0.03	0.69±0.09	0.7±0.1	0.73±0.04	0.2±0.5	0.004	0.412
28	0.79±0.08	0.75±0.08	0.7±0.2	0.7±0.3	0.74±0.06	0.1±2	0.688	0.712
35	0.8±0.1	0.7±0.1	0.7±0.2	0.7±0.4	0.75±0.07	0.2±2	0.320	0.760
42	0.7±0.2	0.7±0.2	0.7±0.5	0.7±0.9	0.7±0.1	0.2±3	0.723	0.957

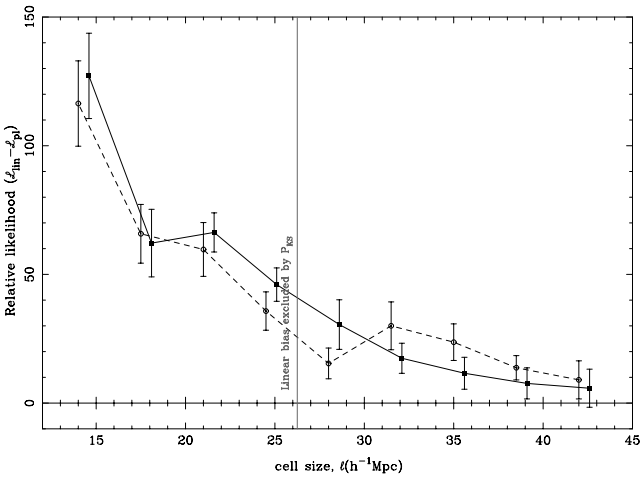

Figure 17. The relative likelihoods of the two deterministic bias models when fitted to the joint counts-in-cells distribution. A value of 0 implies that both models are equally good fits to the data; positive values indicate that the power-law model is a better fit. Results are shown for the NGP (black squares) and SGP (open circles). The dashed line shows the limit in ℓ beyond which a linear bias models is not ruled out by the KS tests. The results, and the error bars shown, are obtained from fits to cell divisions shifted from the original cell division in the same manner as was used for the Tegmark tests in Section 5.2.

5.3.2 Non-linear and stochastic bias

In the case of linear and deterministic bias, all three bias parameters described at the start of the section ($\hat{b}$, $\tilde{b}$ and $b_{\rm var}$), are equal to the parameter b_1 in our model (equation 41). We tabulate the values of $\hat{b}$ and $\tilde{b}$ for our best-fitting power-law bias model below. In the absence of stochasticity we would have $b_{\rm var} = \tilde{b}$, but since we already have an estimated value for $b_{\rm var}$ from the independent fitting of lognormal models to the early- and late-type counts in cells, we can instead ask what value the stochastic bias parameter, $\sigma_{\rm b}$ (equation 33), should take under the rather strong assumption that a power-law bias completely describes any non-linearity, and that the fit is not biased by any stochasticity. The average biasing parameters under this assumption are summarized in Table 4. $\sigma_{\rm b}$ is generally $\sim$0.2, although the large errors mean that we cannot claim to require excess stochasticity above Poisson noise within our errors. A detailed model of stochastic relative bias is discussed by Wild et al. (2004); our results are consistent with the more accurate measurements presented in that paper. Similarly, the non-linearity quantified by $\tilde{b}/\hat{b}$ from our measurements is entirely consistent with that measured by Wild et al. (2004). Szapudi & Pan (2004) describe an interesting technique, based on the cumulative distribution functions, which can in principle recover the full non-linear bias function. This would enable a model-independent measurement of non-linearity and stochasticity, although in unmodified form the technique is not applicable to a flux-limited sample.

6 DISCUSSION

In this section we summarize and discuss in turn some of the main results of the paper: the measurement of the variance and results from fitting lognormal models to the one-point distribution function, a comparison of the linear relative bias parameters obtained by each of our measurement techniques, and finally a discussion of the evidence for scale dependence of the relative bias.

6.1 Variances and the one-point distribution function

We have presented the variance of the counts in cells using the maximum-likelihood technique of Efstathiou et al. (1990), which we have shown is subject to a significant bias when dividing the data into redshift shells of low volume. We have shown that the method can be corrected for this integral constraint bias using the approximation of Hui & Gaztañaga (1999).

The one-point distribution of the counts in cells for early- and late-type galaxies, and the distribution for η-types 1 and 2 combined, has been fitted by lognormal models, using a maximum-likelihood technique. The variances found using this technique are significantly biased on small scales when empty cells are included in the analysis, and we have been able to measure reliable variances only by fitting to counts in cells with empty cells removed. We have corrected our results on small scales to compensate for the inevitable bias resulting from the removal of empty cells. We find that the lognormal model is in general an adequate fit to the distribution functions, as measured by a Kolmogorov–Smirnov test. However, the values for the variance implied by the best-fitting model parameters are slightly high in comparison with both predictions from the correlation functions and relative to the direct counts-in-cells variance measurements presented in this paper. The fact that this bias can be corrected by introducing a weighting scheme giving more weight to regions of higher density contrast suggests that the lognormal model is not an ideal description of the true one-point distribution. It is likely that a generalized lognormal model, such as the 'skewed' lognormal model (SLNDFk) (Colombi 1994; Ueda & Yokoyama 1996), would be a better approximation. Unfortunately, the SLNDFk cannot be used in our maximum-likelihood approach since it is not positive definite, and therefore is not strictly speaking a distribution function.

6.2 Comparison of relative bias parameters

We show in Table 5 a comparison of the relative bias parameters from all of the measurements presented in the paper. We have averaged the bias measurements for all scales with $\ell > 14\,h^{-1}$ Mpc

Table 5. Average bias parameters over all scales from $\ell > 14\,h^{-1}$Mpc for all of the measurements presented in the paper. Error bars are derived assuming measurements for adjacent bins in ℓ scales are correlated.

Bias measurement	NGP	SGP
$1/b_{\rm var}$ (from Efstathiou $\sigma(\ell)$)	1.24±0.06	1.26±0.04
$1/b_{\rm var}$ (from $\sigma_{\rm LN}$ fits)	1.28±0.05	1.27±0.04
$1/b_{1,\rm lin}$ (maximum likelihood)	1.27±0.04	1.17±0.04
$f \approx 1/b_{1,\rm lin}$ (Tegmark test)	1.28±0.03	1.16±0.03
$1/b_{1,\rm PL}$	1.36±0.05	1.29±0.04

which we expect to be unaffected by biases from empty cells. The error bars on each average are obtained assuming that the measurements in adjacent bins in ℓ are perfectly correlated, which is a better approximation than assuming the measurements on separate scales are independent. Where relevant we have also used the more realistic error bars obtained from our Rayleigh–Lévy flight models.

As previously noted, the results for $1/b_{\rm var}$ are consistent between regions and also consistent between measurements from direct variance estimation and fitting lognormal models to the one-point distribution.

Comparing the two estimates of the linear relative bias parameter $1/b_{1,\rm lin}$, from the maximum-likelihood method and from the Tegmark test, we find in both cases a significant discrepancy between NGP and SGP regions. The magnitude of this discrepancy is around 2σ. On the other hand the power-law bias measurements are approximately consistent between regions at a value of $b_{1,\rm PL}$ which is further from unity. As we noted in Section 5.3.1 the assumption of linear bias when fitting to joint counts in cells which contain a significant degree of non-linearity pushes the best-fitting relative bias closer to unity. This effect was also noted by Wild et al. (2004). It is possible that the apparent discrepancy between NGP and SGP linear bias parameters is also partly an artefact produced when non-linear joint distributions are fitted with a linear bias model.

6.3 Scale dependence of the relative bias

In general, the relative bias is expected to be scale dependent on small scales ($r \lesssim r_0$). The scale at which the bias relation becomes scale independent depends on the scales over which the biasing mechanism(s) operates. Non-local bias models (Bower et al. 1993; Matsubara 1999) are those on which the physical processes acting to produce the bias act on scales larger than those defined by the movement of massive particles, for example those models where radiation from QSOs has a significant effect. Local bias models (e.g. Narayanan, Berlind & Weinberg 2000) are those which are defined by some property of the local matter field, for example its density.

Narayanan et al. (2000) determine the variation with scale of a number of local and non-local bias models applied to N-body simulations. A general conclusion of this work is that local bias models are generically unable to influence the biasing relation on scales greater than $r = 8\,h^{-1}$ Mpc, which corresponds to $\ell \approx 12\,h^{-1}$ Mpc in this work. Although there does appear to be some variation of the best-fitting linear bias parameter on scales $\ell \gtrsim 15\,h^{-1}$, when we factor in the larger error bars derived from models including selection function variation across cells in a more realistic way, the significance of any variation becomes negligible. Even if the variation in the linear bias parameter were significant it would not necessarily imply scale dependence of the bias since there is no significant scale dependence of the best-fitting power-law bias parameter. This illustrates the interdependence of non-linear, non-local and stochastic biasing effects. We conclude that any non-local contribution to the relative bias cannot be a dominant effect on large scales.

A special case of local relative bias which was considered in detail by Narayanan et al. (2000) is a local morphology density relation, of the type measured in the local environment of clusters and groups by, e.g. Postman & Geller (1984). A general conclusion for the relative bias produced by such a local effect is that the constant bias factor to which the scale-dependent bias asymptotes on large scales is not equal to unity; assigning galaxy types based on local density produces a difference in clustering strength of the different galaxy types on all scales. Our results are fully consistent with this picture, which leads on to the question of whether the relatively well-studied morphology–density relation can be held solely responsible for the relative bias measured in the 2dFGRS.

7 SUMMARY

We have measured the linear relative bias factor between early- and late-type galaxies in 2dFGRS using a number of methods based on the galaxy counts in cells and over a wide range of scales. In general, all of the methods give consistent values for the relative bias of ≈1.3. We do not find any evidence for significant scale dependence of the bias over the range of scales studied here. We do find, however, that a linear deterministic bias model is a poor approximation for the relative bias, particularly on small scales. We have used a power-law bias model as an example of a deterministic bias model including non-linearity. We find that a power-law bias with an index $b_1 \approx 0.75$ is a consistently better fit to the data than linear bias.

The results presented in this paper are consistent with those presented in the complementary study of Wild et al. (2004) who measure a similar degree of non-linearity and stochasticity ($\tilde{b}$ and $\sigma_{\rm b}$). Their analysis uses smaller, volume-limited subsets of the 2dFGRS, segregated by both spectral type and rest-frame colour. The fact that our results, using the flux-limited catalogue, are consistent with those of Wild et al., using a volume-limited sample, indicates that our results are not significantly biased by systematic effects introduced by the survey selection function. Wild et al. also measure a slight variation in the non-linearity and stochasticity as a function of scale; unfortunately the magnitude of the variation is too small to be measured within our errors. It should be noted that the actual values of the non-linearity and stochasticity parameters measured both in this paper and in Wild et al. are model dependent. Perhaps the most robust result of both papers is the conclusion that both non-linearity and stochasticity are present in the relative bias even on quite large scales.

Wild et al. (2004) also analyse a number of semi-analytic simulations using similar techniques to those used in this paper. The models analysed show similar degrees of non-linearity and stochasticity as exhibited in the data. Complex relative bias is clearly a feature of the galaxy distribution, resulting from the processes of galaxy formation, which has effects over a wide range of scales. More detailed analyses of future simulations should enable the precise causes to be determined and in so doing provide a more thorough understanding of galaxy formation, and of bias in general.

ACKNOWLEDGMENTS

The 2dFGRS was made possible through the dedicated efforts of the staff of the Anglo-Australian Observatory, both in creating the 2dF instrument and in supporting it on the telescope. JAP and OL thank PPARC for their Senior Research Fellowships.

REFERENCES

Baugh C. M., Croton D. J., Gaztañaga E., Norberg P., Colless M., 2004, MNRAS, 351, L44
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Bernardeau F., 1994, A&A, 291, 697
Blanton M., 2000, AJ, 544, 63
Bouchet F. R., Strauss M. A., Davis M., Fisher K. B., Yahil A., Huchra J., 1993, ApJ, 417, 36
Bower R. G., Coles P., Frenk C. S., White S. D. M., 1993, ApJ, 405, 403
Broadhurst T. J., Taylor A. N., Peacock J. A., 1995, ApJ, 438, 49
Carruthers P., 1991, ApJ, 380, 24
Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945
Coles P., Jones B., 1991, MNRAS, 248, 1
Colless M. et al. (the 2dFGRS Team), 2001, MNRAS, 328, 1039
Colless M. et al. (the 2dFGRS Team), 2003, preprint (astro-ph/0306581)
Colombi S., 1994, ApJ, 435, 536
Croton D. J. et al. (the 2dFGRS Team), 2004a, MNRAS, 352, 828
Croton D. J. et al. (the 2dFGRS Team), 2004b, MNRAS, 352, 1232
Dekel A., Bertschinger E., Faber S. M., 1990, ApJ, 364, 349
Dekel A., Lahav O., 1999, ApJ, 520, 24
Dressler A., 1980, ApJ, 236, 351
Efstathiou G., Kaiser N., Saunders W., Lawrence A., Rowan-Robinson M., Ellis R. S., Frenk C. S., 1990, MNRAS, 247, 10P
Efstathiou G. et al. (the 2dFGRS Team), 2002, MNRAS, 330, L29
Elgarøy Ø., Lahav O., 2003, J. Cosmol. Astro-Particle Phys., 4, 4
Fry J. N., 1986, ApJ, 306, 358
Gaztañaga E., Yokoyama J., 1993, ApJ, 403, 450
Hawkins E. et al. (the 2dFGRS Team), 2003, MNRAS, 346, 78
Hui L., Gaztañaga E., 1999, ApJ, 519, 622
Juszkiewicz R., Weinberg D., Amsterdamski P., Chodorowski M., Bouchet F. R., 1995, ApJ, 442, 39
Kaiser N., 1987, MNRAS, 338, 913
Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795
Kayo I. et al., 2004, PASJ, 56, 415
Loveday J., 1996, MNRAS, 278, 1025
Loveday J., Maddox S. J., Efstathiou G., Peterson B. A., 1995, ApJ, 442, 457
Maddox S. J., Efstathiou G., Sutherland W. J., 1990a, MNRAS, 242, 43P
Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990b, MNRAS, 243, 692
Maddox S. J., Efstathiou G., Sutherland W. J., 1990c, MNRAS, 246, 433
Maddox S. J., Efstathiou G., Sutherland W. J., 1996, MNRAS, 283, 1227
Madgwick D. S., 2003, MNRAS, 338, 197
Madgwick D. S. et al. (the 2dFGRS Team), 2002, MNRAS, 333, 133
Madgwick D. S., Lahav O., Somerville R. S., Ellis R. S., 2003a, MNRAS, 343, 871
Madgwick D. S. et al. (the 2dFGRS Team), 2003b, MNRAS, 344, 847
Matsubara T., 1999, ApJ, 525, 543
Narayanan V. K., Berlind A. A., Weinberg D. H., 2000, ApJ, 528, 1
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Norberg P. et al. (the 2dFGRS Team), 2002a, MNRAS, 332, 827
Norberg P. et al. (the 2dFGRS Team), 2002b, MNRAS, 336, 907
Peacock J. A., 1983, MNRAS, 202, 615
Peacock J. A., 1999, Cosmological Physics. Cambridge Univ. Press, Cambridge
Peacock J. A. et al. (the 2dFGRS Team), 2001, Nat, 410, 169
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Percival W. J. et al. (the 2dFGRS Team), 2002, MNRAS, 337, 1068
Percival W. J. et al. (the 2dFGRS Team), 2001, MNRAS, 327, 1297
Postman M., Geller M. J., 1984, ApJ, 281, 95
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C, 2nd edn. Cambridge Univ. Press, Cambridge
Rowan-Robinson M. et al., 1990, MNRAS, 247, 1
Seaborne M. D. et al., 1999, MNRAS, 309, 89
Somerville R. S., Lemson G., Sigad Y., Dekel A., Kauffmann G., White S. D. M., 2001, MNRAS, 320, 289
Szapudi I., Pan J., 2004, ApJ, 602, 26
Tegmark M., 1999, ApJ, 519, 513
Tegmark M., Bromley B. C., 1999, ApJ, 518, L69
Ueda H., Yokoyama J., 1996, MNRAS, 280, 754
Verde L. et al. (the 2dFGRS Team), 2002, MNRAS, 335, 432
Verde L. et al., 2003, ApJS, 148, 195
Wild V. et al. (the 2dFGRS Team), 2004, preprint (astro-ph/0404275)
Zehavi I. et al. (the SDSS team), 2002, ApJ, 571, 172

This paper has been typeset from a TEX/LATEX file prepared by the author.