The Canada-France deep fields survey-II: Lyman-break galaxies and galaxy clustering at z 3

S. Foucaud, H. J. Mccracken, O. Le Fèvre, S. Arnouts, M. Brodwin, S. J. Lilly, D. Crampton, Y. Mellier

A&A 409, 835–850 (2003)
DOI: 10.1051/0004-6361:20031181

Astronomy & Astrophysics

The Canada-France deep fields survey–II: Lyman-break galaxies and galaxy clustering at $z \sim 3^{\star}$

S. Foucaud[1,9], H. J. McCracken[1,8], O. Le Fèvre[1], S. Arnouts[2,1], M. Brodwin[3], S. J. Lilly[4], D. Crampton[5], and Y. Mellier[6,7]

[1] Laboratoire d'Astrophysique de Marseille, Traverse du Siphon, 13376 Marseille Cedex 12, France
[2] ESO – European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
[3] University of Toronto, Department of Astronomy, 60 St. George Street, Toronto, Ontario, Canada M5S 3H8
[4] Institute of Astronomy – ETH Hoenggerberg, HPF D8, 8093 Zurich, Switzerland
[5] Herzberg Institute for Astrophysics, 5071 West Saanich Road, Victoria, British Colombia, Canada V9E 2E7
[6] Institut d'Astrophysique de Paris, 98bis boulevard Arago, 75014 Paris, France
[7] Observatoire de Paris, LERMA, 61 avenue de l'Observatoire, 75014 Paris, France
[8] Present address: University of Bologna, Department of Astronomy, via Ranzani 1, 40127 Bologna, Italy
[9] Present address: Istituto di Astrofisica Spaziale e Fisica cosmica – Sezione di Milano, via Bassini 15, 20133 Milano, Italy

Received 27 February 2003 / Accepted 25 July 2003

Abstract. We present a large sample of $z \sim 3$ $U-$band dropout galaxies extracted from the Canada-France deep fields survey (CFDF). Our catalogue covers an effective area of ∼1700 arcmin^2 divided between three large, contiguous fields separated widely on the sky. To $I_{AB} = 24.5$, the survey contains 1294 Lyman-break candidates, in agreement with previous measurements by other authors, after appropriate incompleteness corrections have been applied to our data. Based on comparisons with spectroscopic observations and simulations, we estimate that our sample of Lyman-break galaxies is contaminated by stars and interlopers (lower-redshift galaxies) at no more than ∼30%. We find that $\omega(\theta)$ is well fitted by a power-law of fixed slope, $\gamma = 1.8$, even at small ($\theta < 10''$) angular separations. In two of our three fields, we are able to fit simultaneously for both the slope and amplitude and find $\gamma = 1.8 \pm 0.2$ and $r_0 = (5.3^{+6.8}_{-2.2})h^{-1}$ Mpc, and $\gamma = 1.8 \pm 0.3$ and $r_0 = (6.3^{+17.9}_{-2.8})h^{-1}$ Mpc (all spatially dependent quantities are quoted for a Λ-flat cosmology). Our data marginally indicates in one field (at a 3σ level) that the Lyman-break correlation length r_0 depends on sample limiting magnitude: brighter Lyman-break galaxies are more clustered than fainter ones. For the *entire* CFDF sample, assuming a fixed slope $\gamma = 1.8$ we find $r_0 = (5.9 \pm 0.5)h^{-1}$ Mpc. Using these clustering measurements and prediction for the dark matter density field computed assuming cluster-normalised linear theory, we derive a linear bias of $b = 3.5 \pm 0.3$. Finally we show that the dependence of the correlation length with the surface density of Lyman-break galaxies is in good agreement with a simple picture where more luminous galaxies are hosted by more massive dark matter halos with a simple one-to-one correspondence.

Key words. cosmology: observations – galaxies: high-redshift – galaxies: evolution – cosmology: large-scale structure of universe

1. Introduction

Surveys of the local Universe, such as the Two-degree Field Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS) (Stoughton et al. 2002), are now providing ever-more detailed pictures of the distribution of galaxies on scales of several hundred Mpc. We have a good paradigm of large-scale structure formation in which small fluctuations of matter density grow under the influence of gravity to form large-scale structures and galaxy halos. Furthermore perturbation theory and numerical simulations provide useful predictions which can be challenged against observations. Making some assumptions on how dark matter traces luminous objects at large scales, we can produce a picture of how galaxies are distributed on large scales locally which match observational data remarkably well.

However, predicting the evolution of clustering to higher redshift is still challenging. We may either attempt to construct

Send offprint requests to: S. Foucaud,
e-mail: foucaud@mi.iasf.cnr.it

⋆ Based on observations obtained at the Canada–France–Hawaii Telescope (CFHT) which is operated by the National Research Council of Canada, the Institut des Sciences de l'Univers (INSU) of the Centre National de la Recherche Scientifique and the University of Hawaii, and at the Cerro Tololo Inter–American Observatory and Mayall 4-meter Telescopes, divisions of the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc. under cooperative agreement with the National Science Foundation.

a fully self-consistent model of galaxy formation which links the dark matter distribution and the luminous galaxies (e.g. Cole et al. 1994; Baugh et al. 1999) or, alternatively, to postulate a relationship between dark matter halos and luminous galaxies (the bias) and use this to predict the galaxy distribution (e.g. Matarrese et al. 1997; Mo et al. 1999). One simple version of this method has been to postulate a linear relationship between the galaxy density, δ_0, and the dark matter one, $\delta_{\rm m}$: $b = \delta_0/\delta_{\rm m}$, where b is the bias parameter (Kaiser 1984). However, until recently, comparing these models to available observations has not been straightforward. For example, angular clustering analyses (e.g. Roche et al. 1993; McCracken et al. 2000) are usually based on magnitude limited samples that typically contain a mixture of galaxy types within a range of redshifts and thus require additional information on the evolution of the galaxy population to allow us to draw meaningful conclusions about the evolution of galaxy clustering.

A much more powerful technique is to measure the clustering of galaxies isolated in different redshift intervals. Spectroscopic surveys, such as the Canada-France Redshift Survey (CFRS) (Lilly et al. 1995; Le Fèvre et al. 1996) and the Canadian Network for Observational Cosmology survey (CNOC) (Carlberg et al. 2000), allow us to directly measure the evolution of the galaxy correlation length r_0 as a function of redshift. Alternatively, the photometric redshift technique has enabled similar analyses up to fainter magnitudes and higher redshifts (e.g. Arnouts et al. 1999; Brunner et al. 2000; Arnouts et al. 2002). Although these various samples are subject to different selection effects and cosmic variance, the results on the clustering measurements agree in showing a general decline of the comoving correlation length r_0 with redshift from $z \sim 0$ to $z \sim 1$. While the clustering amplitude of the underlying dark matter is also expected to decrease with look-back time with a rate depending on the cosmological parameters, the above observations cover a too small redshift range to provide constraints on the evolution of galaxy clustering.

In the early 1990's, several studies attempted to photometrically isolate high redshift ($z \sim 3$) galaxies using very deep $U-$band imaging (Guhathakurta et al. 1990; Steidel & Hamilton 1993). The Lyman limit discontinuity in the emission light of these (star-forming) galaxies, combined with absorption by the intergalactic medium along the line of sight (Madau 1995; Bershady et al. 1999) means these objects are expected to have extremely red ($U - B$) colours, and ($V - I$) colours about zero (Madau et al. 1996). However, it was the advent of 10-m telescopes which allowed the redshifts of these galaxies to be spectroscopically confirmed (Steidel et al. 1996b). Today, a thousand or so of these bright galaxies (i.e. those with $L \sim L^*$) have been spectroscopically confirmed at redshift $z \sim 3$ (Lyman-break galaxies – Steidel et al. 1999), whereas previously only peculiar objects such as QSOs or radio-galaxies were known at this epoch.

The most recent works on $z \sim 3$ galaxies have focused on their physical properties: for example, Adelberger et al. (2003) investigated the cross-correlation between Lyman-break galaxies and the intergalactic medium whereas Shapley et al. (2003) studied their rest-frame UV spectroscopic properties. Properties of these objects at other wavelengths have also been investigated (Nandra et al. 2002; Webb et al. 2003).

More recently, the focus has shifted to replicating the selection of high-redshift objects using the drop-out technique at $z \sim 4$ and beyond (Steidel et al. 1999; Stevens & Lacy 2001; Ouchi et al. 2001; Lehnert & Bremer 2002).

Early studies of clustering measurements of Lyman-break galaxies selected photometrically (Giavalisco et al. 1998) and spectroscopically (Adelberger et al. 1998) indicated they have a correlation length of $\sim 4h^{-1}$ Mpc, comparable to nearby massive galaxies ; a result confirmed by more recent studies (e.g. Adelberger et al. 2003). Since the strength of clustering for dark matter is expected to continuously decrease towards higher redshifts, the high clustering amplitudes found at $z \sim 3$ implies that Lyman-break galaxies are biased tracers of the underlying dark matter distribution, and furthermore suggests that they form preferentially in massive dark matter halos. In the current theoretical paradigm, more massive objects, which form at rarer peaks in the underlying dark matter distribution, have clustering amplitudes much higher than those of less massive, less luminous galaxies (Kaiser 1984; Bardeen et al. 1986). More recent analyses of these Lyman-break galaxies datasets focused on the dependence of clustering amplitude on apparent magnitude selection (Giavalisco & Dickinson 2001) or the behaviour of the galaxy clustering at small angular scales (Porciani & Giavalisco 2002). However, the angular scales probed are generally small, as these surveys consist of many non-contiguous fields each of which covers $\sim$50 arcmin^2. These samples generally contained too few objects to allow a reliable detection of clustering dependence on apparent magnitude or to place useful constraints on the slope of the galaxy correlation function. Furthermore there is also a large spread of measurements made at the same limiting magnitude suggesting the presence of systematic effects or cosmic variance in these surveys.

In this paper, the second in a series, we report new measurements of number counts and clustering properties of Lyman-break galaxies selected in the Canada-France Deep Fields survey (CFDF). The CFDF is a deep, wide-field multi-wavelength survey of four unconnected fields covering three of the CFRS fields. In McCracken et al. (2001), hereafter referred as Paper I, we described the global properties of the CFDF sample and presented measurements of the two-point galaxy correlation function $\omega(\theta)$ as a function of angular scale, limiting I_{AB} magnitude and $(V - I)_{AB}$ colour.

Our wide field optical imaging, combined with deep $U-$band imaging, covering $\sim$0.65 deg^2, allows us to construct the largest sample of photometrically selected Lyman-break galaxies to date. Using spectroscopic observations and simulated catalogues we demonstrate our selection criterion is robust and estimate the degree of contamination in our catalogues. Our three fields, each covering scales of 28′ and separated widely on the sky, allow us to make a robust estimation of the effect of cosmic variance on our results. Additionally the large angular scale of each CFDF field allows us to probe comoving separation at least twice larger than previous works ($\sim 9h^{-1}$ Mpc at redshift $z \sim 3$ for a Λ-flat cosmology with $\Omega_0 = 0.3, \Omega_\Lambda = 0.7$).

This paper is organised as follows: in Sect. 2 we briefly describe the observations which comprise the CFDF survey; in Sect. 3 we outline how Lyman-break galaxies were selected in the CFDF, and present an estimate of the robustness of this selection; in Sect. 4 we present our clustering measurements of the CFDF Lyman-break sample; in Sect. 5 we compare these observations to a range of theoretical predictions, and present our interpretation. Finally, conclusions are summarised in Sect. 6. Unless stated otherwise, throughout this paper we use a Λ-flat cosmology ($\Omega_0 = 0.3, \Omega_\Lambda = 0.7$ to compute spatial quantities and we assume $h = H_0/100$ km s^{-1} Mpc^{-1}).

2. Observations, data reductions and catalogue preparation

2.1. Observations and data reductions

The CFDF survey comprises four separate $28' \times 28'$ fields; and for two and half of these fields we have complete *UBVI* photometry. In total these fields cover $\sim$0.65 deg^2 and include the 03 hr, 14 hr and 22 hr fields of the CFRS survey (Lilly et al. 1995). Lyman-break studies have already been carried out in several subareas of the CFDF-14 hr (the "Groth strip") and the CFDF-22 hr fields by Steidel et al. (1999).

Full details of the CFDF *BVI* observations and the data reduction procedures are given in Paper I. These observations were carried out using the University of Hawaii's 64-megapixel mosaic camera (UH8K) at the Canada-France Hawaii Telescope (CFHT) in a series of runs from 1996 to 1997. In Paper I we demonstrated that the I_{AB} zero-point rms magnitude variation across each UH8K pointing is $\sim$0.04 mag, and that our internal rms astrometric accuracy (between images taken in separate filters) is $\sim$0.05″. This allows us to measure accurately galaxy colours by using the same aperture at the same (x, y) position on stacks constructed from different filters without the needing to positionally match our catalogues.

The unthinned Loral-3 CCDs used in UH8K has very poor response blueward of 4000 Å. For this reason, separate $U-$band observations were carried out at the Cerro Tolo Inter-American Observatory (CTIO) and at the Kitt Peak National Observatory (KPNO) 4-m telescopes. The detectors used were TEK 2048×2048 thinned CCDs with a pixel scale of 0.42″ pixel^{-1}. To cover each $28' \times 28'$ UH8K field, four separate pointings were required. Total integration per pointing was approximatively 10 hours with 10 to 15 exposures. Within each pointing, the airmass varied between 1.0 and 1.6 and seeing ranged from 1.0″ to 1.4″. Reduction of these data followed the usual steps of bias and overscan removal followed by flat-fielding. Each exposure in each pointing was then stacked and scaled so that all have the same photometric zero-point. A coordinate transformation between each of the four sub-pointings and the CFDF $I-$band was then computed. These sub-pointings were then resampled using this transformation to the pixel scale of UH8K (0.205″ pixel^{-1}). Finally, each sub-pointing was coadded to make a single large mosaic covering the entire field of UH8K. The 14 hr and 03 hr fields consist of four separate $U-$band sub-pointings, whereas we have only two for the 22 hr data.

Table 1. Details of the CFDF images used in this study. For the 03 hr and 14 hr fields, we list the 3σ detection limit inside an aperture of 3″ for images convolved to the worst seeing (i.e. 1.3″ for 03 hr and 1.4″ for 14 hr). For the 22 hr field, we list the 3σ detection limit inside an aperture of 4″ for un-convolved images.

Field	RA (J2000)	Dec (J2000)	Band	3σ limit (*AB* mags)
0300+00	03:02:40	+00:10:21	*U*	26.98
			B	26.38
			V	26.40
			I	25.62
1415+52	14:17:54	+52:30:31	*U*	27.71
			B	26.23
			V	25.98
			I	25.16
2215+00	22:17:48	+00:17:13	*U*	27.16
			B	25.76
			V	26.18
			I	25.22

2.2. Catalogue preparation

As described in Paper I, we prepared catalogues using the χ^2 technique outlined in Szalay et al. (1999). This method provides an optimal way for detecting faint objects in multi-colour space. We did not use this method for our 22 hr data as the seeing differs greatly across the $U-$band images; 22 hr $U-$band images are composed of two different pointings taken at CTIO, one has a seeing of 1.2″ and the second 1.4″. Application of the χ^2 technique would involve convolving all images in all bands to the worst seeing, which we would prefer to avoid. Instead, we use an object detection list generated from the $I-$band image and measure colours using apertures at these positions for the other four images. To account for the poorer seeing in these images we use a slightly larger aperture of 4″ to measure galaxy colours; for the other fields we use an aperture of 3″. As we will see later, the slightly different reduction procedures used for the 22 hr field does not affect our clustering measurements. Throughout, galaxy magnitudes are measured using Kron (1980) total magnitudes (`SExtractor` parameter `MAG_AUTO`, Bertin & Arnouts 1996). Table 1 gives the central coordinates of the three fields and the limiting magnitudes in the different bands, taking into account the different aperture sizes and extraction methods used to prepare each catalogue.

Polygonal masks were created covering regions near bright stars, or with lower signal-to-noise, and objects inside these areas were rejected. The total area, after masking, is given in Table 2. As explained in Paper I (Sect. 5.1), we have conducted extensive tests with both correlated and uncorrelated mock datasets to demonstrate that the masking procedure does not affect the estimated correlation amplitudes.

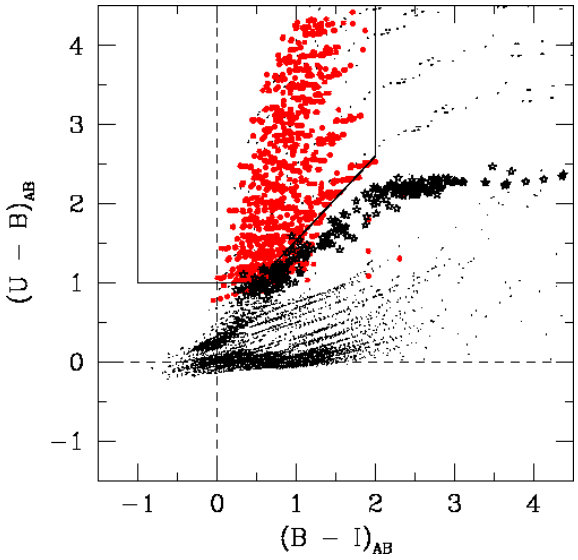

Fig. 1. Galaxy evolutionary tracks (dots) used to define our selection box, represented as the solid line. Filled symbols indicate galaxies in the range $2.9 < z < 3.5$. Star symbols represent simulated colours for galactic stars with $I_{AB} < 20.0$.

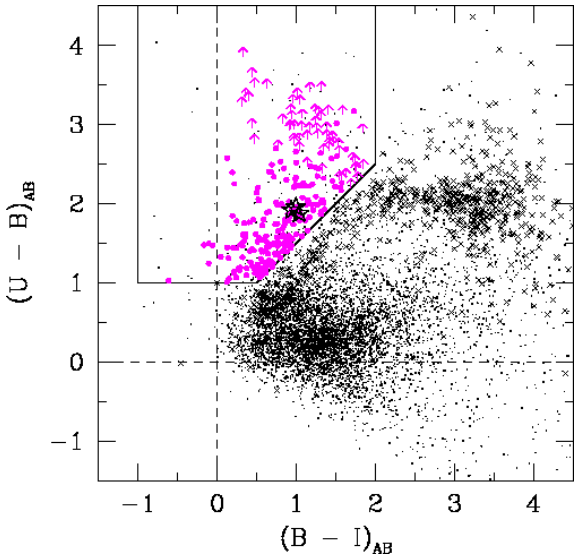

Fig. 2. $(U-B)_{AB}$ against $(B-I)_{AB}$ for galaxies with $I_{AB} < 24$ in the CFDF-03 hr field. Almost 14 000 objects are represented; for clarity only half of all objects in this field are shown. The solid line represents the selection box given in Eq. (1). There are 269 candidates (filled circles) which satisfy our selection criteria. The arrows indicate Lyman-break candidates which have a 1σ upper limit in U. Crosses indicate star-like objects (identified on the basis of their compactness). The three spectroscopically confirmed Lyman-break galaxies are shown with star symbols.

3. The sample of Lyman-break galaxies

3.1. Selecting Lyman-break galaxy candidates

Lyman-break candidates were selected by isolating the Lyman-break feature at 912 Å in a colour-colour diagram (Steidel & Hamilton 1993). To define our selection box we examine the path of synthetic evolutionary tracks in the $(U-B)$ vs. $(B-I)$ colour-colour space defined by the CFDF filter set. These tracks are derived from a set of spectral energy distribution templates (Bruzual & Charlot 1993) assuming a Λ-flat cosmology.

Figure 1 illustrates the tracks used; each track represents a different combination of galaxy type, age, metallicities and reddening. Internal extinction is modelled using a relation appropriate for starburst galaxies (Calzetti et al. 1994). We have also included the Lyman absorption produced by the intergalactic medium following Madau (1995). Colours of field stars are estimated using the galactic model of Robin & Creze (1986) transformed to our instrumental system (magnitude errors are not include in this figure).

Based on these considerations, we define our selection box as

$$\begin{aligned} &1.0 \leq (U-B)_{AB}, \\ &-1.0 \leq (B-I)_{AB} \leq 2.0, \\ &(B-I)_{AB} + 0.5 \leq (U-B)_{AB}, \\ &(V-I)_{AB} \leq 1.0. \end{aligned} \quad (1)$$

We estimate the redshift range of our Lyman-break sample to be $2.9 < z < 3.5$, quite close to the $2.7 < z < 3.4$ interval sampled by Steidel et al. (1996a).

The criterion $(V-I)_{AB} \leq 1.0$ reduces contamination by stars and avoids contamination by elliptical galaxies $z \sim 1.5$. Additionally, we require that our Lyman-break candidates are detected in B, V and I. Finally, *all* candidates are visually inspected in all five channels ($UBVI$ and the χ^2 detection image) before they are added to the source catalogue. About ten percent of the Lyman-break sources were rejected as spurious; these objects are typically detections on bad columns or other cosmetic defects which had not been removed by the masking process. Given the detection limits in U and I presented in Table 1, selecting Lyman-break galaxies to $I_{AB} = 24.5$ is feasible for all our fields.

In Fig. 2 we show the $(U-B)$ vs. $(B-I)$ colour-colour diagram for galaxies with $I_{AB} < 24$ in the 03 hr field, with Lyman-break candidates identified using the selection box defined in Eq. (1). Redshifts for three of these galaxies were spectroscopically confirmed ($z = 3.07, 3.08$ and 3.27 respectively) with data taken at CFHT in November 1997 using the Multi-Object Spectrograph. These galaxies are plotted in Fig. 2 as open stars. A spectrum of one of these galaxies is shown in Fig. 3.

As we have limited spectroscopy on CFDF $z \sim 3$ galaxies, we have carried out extensive simulations, described in the Sect. 3.3, to ensure the robustness of our selection box. As we will see, these simulations allow us to quantify how much contamination we expect from stars and lower-redshift interlopers.

3.2. Source counts of Lyman-break galaxies

In Fig. 4 we present our raw and corrected Lyman-break galaxy counts as a function of I_{AB} magnitude (dotted and solid symbols). Table 2 presents the surface densities of our Lyman-break galaxy sample for a ranges of limiting magnitudes. In addition, we present counts from Metcalfe et al. (2001) and Steidel et al. (1999) samples of Lyman-break galaxies. The latter

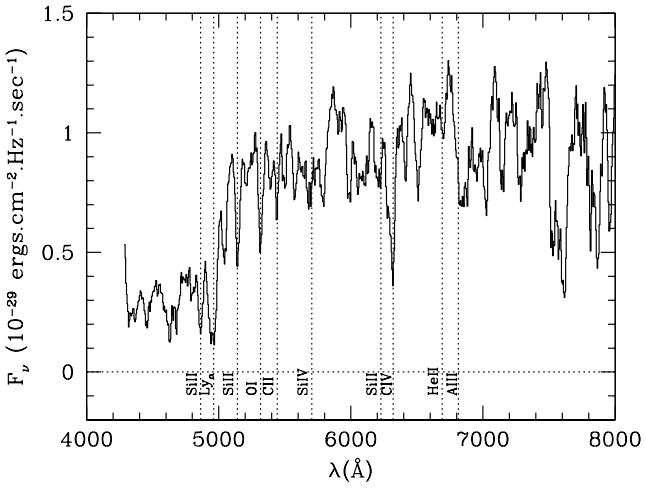

Fig. 3. Spectrum of a confirmed Lyman-break galaxy at redshift $z = 3.08$ observed at CFHT using the MOS spectrograph. Spectral features are indicated with the dotted lines.

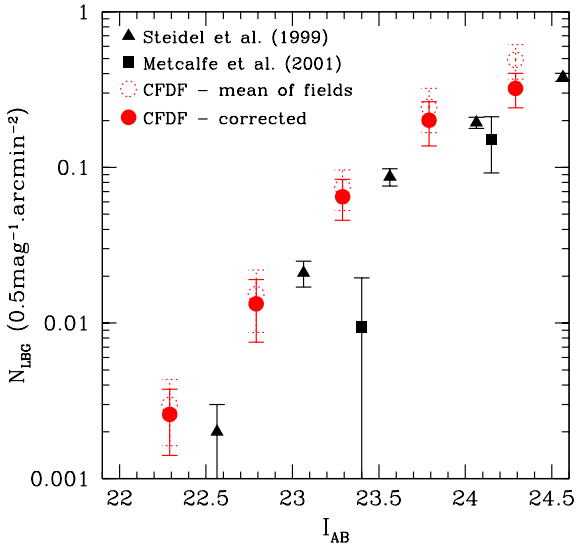

Fig. 4. Raw and corrected number counts of Lyman-break galaxies in the CFDF (open and filled circles respectively). The errorbars on each point is computed from the amplitude of the field-to-field variance. We also show colour-selected Lyman-break galaxy counts from Metcalfe et al. (2001) (filled squares) and Steidel et al. (1999) (filled triangles).

compilation contains a correction for contamination by stars and AGN estimated from spectroscopic observations.

To convert these $\mathcal{R}_{AB}$-selected observations to our I_{AB} magnitudes, we estimate the mean colour of Lyman-break galaxies at redshift $z \simeq 3$ to be $(R-I)_{AB} \simeq 0.3$. In making this transformation we assume that the colours of the Lyman-break population do not evolve with magnitude. Combining this with the conversion between $\mathcal{R}_{AB}$ and R_{AB} given in Steidel & Hamilton (1993), we estimate that $(\mathcal{R}-I)_{AB} \simeq 0.19$.

At bright magnitudes, our counts are in good agreement with the literature compilation: however, at fainter bins $24.0 < I_{AB} < 24.5$ they exceed the literature comparisons by a factor 1.3−1.5. Essentially this is due to higher contamination in our sample. According to simulations (which we describe fully in Sect. 3.3) this contamination, arising from the shallower depth of our *UBVI* data compared to Steidel et al.'s $U_nG\mathcal{R}$ data, amounts to ∼30% in the faintest bins. Counts corrected for this contamination are indicated as the dotted symbols in Fig. 4. After this correction our counts are in closer agreement with the literature.

We note that the dispersion in Lyman-break counts between our three fields is larger than one would expect based on purely Poissonian errors. We suggest several possible explanations for this dispersion. Firstly, Lyman-break galaxies are strongly clustered objects: at the magnitude limit of the survey, this clustering can produce count fluctuations of ∼15%. Secondly, the absolute photometric calibration between each of the three fields (which were all taken in different observing runs, in different seasons, and in some cases with different U−band imagers) differs by at worst ∼0.1 mag (although, as demonstrated in Fig. 9 of Paper I, the field–to–field variation in galaxy counts is still very small).

How large an effect could a systematic error of ∼0.1 mag have on the Lyman-break number counts? To address this question we have carried out a set of simulations in which a small, Gaussian error of $\sigma = 0.1$ is added to each filter, i.e., new magnitudes are computed according to $M' = M + \delta M$. The number counts of galaxies falling in the selection box is recomputed at each iteration. From this experiment we find that magnitude errors of $\sigma = 0.1$ can produce a fluctuation in Lyman-break counts of ∼15%. Adding the contribution from the clustered nature of Lyman-break galaxies, this leads to a total expected field-to-field fluctuation of ∼20%, large enough to explain the deviation between the 14 hr and 22 hr fields. We have examined the 03 hr field in more detail, and we find that one quadrant has a ∼50% higher density of Lyman-break candidates than the other three: if this quadrant is removed, the fluctuations between the 03 hr field and the other two can be explained by the sources of errors listed above. The effect of this overdense quadrant on $\omega(\theta)$ is to increase the amount of power at ∼0.1° scales but, as we will see in Sect. 4, at the scales we normally measure galaxy correlation amplitudes, the field-to-field variation in $\omega(\theta)$ is still less than the amplitude of the Poissonian error in $\omega(\theta)$.

3.3. Estimating the reliability of the CFDF Lyman-break selection box

To estimate the level of contamination by stars and interlopers (lower-redshift galaxies) and the fraction of Lyman-break galaxies which could be missed in our sample, we construct multi-colour mock catalogues which incorporate all the observational uncertainties.

We use the model[1] of Robin & Creze (1986) to generate our stellar catalogue at the galactic latitude of the 14 hr field. The catalogue's *UBVI* Johnson-Cousins colours were transformed to our instrumental system and then convolved with a function describing the dependence of magnitude errors with magnitude for each passband. In Fig. 1 star symbols show objects from this catalogue; for clarity, only stars with $I_{AB} < 20.0$ and without

[1] http://www.obs-besancon.fr/www/modele/

Table 2. Differential number counts, N, and surface density, n (in arcmin^{-2}), of Lyman-break galaxies in the CFDF fields, for a range of I_{AB}-selected slices. The mean surface density, labelled CFDF, is also given. Errors in the surface density measurements for each individual field are computed using Poisson counting statistics; field–to–field variance is used to estimate the error in the mean.

	0300+00		**1415+52**		**2215+00**		**CFDF**
	$A = 646$ **arcmin**^{2}		$A = 708$ **arcmin**^{2}		$A = 317$ **arcmin**^{2}		
magnitude range (I_{AB})	n	(N)	n	(N)	n	(N)	n
20.0–22.5	0.003 ± 0.003	(2)	0.010 ± 0.004	(7)	0.006 ± 0.006	(2)	0.006 ± 0.004
22.5–23.0	0.020 ± 0.006	(12)	0.008 ± 0.004	(6)	0.013 ± 0.008	(4)	0.014 ± 0.006
23.0–23.5	0.08 ± 0.01	(55)	0.05 ± 0.01	(36)	0.07 ± 0.02	(22)	0.07 ± 0.02
23.5–24.0	0.31 ± 0.02	(200)	0.16 ± 0.02	(116)	0.22 ± 0.03	(69)	0.23 ± 0.08
24.0–24.5	0.59 ± 0.03	(379)	0.34 ± 0.02	(242)	0.45 ± 0.04	(142)	0.46 ± 0.13

magnitude errors are shown. Fainter objects occupy the same region in colour-colour space.

For the galaxy catalogues, we use the empirical approach developed by Arnouts et al. (2003, in preparation); the main components of which are as follows. To characterise the spectral energy distribution (SEDs) of galaxies, we use the four observed SEDs of Coleman et al. (1980) (corresponding to Elliptical, Sbc, Scd and Irregular local galaxy types), and two SEDs corresponding to star-forming galaxies with ages of 0.05 and 2 Gyrs. These SEDs were computed using the GISSEL model (Bruzual & Charlot 1993) assuming solar metallicity, a Salpeter initial mass function and constant star formation rate. Following the approach adopted by Sawicki et al. (1997), we interpolated between the 6 original spectra to provide a finer grid of the spectral-type coverage producing a total number of 61 templates. We derive the density of objects for given magnitude and redshift interval using the luminosity function parameters from the $R-$band ESO-Sculptor Survey to $z \simeq 0.6$. Galaxies are divided into three spectral classes: early, intermediate and late types (de Lapparent et al. 2002, in preparation). At higher redshift the luminosity function parameters have been adjusted in order to reproduce the observed redshift distributions of the CFRS (Crampton et al. 1995) and the North and South Hubble Deep Fields (HDF-N and -S) (Arnouts et al. 1999, 2002). We derive magnitudes in other passbands using these SEDs to compute the appropriate $k-$correction. A model for the "observed" magnitudes is obtained by taking into account the luminosity profile of the galaxy and observational conditions (such as seeing and surface brightness limits) and computing the fraction of light lost according to the magnitude scheme employed. We derive photometric errors using the observed dependence of error with magnitude in each passband. This empirical method reproduces the main observables such as counts, colours and redshift distributions. Special attention is paid to the redshift distributions to ensure a reasonable description of the relative fraction of galaxies at low and high redshift which is the first step in quantifying how target selection in a colour-colour diagram can be subject to contamination effects.

In Table 3 we show the surface densities of all the simulated objects (computed for an area of ~1 deg^2) and compare them to observations in the CFDF-14 hr field. The total surface densities of objects found in the selection box from simulations and observations are close, reflecting our requirement that the models match observed redshift distributions. According to the simulations, the contamination by stars decreases from 6% to 3%, while the contamination by galaxies outside our chosen redshift range increases from 8% to 25% for $I_{AB} < 23.5$ to $I_{AB} < 24.5$ respectively. Furthermore we find that the class of interlopers changes as a function of limiting magnitude. For $I_{AB} < 23.5$, about 70% of the galaxy interlopers are expected to be at $z \geq 2$ and the remaining at $z < 2$. At $I_{AB} < 24.5$ the situation is different, due to the larger uncertainties in the colours: about 60% of interlopers are expected to be $z < 2$ and a large part of the remainder (~25%) are at $z \geq 2$. In the following section we assess the reliability of these simulations by direct comparison with spectroscopic observations.

Our 14 hr field covers the "Groth strip" field. C. Steidel has kindly provided us with spectroscopic redshifts for 335 photometrically selected objects in this area and we have used this dataset to assess the reliability of our selection box. There are 315 objects in common (based on a simple positional match) between the two catalogues, and for these objects, selected using $U_nG\mathcal{R}$ photometry, we have spectroscopic redshifts in addition to CFDF $UBVI$ photometry. Table 4 shows the redshift distribution for galaxies with $I_{AB} < 24.5$ before and after the application of our selection box.

In total we retrieve 59.6% (31/52) of galaxies at $2.9 < z < 3.5$ after applying our selection box. Given that the redshift distribution of the two samples is different (with mean redshifts of $\overline{z} = 3.04$ and $\overline{z} = 3.2$ respectively) this is to be expected, assuming the underlying distributions are Gaussian with the same dispersion.

Although the photometric selection of the Steidel et al. sample is different from ours, we can attempt to estimate the amount of contamination in our catalogue by galaxies outside our redshift range after the application of our selection box. Inside our selection box, galaxies at lower redshifts ($2.0 < z < 2.9$) amount to 25% of the total. Spectroscopically identified stars account for a further 3.8% of objects, in broad agreement with the results of our simulated catalogues. However, we note that the spectroscopic sample contains no objects with $z < 2$, in disagreement with our simulations. Finally, 9.6% of our candidates have no redshift.

Table 3. Observed and simulated surface densities of objects (in arcmin^{-2}) recovered using the selection box (Eq. (1)), based on simulations described in Sect. 3.3 and observations in the CFDF-14 hr field. We also estimate the fraction of Lyman-break galaxies (LBG) recovered using this selection box from the total galaxy population in this redshift range. Additionally, we present the fraction of contaminants within this selection box by interlopers (lower-redshift galaxies) outside our redshift range ($2.9 < z < 3.4$) and by interlopers with $z < 2$ and by stars.

magnitude range	density of objects		LBG	contamination from		
(I_{AB})	in the selection box from		found in the	interlopers outside	interlopers	stars
	observations	simulations	selection box	our $2.9 < z < 3.5$ range	with $z < 2$	
20.0–23.5	0.069	0.073	91.8%	7.7%	2.2%	5.5%
20.0–24.0	0.233	0.233	89.7%	11.9%	6.3%	4.3%
20.0–24.5	0.574	0.652	83.9%	24.5%	17.3%	3.5%
23.5–24.0	0.164	0.160	88.7%	13.9%	8.2%	3.8%
24.0–24.5	0.341	0.419	80.2%	31.4%	23.5%	3.1%

Table 4. Comparison for different redshift ranges between objects falling within Steidel et al.'s selection box and objects selected in Steidel et al.'s box which also lie within the CFDF selection box (Eq. (1)) for $I_{AB} < 24.5$.

redshift range	Steidel et al. box	combined CFDF/Steidel box
total	108	52
$2.9 < z < 3.5$	52	31
$2.0 < z < 2.9$	26	13
$z \leq 2.0$	0	0
$z \geq 3.5$	1	0
stars	6	2
QSOs	3	1
no z	20	5

Furthermore, the full spectroscopic catalogue, there are no objects with $z < 2$; however, in ∼18.5% candidates have no measured redshift. Determining redshifts for galaxies in the range $1 < z < 2$ is difficult, so it is possible some of these unidentified objects *could* be galaxies in this redshift range. But as the main fraction of these object simply have not been attempted yet, this couldn't account for some of the ∼17% of contamination by interlopers with redshift $z < 2$ indicated by our simulations at $I_{AB} < 24.5$.

Although our U data is approximately as deep as Steidel et al.'s, our BI data is somewhat shallower than their $G\mathcal{R}$ images. For example, detection limits of the Steidel et al. data are approximately $(U_nGR)_{AB} \sim 27.3, 27.3, 26.8$ (Adelberger et al. 2003) compared CFDF limits of $(UBVI)_{AB} = 27.0, 26.4, 26.4, 25.6$ (3σ limits, $3''$ diameter aperture, 03 hr field; see Paper I for more details). At fainter magnitudes the shallowness of our B images is expected to increase our contamination by lower-redshift galaxies. This can explain the discrepancy between our raw number counts and the number counts of Steidel et al. as shown in Fig. 4, and the fact that they are in good agreement after correction from contamination in our sample.

In summary, we estimate that our sample is contaminated at the level of 15% to 30% between $I_{AB} < 23.5$ and $I_{AB} < 24.5$ respectively. Our selection box allows us to recover a large fraction of simulated Lyman-break galaxies, ranging from 95% to 80% between $I_{AB} < 23.5$ and $I_{AB} < 24.5$ respectively. Comparisons with a large sample of galaxies with spectroscopic redshifts (preselected, however, using a different photometric criterion from ours) indicate we recover, in this case, ∼60% of the Lyman-break galaxies. We attribute this discrepancy to the different underlying redshift distributions for the two photometrically selected samples.

4. Clustering of the Lyman-break galaxies

4.1. Estimating $\omega(\theta)$ and A_ω

To measure $\omega(\theta)$, the two-point projected galaxy correlation function, we use the Landy & Szalay (1993) estimator,

$$\omega(\theta) = \frac{\mathrm{DD} - 2\mathrm{DR} + \mathrm{RR}}{\mathrm{RR}}, \quad (2)$$

where the DD, DR and RR terms refer to the number of data-data, data-random and random-random galaxy pairs having angular separations between θ and $\theta + \delta\theta$.

In the weak clustering regime this estimator has a nearly Poissonian variance (Landy & Szalay 1993),

$$\delta\omega(\theta) = \sqrt{\frac{1 + \omega(\theta)}{\mathrm{DD}}}. \quad (3)$$

Section 4.6 addresses the reliability of this error measurement for our present dataset. To determine A_ω, the amplitude of $\omega(\theta)$ at 1 degree, we assume that $\omega(\theta)$ is well represented by a power-law of slope δ, i.e. $\omega(\theta) = A_\omega\theta^{-\delta}$ (Groth & Peebles 1977). In what follows, we assume $\delta = 0.8$; in Sect. 4.2 we explore this assumption in more detail. This fitted amplitude must be adjusted to take into account the "integral constraint" correction, which arises from the fact that the mean background density of galaxies is estimated from the sample itself. We estimate this term as follows (Roche et al. 1993),

$$C = \frac{1}{\Omega^2}\int\int \omega(\theta)\mathrm{d}\Omega_1\mathrm{d}\Omega_2, \quad (4)$$

where Ω is the area subtended by the survey field. To determine C we numerically integrate this expression over each

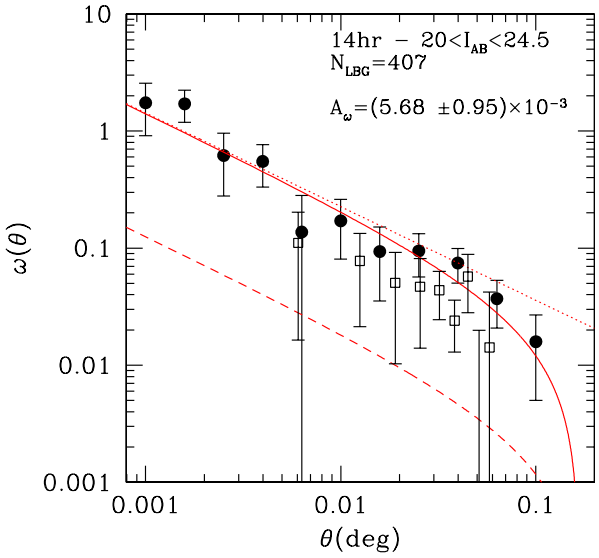

Fig. 5. The amplitude of the angular correlation function, $\omega(\theta)$, as a function of the angular separation in degrees, for a $20 < I_{AB} < 24.5$ limited Lyman-break sample of the CFDF-14 hr field (filled black circles). Errorbars represent normal Poisson errors (Eq. (3)). The solid line shows the fitted correlation amplitude, derived using Eq. (5) and assuming a power-law slope of $\delta = 0.8$ and a value of $C = 4.2A_\omega$ for the integral constraint term. The dotted line shows the fitted power-law without the integral constraint correction. The long dashed line shows the fitted correlation amplitude (from Paper I) for field galaxies selected with the same limiting magnitude. Open squares are the $\omega(\theta)$ measurements from the Giavalisco et al. (1998) $\mathcal{R} < 25.5$ selected Lyman-break sample.

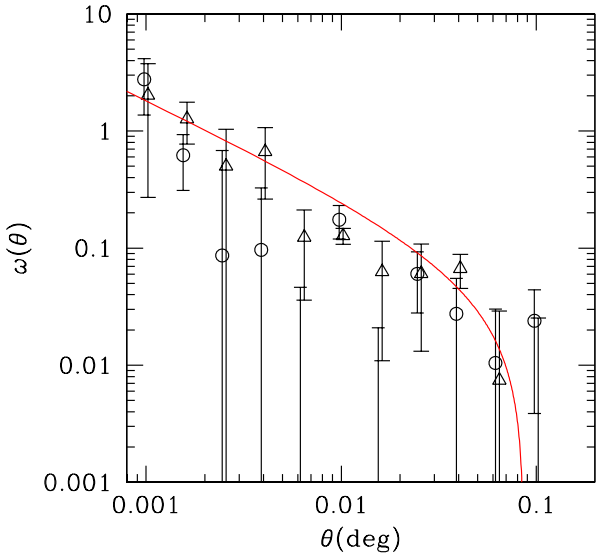

Fig. 6. The quadrant-averaged Lyman-break correlation function $\omega(\theta)$ for galaxies in the 03 hr and 22 hr fields (triangles and circles respectively). The amplitude of these error bars corresponds to the quadrant-to-quadrant variation in $\omega(\theta)$. The solid line shows the fitted correlation amplitude for Lyman break galaxies in the 14 hr field with a fixed slope applied for one quadrant (Fig. 5).

field, excluding masked regions. We find $C \simeq 4.2A_\omega$ for the 14 hr and 03 hr fields. For the 22 hr field, which has half the coverage, we derive $C \simeq 5.5A_\omega$. Then we determine A_ω fitting the expression:

$$\omega_{\rm obs}(\theta) = A_\omega \theta^{-\delta} - C. \quad (5)$$

Figure 5 shows the angular correlation function, $\omega(\theta)$, as a function of the angular separation in degrees for Lyman-break galaxies with $20.0 < I_{AB} < 24.5$ in the CFDF-14 hr field. Here the errorbars have been estimated with the normal Poisson errors (Eq. (3)). The fitted amplitude derived from Eq. (5) is represented by the solid line. The long dashed line shows the A_ω value computed from the CFDF field galaxy sample at the same limiting magnitude (from Paper I). The A_ω for the Lyman-break sample is ∼10 higher than the field galaxy sample. In Table 5 we summarise our Lyman-break A_ω measurements for a range of limiting magnitudes in the CFDF.

In Fig. 5 we compare our measurements of $\omega(\theta)$ to those of Giavalisco et al. (1998). As the largest CFDF fields are ∼9 times larger than those used in this study, our measurements of $\omega(\theta)$ cover a much larger range of angular separations. We note that our amplitude measurements are ∼2 times larger those of Giavalisco et al.; we expect this arises from the greater depth of the Giavalisco et al. study compared to the CFDF. To test that the origin of this discrepancy in amplitude did *not* arise from inhomogeneities within our fields, we extracted, from each CFDF field, sub-fields covering the same $9' \times 9'$ area as subtended by the Giavalisco et al. work. In total we extracted 21 fields of these dimensions. We fitted each sub-field individually and found a median correlation amplitude over all fields of $(6.9 \pm 5.1) \times 10^{-3}$, which is in good agreement with the full field value of $(7.4 \pm 1.0) \times 10^{-3}$ quoted in Table 5. The results of this test are consistent with the simulations carried out in Paper I in which we demonstrated that measurements of $\omega(\theta)$ for $I_{AB}-$ limited samples in the CFDF are unaffected by sensitivity variations across the mosaics to at least $I_{AB} \sim 25$. Finally, we also note that our measurements of $\omega(\theta)$ follow a power-law behaviour over the entire range $0.001^\circ < \theta < 0.02^\circ$ accessible to our survey and there is no evidence of an excess of power on large scales (with the exception of the 03 hr field), as one might expect if residual inhomogeneities existed within individual field.

We also measured $\omega(\theta)$ in four separate sub-areas on each of our three fields. Each sub-areas corresponds to the size of the individual U-band pointings. In each sub-area we measure $\omega(\theta)$ separately and then determine the mean and the variance: this is illustrated in Fig. 6. Measuring $\omega(\theta)$ is these sub-areas is more challenging as the numbers of galaxies involved is much smaller. However, the fitted amplitudes in each of these sub-areas agrees very well with the full-field values presented in Table 5.

4.2. Measuring the slope

Is the slope of the $\omega(\theta)$ for Lyman-break galaxies really $\delta = 0.8$? In earlier works (Adelberger et al. 1998; Arnouts et al. 1999), a value of $\delta = 0.8$ was used based on results from local large surveys (Groth & Peebles 1977). In contrast, Giavalisco et al. (1998) measured $\delta = 1.0 \pm 0.3$. The large angular coverage of the CFDF fields allow estimate δ; in Fig. 5, we can easily

Table 5. The amplitude of $\omega(\theta)$ at 1 degree, A_ω, the slope δ and the comoving correlation length r_0 (in h^{-1} Mpc), for each field and for different magnitude limited samples considered in this paper. r_0 measurements are computed for three standard cosmological models. To derive r_0 we assume a top-hat redshift distribution centred at $\overline{z} = 3.2$ and the best fitted value of the slope. Result marked as CFDF mean are computed from the mean over all three fields. The error bars shown correspond to Poisson error bars. An extra systematic errorbar arising from our uncertainty in the underlying redshift range of our Lyman-break sources of ±0.2 for the entire faint samples and of ±0.4 for the bright sample should be added. (Our principal results are highlighted in bold.)

Field	magnitude limit (I_{AB})	A_ω(1 deg) $\times 10^{-3}$	δ	r_0 (h^{-1} Mpc) $\Omega_0 = 1.0$, $\Omega_\Lambda = 0.0$	r_0 (h^{-1} Mpc) $\Omega_0 = 0.2$, $\Omega_\Lambda = 0.0$	r_0 (h^{-1} Mpc) $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$
CFDF-14	20.0–24.5	$5.9^{+13.2}_{-4.0}$	$\mathbf{0.81}^{+0.21}_{-0.24}$	$3.2^{+4.0}_{-1.2}$	$3.6^{+4.4}_{-1.3}$	$5.3^{+6.6}_{-2.0}$
	20.0–23.5	$7.8^{+166.0}_{-7.7}$	$1.08^{+0.84}_{-0.66}$	$5.8^{+59.8}_{-2.8}$	$6.8^{+69.3}_{-3.2}$	$9.5^{+97.7}_{-4.5}$
	23.5–24.5	$2.3^{+6.4}_{-1.7}$	$0.96^{+0.25}_{-0.26}$	$2.6^{+3.2}_{-1.0}$	$3.0^{+4.2}_{-1.1}$	$4.3^{+6.1}_{-1.6}$
CFDF-22	20.0–24.5	$8.1^{+40.9}_{-6.0}$	$\mathbf{0.81}^{+0.25}_{-0.35}$	$3.8^{+10.7}_{-1.6}$	$4.2^{+11.8}_{-1.7}$	$6.3^{+17.7}_{-2.6}$
CFDF-03	20.0–24.5	8.6 ± 0.6	0.8	3.9 ± 0.2	4.3 ± 0.2	6.4 ± 0.3
CFDF mean	20.0–24.5	7.4 ± 1.0	0.8	3.6 ± 0.3	3.9 ± 0.3	$\mathbf{5.9 \pm 0.5}$
CFDF-14	20.0–23.5	24.9 ± 7.9	0.8	7.0 ± 1.2	7.7 ± 1.4	$\mathbf{11.6 \pm 2.0}$
	23.5–24.5	5.4 ± 1.1	0.8	3.0 ± 0.3	3.3 ± 0.4	$\mathbf{5.0 \pm 0.6}$

detect power in $\omega(\theta)$ to scales of $\sim 0.1°$, making it possible to place constraints on the joint values of A_ω and δ.

To estimate the best-fitting values of A_ω and δ we carry out a χ^2 minimisation on the values of $\omega(\theta)$ determined for all fields, similar to that described in Paper I. This computation accounts for the dependence of the integral constraint C with the slope δ. Figure 7 shows the fit for two of our three fields; our data provides an approximate constraint on δ. We find the mean of the best fitted slopes is $\delta = 0.81^{+0.21}_{-0.24}$ for the CFDF-14 hr field and $\delta = 0.81^{+0.25}_{-0.35}$ for the CFDF-22 hr field for $I_{AB} < 24.5$ (we were not able to fit simultaneously both for the slope and amplitude on the CFDF-03 hr field).

To summarise, our clustering measurements are broadly consistent with a power-law of slope $\delta \sim 0.8$ over all the magnitude ranges accessible to our survey. Our data do not provide any strong evidence for slopes shallower or steeper than this value, as suggested by other authors (Giavalisco & Dickinson 2001; Adelberger et al. 2003).

4.3. The comoving correlation length r_0

We use the spatial correlation function (Groth & Peebles 1977), to derive r_0, the comoving galaxy correlation length based on our angular clustering measurements, given by

$$\xi(r, z) = \left(\frac{r}{r_0(z)}\right)^{-\gamma}, \quad (6)$$

where $\gamma = 1 + \delta$. The redshift dependence is included in the comoving correlation length $r_0(z)$.

Using the relativistic Limber equation (Peebles 1980; Magliocchetti & Maddox 1999), we can derive the correlation length r_0 from the correlation amplitude A_ω, providing we can estimate a redshift distribution for our sources. In what follows we assume that our Lyman-break redshift distribution is well described by a top-hat function spanning the interval $2.9 < z < 3.5$; however, our results are unchanged if we use a Gaussian redshift distribution covering the same interval.

Could our adopted redshift distribution be modified by the presence of interlopers? Assuming a Gaussian distribution of Lyman-break galaxies centred on $2.9 < z < 3.5$ with $\overline{z} = 3.2$ and $\sigma_z = 0.3$, we estimate in the following manner the effect that 30% of contamination on $\overline{z}$: first, we assume the redshift distribution of the interlopers is also a Gaussian with $\overline{z} = 2.5$ and $\sigma_z = 0.3$. Next, adding 30% of these object to our reference distribution we find $\overline{z} = 3.0$ with $\sigma_z = 0.4$, i.e. a variation of 5%. Interlopers at lower redshifts, $\overline{z} = 1.8$ and $\sigma_z = 0.3$ produce a 10% variation in $\overline{z}$. These numbers are unchanged if instead we assume top-hat interloper distribution. Based on this discussion we adopt a 10% as upper limit of to our uncertainty in $\overline{z}$.

In Table 5 we present the values of the comoving correlation length r_0 of Lyman-break galaxies with $20.0 < I_{AB} < 24.5$. If we incorporate the uncertainty in $\overline{z}$ outlined above, an extra error of ±0.2 for the samples with $20.0 < I_{AB} < 24.5$, and of ±0.4 for $20.0 < I_{AB} < 23.5$ must be added. Results are shown for three cosmologies: Einstein-DeSitter ($\Omega_0 = 1.0$, $\Omega_\Lambda = 0.0$), open ($\Omega_0 = 0.2$, $\Omega_\Lambda = 0.0$) and Λ-flat ($\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$). We present correlation lengths computed for each field and for the mean of the three fields. We also show the results for two-parameter fits (slope and amplitude) for the 14 hr and 22 hr fields, and also for a fixed slope $\delta = 0.8$ for the 03 hr field and for the mean of all fields. Errors were computed using the Poissonian statistics (Eq. (3)).

We note that our two-parameter fits for r_0 and slope are not consistent with those of Adelberger et al. (2003) ($r_0 = (4.0 \pm 0.3)h^{-1}$ Mpc and $\delta = 0.55 \pm 0.15$); these measurements fall outside the error ellipses plotted in Fig. 7. Two phenomena could explain this discrepancy: firstly the sample of Adelberger et al. is slightly fainter than ours (which could produce a shift of the contour in Fig. 7 to the right – see Sect. 4.4) and

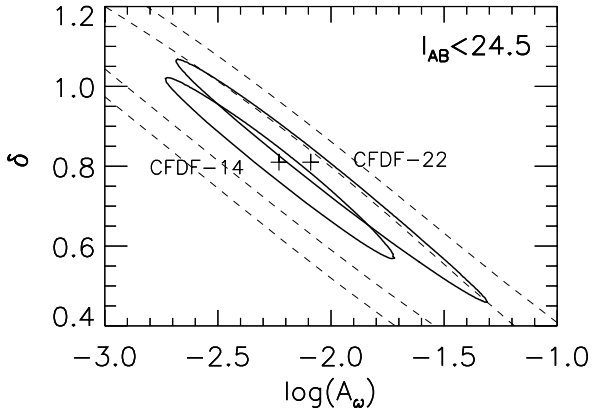

Fig. 7. Contours of χ^2 for the mean $\omega(\theta)$ computed for Lyman-break galaxies selected with $I_{AB} < 24.5$ in the CFDF-14 hr and CFDF-22 hr fields. The plus symbols show the best-fitting amplitudes and slopes, the contours correspond to the 1σ (thick contours) and 3σ confidence levels.

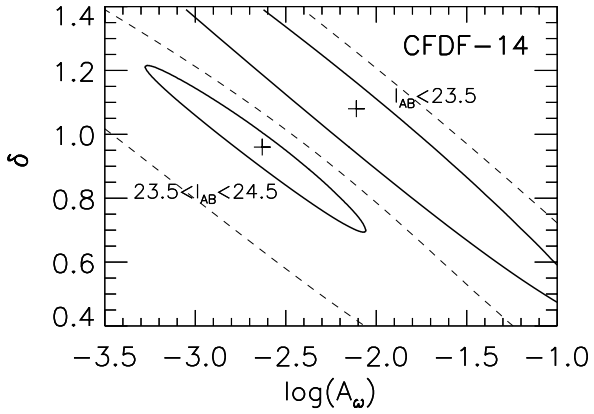

Fig. 8. Contours of χ^2 for $\omega(\theta)$ for two subsamples of Lyman-break galaxies with $20.0 < I_{AB} < 23.5$ and $23.5 < I_{AB} < 24.5$ in the CFDF-14 hr field. The plus symbol shows the best-fitting amplitudes and slope, and two contours correspond to the 1σ (thick contours) and 3σ confidence levels.

secondly their sample is a spectroscopic one and is expected to have a lower level of contamination than ours.

To summarise, for a Λ-flat cosmology, and for $\delta = 0.8$, we derive for the full $20.0 < I_{AB} < 24.5$ sample $r_0 = (5.9 \pm 0.5)h^{-1}$ Mpc, averaged over all three fields. For the two-parameter fits, we find $r_0 = (5.3^{+6.8}_{-2.2})h^{-1}$ Mpc with $\delta = 0.81^{+0.21}_{-0.24}$ and $r_0 = (6.3^{+17.9}_{-2.8})h^{-1}$ Mpc with $\delta = 0.81^{+0.25}_{-0.35}$ respectively.

4.4. Possible dependence on apparent magnitude of the correlation amplitude and length

In the 14 hr field, our r_0 measurements indicate samples with fainter limits magnitudes have *lower* values of r_0. Comparing our brightest $20.0 < I_{AB} < 23.5$ and our faintest $23.5 < I_{AB} < 24.5$ samples, we detect this effect with a 3σ confidence, as shown in Fig. 8. For these two sub-samples we also find approximately the same values of the slope. In Table 5, we show the values of the correlation amplitude and length for the two magnitude-limited samples, with the slope fixed to $\delta = 0.8$ and not fixed.

We note that this dependence of clustering strength with luminosity is also observed at lower redshifts (Norberg et al. 2002; Budavari et al. 2003). Moreover, recent results from the SDSS (Budavari et al. 2003) demonstrate that the slope of galaxy correlation function is independent of galaxy absolute luminosity, consistent with our observations.

4.5. Effect of contaminants on A_ω and r_0

To estimate the effect the contaminating population has on our measurements of r_0 and A_ω, we must make some assumptions of their clustering properties. In the case of the stellar contaminants, this is easy; however, for the interloper population it is less clear. Our selection criterion of $(V - I) < 1$ eliminates $z \sim 1.5$ bright ellipticals which might produce spuriously high correlations (additionally, we find no trend in our samples of $(V - I)$ with I_{AB} magnitude). Moreover, our simulations indicate that most of the interloper population lies at $z \sim 2$.

Assuming *all* the contaminants are unclustered, we can derive upper limits of the effect on the clustering. We find a fraction of contamination (by lower-redshift interlopers and stars) of $f \simeq 0.15$ for $20 < I_{AB} < 23.5$ and $f \simeq 0.3$ for the fainter $20 < I_{AB} < 24.5$ (Sect. 3.3). If these objects are not clustered, the estimates of clustering amplitudes A_ω, assuming a fixed slope of $\delta = 0.8$, have to be readjusted by a factor $1/(1-f)^2 \simeq 1.38$ for the brighter sample and $1/(1-f)^2 \simeq 2.04$ for the fainter one. This implies factors of ≃1.20 and ≃1.49 respectively for the correlation lengths r_0 for bright and faint samples (in our fainter magnitude bins, the interloper population is composed primarily of galaxies, which are more strongly clustered than stars but less strongly clustered than the Lyman-break population; this may further reduce the factor of 1.49).

An empirical way to estimate the effect of the contamination on our measurements is to replace a fraction of our candidates by objects extracted from the whole catalogue. We carried out this exercise for the 14 hr field by replacing 30% of the objects with $20 < I_{AB} < 24.5$ and 15% for $20 < I_{AB} < 23.5$, computing clustering with a slope of $\delta = 0.8$ and for a Λ-flat cosmology. In the first case we find $r_0 = (10.3 \pm 2.2)h^{-1}$ Mpc, i.e. a factor of ≃1.13 times lower, and in the second case $r_0 = (4.1 \pm 0.5)h^{-1}$ Mpc, i.e. a factor of ≃1.22 times lower. Of course in this experiment we cannot control the nature of the contaminants but these results indicate that this level of contamination could not produce the 3σ segregation effect reported in Sect. 4.4.

4.6. Are Poisson errors appropriate for the Landy and Szalay estimator?

In this section we investigate if the errors in the Landy and Szalay estimator (Eq. (3)) can be reliably described by Poissonian statistics. In doing this, we neglect other contributions, such as the finite size of the survey and the clustered nature of the galaxy distribution. We estimate here the relative amount of the various contributions to the error budget, using the analytical expression derived by Bernstein (1994).

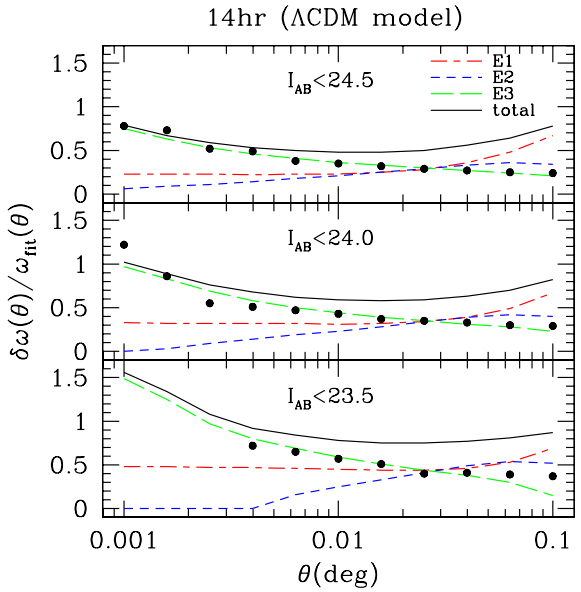

Fig. 9. Comparison of the different errors contributing to the global cosmic errors as a function of angular scales for three different magnitude cuts of the CFDF-14 hr sample. The main source of errors are: the finite volume error $E_1^{1/2}$ (long-short-dashed line); the discreteness errors $E_2^{1/2}$ (short-dashed line), $E_3^{1/2}$ (long-dashed line), and the total cosmic error $E^{1/2} \equiv (E_1+E_2+E_3)^{1/2}$ (solid line) (see Bernstein 1994). Filled black circles are the Poissonian errors estimated from Eq. (3). The analytical errors are computed using a "ΛCDM bias model" (see Sect. 4.6).

This expression has three terms: one reflecting the finite volume error (E_1: "*cosmic variance*"), which is independent of the number of galaxies, and two others related to the discrete nature of the galaxy catalogue: the first one appears only in the case of correlated sets of points (E_2, which cancels if $\omega \to 0$) and the second one includes the pure Poisson error (E_3). The calculation of the cosmic error requires prior knowledge of higher-order statistics (S_3 and S_4) as well as their redshift behaviours. We follow the recipes described in Colombi et al. (2000) and Arnouts et al. (2002) to perform this computation. Of course our total error budget will be dominated by the effects of systematic errors arising from our imperfect knowledge of the source redshift distribution and the precise quantity of contaminants in our sample, as we have discussed extensively elsewhere in our paper.

In Fig. 9 we show the relative magnitudes of the three components $E_1^{1/2}$ (short-long dashed lines), $E_2^{1/2}$ (short dashed lines) and $E_3^{1/2}$ (long-dashed lines) and the total error ($E^{1/2} = \sqrt{E_1+E_2+E_3}$, solid lines) as a function of the angular scale, θ, for three limiting magnitudes. The results are shown for the "ΛCDM bias model" described in Arnouts et al. (2002). The bias values used in this analysis for the different samples are given in Table 5. The theoretical estimates are compared to the observed errors of the CFDF-14 hr field ($\delta\omega/\omega_{\rm fit}$).

The behaviour of the observed errors matches closely the Poisson term E_3 at all angular scales for each of the three magnitude limited samples. At $\theta \leq 0.02°$, E_3 dominates the total error. The contribution of the finite volume error E_1 starts to play a significant role at relatively large scales: $0.04° \leq \theta \leq 0.1°$. The contribution of E_2 is never dominant at any scale. For $0.001° < \theta < 0.02°$, our analysis shows that the total cosmic error (E) is dominated by Poisson noise (E_3) and the amplitude of $E(\theta \sim 0.02°)^{1/2}$ is not more than a factor 1.6 larger than the amplitude of $E_3^{1/2}$. This result justifies the choice of using the nearly Poissonian errors.

5. Discussion and comparison with theory

5.1. Introduction

In this section we compare our measurements of the galaxy correlation length r_0 (in h^{-1} Mpc) with those of other authors and we interpret these derived values in terms of several simple models. Throughout this section, unless stated otherwise, all measurements of r_0 are presented for a Λ-flat cosmology ($\Omega_0 = 0.3$ and $\Omega_\Lambda = 0.7$). When necessary, we transform measurements from other authors to this cosmology using the equations presented in Magliocchetti et al. (2000). As we have already demonstrated in Sect. 4.2, our measured slopes are consistent with $\delta = 0.8$; to comparing our results with literature measurements and models, we use this corresponding value of the slope.

5.2. Tracing the evolution of r_0 with redshift

In Fig. 10, we plot the comoving correlation length r_0 of Lyman-break galaxies in the CFDF. The filled circle shows the mean measurement for all fields, for $20.0 < I_{AB} < 24.5$; the filled square and filled triangle shows measurements at $20.0 < I_{AB} < 23.5$ and $23.5 < I_{AB} < 24.5$ respectively for the CFDF-14 hr field. These measurements are shown at the mean assumed redshift of the CFDF Lyman-break sample $\overline{z} = 3.2$. For clarity each of the samples is slightly offset from each other. In addition, the dotted line on the right of this figure represents the errorbar for the CFDF-14 hr $20.0 < I_{AB} < 24.5$ sample when both δ and the amplitude are fitted, and corresponds to a projection of the 1σ contour plot shown in Fig. 7 along the amplitude axis.

For comparison we also show r_0 measurements for the local Universe from the Stromlo-APM survey (Loveday et al. 1995). Clustering measurements from the $I_{AB} < 22.5$ selected CFRS and the CNOC absolute magnitude-limited $M_R^{k,e} < -20$ surveys provide measurements of the evolution of clustering to $z < 1$ (Le Fèvre et al. 1996; Carlberg et al. 2000). We also show an average of measurements based on photometric redshift studies of the HDF-N and -S (Arnouts et al. 1999, 2002); galaxies in this study have $I_{AB} \leq 28$ (clustering measurements at $z \sim 4$ from this study are not shown because of the very small numbers of galaxies in this redshift bin). Finally, correlation length derived for $z \sim 3.8$ galaxies with $i'_{AB} \leq 26$ in the Subaru deep field (Ouchi et al. 2001) is shown.

A comparison of previous clustering measurements of Lyman-break galaxies are also presented. We note that in the literature there are several different analyses of the same dataset or supersets of the same dataset (either the HDF fields or the fields analysed by Steidel and collaborators). The open stars show measurements from Adelberger et al. (2003), who fit for

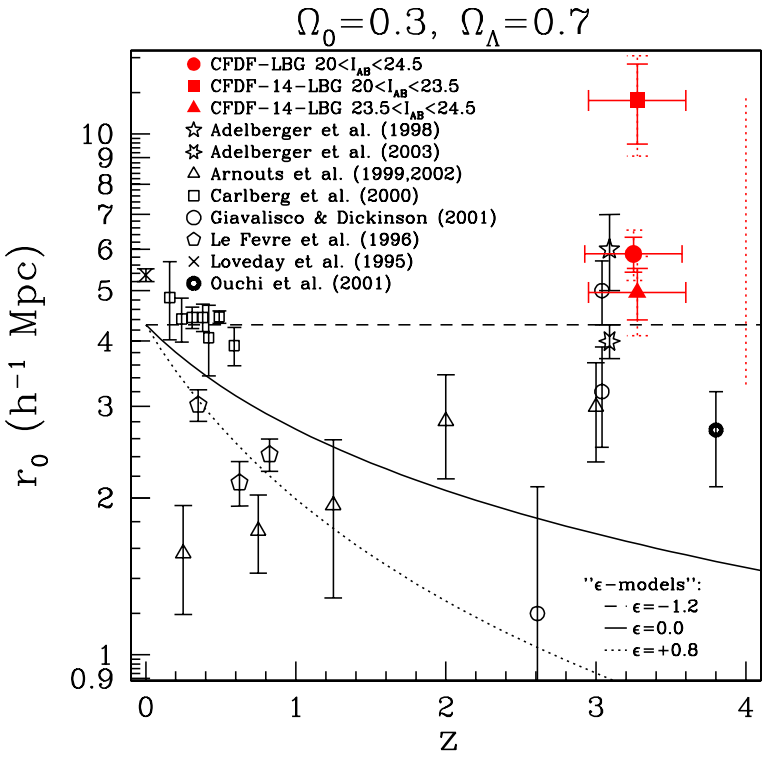

Fig. 10. The comoving correlation length, r_0 (in h^{-1} Mpc), for three samples of Lyman-break galaxies in the CFDF with the slope value fixed to $\delta = 0.8$ (symbols slightly offset for clarity). The circle, square and the triangles represent the mean correlation length for each of the three fields, the correlation length for a magnitude-limited sample of the CFDF-14 hr fields with $20.0 < I_{AB} < 23.5$ and the correlation length for a magnitude-limited sample of the CFDF-14 hr fields with $23.5 < I_{AB} < 24.5$ respectively. We plot a range of r_0 measurements from the literature, which are described in detail in Sect. 5.2. The horizontal error bars represent an uncertainty of 10% in the mean redshift. The solid vertical errorbars on r_0 are computed using Poisson statistics. Dotted vertical error bars represent the addition of this redshift error to the Poissonian component. The dotted line on the right part of the figure represents the errorbars for the CFDF-14 hr whole sample when we fit for both the slope and amplitude.

both slope and amplitude; the Adelberger et al. (1998) measurement was carried out on a subsample of this, with the slope fixed to $\delta = 0.8$ (for clarity those measurements were slightly offset). The three open circles show the clustering measurements from Giavalisco & Dickinson (2001); the upper circle represents their r_0 measurement from a $\mathcal{R}_{AB} \lesssim 25.1$ spectroscopically selected sample of Lyman-break galaxies (their "SPEC" sample), another subset of the Adelberger et al. (2003) sample. The middle open circle is the Giavalisco & Dickinson $\mathcal{R}_{AB} < 25.5$ photometrically selected Lyman-break sample (the "PHOT" sample), and the lower circle is Giavalisco & Dickinson's measurement of Lyman-break galaxies with $V_{AB606} < 27$ in the HDF. We caution that the Giavalisco & Dickinson use a slope of $\delta \sim 1.2$, different from this work. This explains the discrepancy between the HDF clustering measurement by Giavalisco & Dickinson and that of average HDF-N and -S values from Arnouts et al., who computed fitted correlation quantities assuming $\delta = 0.8$.

Figure 10 also shows "ϵ-model" predictions, i.e., $r_0(z) = r_0(z=0)(1+z)^{-(3+\epsilon-\gamma)/\gamma}$, for different values of ϵ, scaled arbitrarily to the value of $r_0 = 4.3h^{-1}$ Mpc at redshift $z = 0$ (Groth & Peebles 1977). In this simple prescription, three values of ϵ are normally considered: $\epsilon = -1.2$ for a slope $\gamma = 1.8$, corresponding to clustering fixed in comoving coordinates; $\epsilon = 0$, representing clustering fixed in proper coordinates; and $\epsilon = 0.8$ which corresponds to the predictions of linear theory, for an Einstein-DeSitter cosmology.

Taken together, these measurements present no clear picture of how r_0 evolves with redshift; for the CNOC survey, it appears that clustering is approximately fixed in comoving coordinates up to $z \sim 0.6$, whereas the results of the CFRS study indicate r_0 declines to $z \sim 1$. The HDF r_0 measurements appear to increase gradually over the entire redshift interval shown in our graph, and are always below the high-redshift values estimated from the CFDF. A number of separate factors contribute to this disparity: firstly, as we have highlighted, each individual sample has a different selection criterion; for example, galaxies at $z < 1$ from the HDF samples are much fainter and less numerous than those selected in the CFRS survey. It is clear from local spectroscopic surveys that galaxy clustering is a sensitive function of spectral type and intrinsic luminosity (Loveday et al. 1995, 1999; Norberg et al. 2002). Secondly, the field of view and the comoving scales probed are very different between each survey. At $z \sim 1$, for example, the HDF probes only $1h^{-1}$ Mpc, and this comoving scale increases at higher redshifts. Lastly, all surveys are subject to sampling and cosmic variance effects.

Precisely because of the effects outlined above it is difficult to directly compare our measurements of r_0 for Lyman-break galaxies to those of other authors. As mentioned

previously, an additional uncertainty is that not all authors adopt the same value of the slope δ, although the strong covariance between A_ω and δ allows us to estimate approximately the effect a changing slope will have on the fitted amplitude (see Fig. 7). Furthermore, all previous measurements of clustering at high redshift, based on photometric samples such as ours, are for fainter magnitudes than the faintest CFDF sample. Given the observed segregation of clustering amplitude with apparent magnitude observed in the CFDF-14 hr field, we would expect these previous studies to measure a lower amplitude than our work, and this is indeed what is observed. The photometric sample of Giavalisco & Dickinson (2001), reaching a half-magnitude fainter than our faintest sample, displays a clustering amplitude approximately twice as low as our faintest bin. However, the *spectroscopic* Lyman-break samples of Adelberger et al. (1998) and Giavalisco & Dickinson (2001) have approximately the same magnitude limits as our work, and we agree quite well with these measurements.

5.3. The surface density dependence of r_0

In this section we discuss the dependence of r_0 with galaxy surface density, a relationship which is more amenable to direct modelling, and discuss in more detail the implications of the segregation of galaxy clustering with apparent magnitude.

Figure 11 shows the comoving correlation length r_0 as a function of surface density for two magnitude limited samples ($20.0 < I_{AB} < 23.5$ and $20.0 < I_{AB} < 24.5$) extracted from the CFDF-14 hr and averaged over all three CFDF fields respectively. Error bars are computed using Poisson statistics. We added two clustering measurement of Lyman-break galaxies taken from the literature: Adelberger et al. (1998), and the average of two measurements for Lyman-break galaxies photometrically selected in the HDF-N and -S (Arnouts et al. 1999, 2002) (here we only show measurements of r_0 computed assuming a slope $\gamma = 1.8$). At densities of $\sim$1 arcmin^{-2} our r_0 measurements are in excellent agreement with those of Adelberger et al. Moreover, our measurements show a trend of increasing correlation length with decreasing galaxy surface density.

As an attempt to interpret these results, we consider the Λ-CDM analytic model of structure evolution presented in Arnouts et al. (1999) and discussed fully in Matarrese et al. (1997) and Moscardini et al. (1998) (their "transient" model). The relevant cosmological parameters for this model are given in Table 6 of Moscardini et al.'s paper. Similar models have also been presented elsewhere (Mo & White 1996; Mo et al. 1999). In this model, each Lyman-break galaxy is associated with one dark matter halo.

To briefly summarise the model's main components, we assume that the clustering of galaxies $\xi_g(r,z)$ is linearly related to the dark matter clustering $\xi_m(r,z)$ through the linear effective bias $b^2_{\rm eff}(z)$. The dark matter clustering is computed in the non-linear regime occupied by our survey using the fitting formulae of Peacock & Dodds (1996) and a power spectrum normalised to correctly reproduce the present-day abundance of bright clusters. The effective bias is calculated by integrating the product of the bias parameter $b(M,z)$ and the Press & Schechter (1974) dark matter halo redshift-mass distribution function over all the masses of halos larger than a typical minimum mass. To improve accuracy, the models use the fitting formulae of Sheth & Tormen (1999) for these quantities based on halos identified in a large N-body simulation. This model is shown as the solid line in Fig. 11.

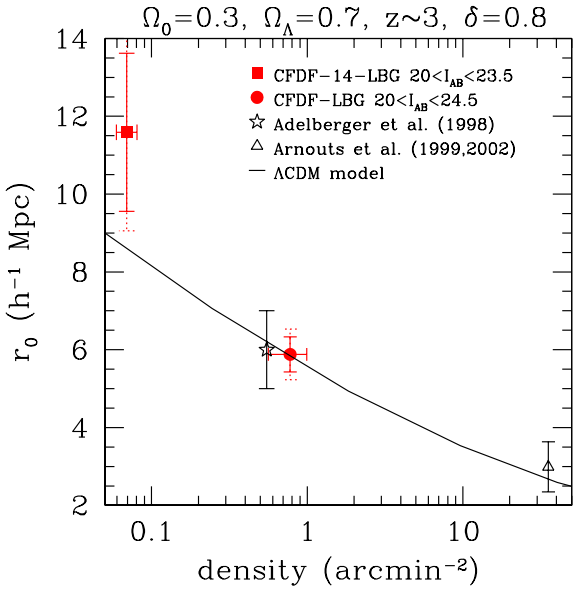

Fig. 11. The comoving correlation length r_0 (in h^{-1} Mpc) for two magnitude-limited CFDF samples (filled circle symbol for the mean over the three fields, and filled square symbol for the CFDF-14 hr field with $I_{AB} < 23.5$), as a function of cumulative surface density on the sky. Measurements from other Lyman-break samples (open symbols), and from the mean of HDF-N and -S (open triangles) are displayed. The solid vertical errorbars on r_0 are computed using Poisson statistics. Dotted vertical error bars represent the addition of the redshift error to the Poissonian component.

Despite its simplicity, this model reproduces quite well the observed strong dependence of r_0 on Lyman-break surface density seen in the CFDF survey, and this agreement continues to very faint $I_{AB} = 28$ measurements at surface densities of $\sim$40 arcmin^{-2} from the combined HDF measurement. Previously, such a dependence had not been unambiguously detected *within* a given survey.

These results argue against models of Lyman-break galaxy clustering such as the "bursting" scenario proposed by Kolatt et al. (1999), in which Lyman-break galaxies become visible as a result of stochastic star-formation activity. These models have difficulty in reproducing the strong dependence of galaxy clustering on surface density observed in our survey.

In the framework of biased galaxy formation, our results are consistent with a picture where more biased galaxies are more luminous and inhabit more massive dark matter halos with a simple one-to-one correspondence. A simple way to explain this relationship could be that there is a direct link between the luminosity of the galaxies and the mass of the halo. As the magnitudes we are measuring correspond to the rest-frame ultraviolet luminosity and as we assume here there is only one galaxy per halo, the most natural explanation of this

Table 6. Bias for each field and for each magnitude limited sample considered in this paper, for $\overline{z} = 3.2$ and for the best fitting value of the slope. The result marked as "CFDF mean" is computed from the mean over all three fields. The error bars shown correspond to Poisson error bars. To account for our uncertainty in the underlying redshift distribution of our Lyman-break sources, an extra systematic error of ±0.1 for the whole and faint samples and of ±0.2 for the bright sample should be added.

Field	magnitude cuts (I_{AB})	γ	b $\Omega_0 = 1.0$, $\Omega_\Lambda = 0.0$	b $\Omega_0 = 0.2$, $\Omega_\Lambda = 0.0$	b $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$
CFDF-14	20.0–24.5	$1.81^{+0.21}_{-0.24}$	$4.6^{+5.2}_{-1.6}$	$1.8^{+2.0}_{-0.6}$	$3.2^{+3.6}_{-1.1}$
	20.0–23.5	$2.08^{+0.84}_{-0.66}$	$8.7^{+92.6}_{-4.3}$	$3.6^{+38.7}_{-1.8}$	$6.3^{+67.5}_{-3.1}$
	23.5–24.5	$1.96^{+0.25}_{-0.26}$	$3.8^{+5.3}_{-1.4}$	$1.5^{+2.1}_{-0.6}$	$2.7^{+3.8}_{-1.0}$
CFDF-22	20.0–24.5	$1.81^{+0.25}_{-0.35}$	$5.4^{+13.7}_{-2.0}$	$2.1^{+5.4}_{-0.8}$	$3.7^{+9.4}_{-1.4}$
CFDF-03	20.0–24.5	1.8	5.5 ± 0.2	2.1 ± 0.1	3.8 ± 0.1
CFDF mean	20.0–24.5	1.8	5.1 ± 0.4	2.0 ± 0.2	$\mathbf{3.5 \pm 0.3}$
CFDF-14	20.0–23.5	1.8	9.9 ± 1.5	3.6 ± 0.6	6.4 ± 1.0
	23.5–24.5	1.8	4.3 ± 0.4	1.7 ± 0.2	3.0 ± 0.3

relationship could be a direct link between the stellar masses of the Lyman-break galaxy population and the rest frame ultraviolet luminosity (Papovich et al. 2001). However, stellar population synthesis modelling of Lyman-break galaxies population has failed to definitively establish such a relationship: as suggested by Shapley et al. (2001) these models are dependent on the assumed extinction law, which is currently unknown for Lyman-break galaxies.

How realistic is our assumption that each Lyman-break galaxy traces exactly one dark-matter halo? Applying semi-analytic models of galaxy formation to the clustering of Lyman-break galaxies, Baugh et al. (1999) found that, at higher redshifts, these models gave almost identical clustering amplitudes to these simpler "massive halo models". However, a more important question is how this clustering strength scales with halo abundance. More recent work has shown how the halo occupation function – the number of objects per halo – affects sensitively the slope of the model curve in Fig. 11 (Wechsler et al. 2001; Bullock et al. 2002). Models in which many Lyman-break galaxies inhabit a single halo show a weak dependence of clustering strength with object abundance and have difficulty reproducing the strong trend seen in our data.

It is also interesting to investigate the small-scale behaviour of $\omega(\theta)$ which can provide information on the halo occupation function (Bullock et al. 2002). It has been claimed that at small ($\theta < 10''$) separations $\omega(\theta)$ no longer follows a power law (Porciani & Giavalisco 2002). For the full $20 < I_{AB} < 24.5$ CFDF Lyman-break sample we have computed the ratio of pairs at small separation $N_{\rm p}(\theta < 10'')$ to those at larger separation, $N_{\rm p}(10'' < \theta < 60'')$. Based on the fitted values of $\omega(\theta)$ given in Table 5, we expect the ratio $N_{\rm p}(10'' < \theta < 60'')/N_{\rm p}(\theta < 10'')$ to be around 19. In the CFDF data (for a weighted average over all fields) we find this pair fraction is 26 ± 6. Based on these statistics, we conclude that the CFDF dataset provides no convincing evidence for a small-scale departure from a power-law behaviour with $\delta = 0.8$, a conclusion consistent with the observed small-scale behaviour of $\omega(\theta)$ in Fig. 5.

We note that our bright measurement in Fig. 11 deviates from our model curve at the $\sim 1.5\sigma$ level. We investigate the origin of this effect, measuring the median $(V-I)_{AB}$ colour for each of our magnitude-limited samples. Our brighter samples are no redder than our fainter samples, suggesting that contamination by nearby bright ellipticals in this sample is minimal (furthermore, all magnitude limited samples are subject to the criterion $(V-I)_{AB} < 1.0$, from Eq. (1)). A more likely origin for this discrepancy is that in computing r_0, we assume that the redshift distribution of each magnitude limited sample is the same; a slightly lower mean redshift would imply a lower value for r_0.

Finally, we remark that in our fainter bin, our stated level of incompleteness of ~20% at $20.0 < I_{AB} < 24.5$ (Table 3) indicates that our surface densities may be underestimated by around ~0.2. Furthermore, if Lyman-break galaxies were especially dusty (although this is not supported by current observations; see Webb et al. 2003) we would expect the true Lyman-break galaxy density to be further underestimated. However, these considerations do not affect the principal conclusions of this work, as these effects are expected to be much smaller than the observed variation of clustering strength with apparent magnitude.

5.4. Linear bias estimates for the CFDF Lyman-break sample

The theoretical procedures described in the previous section can also be used to estimate of the effective bias, b, of the Lyman-break galaxy sample. From the comoving correlation length r_0 we can compute the observed rms galaxy density fluctuation within a sphere of $8h^{-1}$ Mpc, $\sigma_8^{\rm gal}$ (Magliocchetti et al. 2000). Dividing this quantity by the rms mass density fluctuation, computed from cluster-normalised models assuming the linear theory, we may derive the linear bias b. In Table 6 we

present these results, together with Poisson errors, for a range of cosmologies.

In comparison, Adelberger et al. (1998), with a sample of spectroscopically confirmed Lyman-break galaxies at $z \simeq 3$ for $\mathcal{R}_{AB} < 25.5$, find $b = 4.0 \pm 0.7$ for $\Omega_0 = 0.3$ and $\Omega_\Lambda = 0.7$. From the average of the fainter $I_{AB} \leq 28$ galaxies selected in the HDF-N and -S, Arnouts et al. (1999, 2002) find $b = 1.9 \pm 0.4$ for $\Omega_0 = 0.3$ and $\Omega_\Lambda = 0.7$.

Many studies agree on the strongly biased nature of the Lyman-break galaxy population, and provide evidence for a picture in which structures form hierarchically and massive objects form at highest peaks in the underlying density field (Kaiser 1984; Bardeen et al. 1986). Our measurements of Lyman-break galaxies at $I_{AB} = 24.5$ appear to support this picture. For very bright Lyman-break galaxies, at $I_{AB} = 23.5$, we find correlation lengths of $>10h^{-1}$ Mpc and a linear bias of $b \sim 6$ in the Λ-flat cosmology. These biases would imply underlying dark matter halo masses for the Lyman-break galaxy of around $10^{13}h^{-1}\ M_\odot$, about a factor of ten above the most massive haloes of Lyman-break galaxy observed to date, but still comparable to the masses of present day $M^\star$ galaxies. We note that the clustering lengths of our brighter Lyman-break galaxies are comparable to those of the "extremely red object" (ERO) population (e.g. $r_0 = 13.8 \pm 1.5h^{-1}$ Mpc in a Λ-flat cosmology – Daddi et al. 2001) and we speculate that, unlike the fainter Lyman-break objects studied previously, some fraction of these bright Lyman-break galaxies may evolve into EROs by $z \sim 1$, according to a galaxy conservation model with a fixed bias at burst (Mo & White 1996).

6. Summary and conclusions

We have extracted a large sample of $z \sim 3$ Lyman-break galaxies from the Canada-France Deep Fields survey. Our catalogues cover an effective area of $\sim$1700 arcmin^2 in three separate large, contiguous fields. In total the survey contains 1294 Lyman-break candidates to a limiting magnitude of $I_{AB} = 24.5$. Our conclusions are as follows (assuming $\Omega_0 = 0.3$, $\Omega_\Lambda = 0.7$):

1. Number counts and surface densities of $z \sim 3$ galaxies selected in the CFDF agrees very well with literature measurements over the entire $20.0 < I_{AB} < 24.5$ mag range of our survey.

2. Using simulated catalogues, we demonstrate that at the limiting magnitude our catalogue contains contaminants at a level of $\sim$30% or less.

3. We measure the two-point galaxy correlation function $\omega(\theta)$ of Lyman-break galaxies and show it is well described in term of a power law of slope $\delta = 0.8$ even at small angular separations, where no excess of close pairs is found.

4. Assuming that Lyman-break galaxies in the CFDF survey are at $\overline{z} = 3.2$, we derive the comoving correlation length, r_0, for a range of magnitude limited samples. For the whole $20.0 < I_{AB} < 24.5$ sample, we find $r_0 = (5.9 \pm 0.5)h^{-1}$ Mpc with the slope fixed to $\gamma = 1.8$. For simultaneous fits of the slope and amplitude , we find for the CFDF-14 hr field $\gamma = 1.8 \pm 0.2$ and $r_0 = (5.3^{+6.8}_{-2.2})h^{-1}$ Mpc, and for the CFDF-22 hr field $\gamma = 1.8 \pm 0.3$ and $r_0 = (6.3^{+17.9}_{-2.8})h^{-1}$ Mpc, in good agreement with the values determined with the slope fixed.

5. In the CFDF-14 hr field, we find a marginal dependence of r_0 on apparent magnitude: for Lyman-break galaxies with $20.0 < I_{AB} < 23.5$, we derive $r_0 = (11.6 \pm 2.0)h^{-1}$ Mpc, whereas for $23.5 < I_{AB} < 24.5$ we find $r_0 = (5.0 \pm 0.6)h^{-1}$ Mpc (in both cases for slopes fixed to $\gamma = 1.8$). Allowing both the slope and amplitude to vary, this segregation is still detected at the 3σ-level.

6. We investigate the dependence of r_0 on surface density, n, and find a strong correlation. For $n = (0.09 \pm 0.02)$ arcmin^{-2}, $r_0 = (11.6 \pm 2.0)h^{-1}$ Mpc, whereas for $n = (0.78 \pm 0.24)$ arcmin^{-2}, we find $r_0 = (5.9 \pm 0.5)h^{-1}$ Mpc.

7. A simple analytic model in which each Lyman-break galaxy traces one dark matter halo is able to reproduce the observed dependence of correlation length on abundance quite well, except for our very bright sample of Lyman-break galaxies, which deviates from the predictions of our models by around $\sim 1.5\sigma$.

8. We derived a linear bias b by dividing the measured rms galaxy density fluctuation $\sigma_8^{\rm gal}$ by the rms mass fluctuation $\sigma_8^{\rm m}$ computed by assuming cluster-normalised linear theory. For our sample of Lyman-break galaxies, we find for $20.0 < I_{AB} < 24.5$, $b = 3.5 \pm 0.3$.

Acknowledgements. SF and HJMCC wish to acknowledge the use of TERAPIX computer facilities at the Institut d'Astrophysique de Paris, where part of the work for this paper was carried out. We would like to thank Chuck Steidel for providing us with his Lyman-break catalogue covering the 14 hr field. We would also like to thank our referee for a detailed and thorough report which improved our paper. HJMCC's work has been supported by MIUR postdoctoral grant COFIN-00-02 and a VIRMOS postdoctoral fellowship.

References

Adelberger, K. L., Steidel, C. C., Giavalisco, M., et al. 1998, ApJ, 505, 18
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355
Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
Baugh, C. M., Benson, A. J., Cole, S., Frenk, C. S., & Lacey, C. G. 1999, MNRAS, 305, L21
Bernstein, G. M. 1994, ApJ, 424, 569
Bershady, M. A., Charlton, J. C., & Geoffroy, J. M. 1999, ApJ, 518, 103
Bertin, E., & Arnouts, S. 1996, A&A, 117, 393
Brunner, R. J., Szalay, A. S., & Connolly, A. J. 2000, ApJ, 541, 527
Bruzual, G., & Charlot, S. 1993, ApJ, 405, 538
Budavari, T., Connolly, A. J., Szalay, A. S., et al. 2003, ApJ, accepted [astro-ph/0305603]
Bullock, J. S., Wechsler, R. H., & Somerville, R. S. 2002, MNRAS, 329, 246
Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582
Carlberg, R. G., Yee, H. K. C., Morris, S. L., et al. 2000, ApJ, 542, 57

Cole, S., Aragon-Salamanca, A., Frenk, C., Navarro, J., & Zepf, S. 1994, MNRAS, 271, 781
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Colombi, S., Charlot, S., Devriendt, J., Fioc, M., & Szapudi, I. 2000, in Clustering at High Redshift, ASP Conf. Ser., 200, 153
Crampton, D., Le Fèvre, O., Lilly, S. J., & Hammer, F. 1995, ApJ, 455, 96
Daddi, E., Broadhurst, T., Zamorani, G., et al. 2001, A&A, 376, 825
Giavalisco, M., & Dickinson, M. 2001, ApJ, 550, 177
Giavalisco, M., Steidel, C. C., Adelberger, K. L., et al. 1998, ApJ, 503, 543
Groth, E. J., & Peebles, P. J. E. 1977, ApJ, 217, 385
Guhathakurta, P., Tyson, J. A., & Majewski, S. R. 1990, ApJ, 357, L9
Kaiser, N. 1984, ApJ, 284, L9
Kolatt, T. S., Bullock, J. S., Somerville, R. S., et al. 1999, ApJ, 523, L109
Kron, R. G. 1980, ApJS, 43, 305
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Le Fèvre, O., Hudon, D., Lilly, S. J., et al. 1996, ApJ, 461, 534
Lehnert, M. D., & Bremer, M. 2002, AJ, submitted [astro-ph/0212431]
Lilly, S. J., Le Fèvre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, ApJ, 455, 50
Loveday, J., Maddox, S. J., Efstathiou, G., & Peterson, B. A. 1995, ApJ, 442, 457
Loveday, J., Tresse, L., & Maddox, S. 1999, MNRAS, 310, 281
Madau, P. 1995, ApJ, 441, 18
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388
Magliocchetti, M., Bagla, J. S., Maddox, S. J., & Lahav, O. 2000, MNRAS, 314, 546
Magliocchetti, M., & Maddox, S. J. 1999, MNRAS, 306, 988
Matarrese, S., Coles, P., Lucchin, F., & Moscardini, L. 1997, MNRAS, 286, 115
McCracken, H. J., Le Fèvre, O., Brodwin, M., et al. 2001, A&A, 376, 756 (Paper I)
McCracken, H. J., Shanks, T., Metcalfe, N., Fong, R., & Campos, A. 2000, MNRAS, 318, 913
Metcalfe, N., Shanks, T., Campos, A., McCracken, H. J., & Fong, R. 2001, MNRAS, 323, 795
Mo, H. J., Mao, S., & White, S. D. M. 1999, MNRAS, 304, 175
Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Moscardini, L., Coles, P., Lucchin, F., & Matarrese, S. 1998, MNRAS, 299, 95
Nandra, K., Mushotzky, R. F., Arnaud, K., et al. 2002, ApJ, 576, 625
Norberg, P., Baugh, C. M., Hawkins, E., et al. 2002, MNRAS, 332, 827
Ouchi, M., Shimasaku, K., Okamura, S., et al. 2001, ApJ, 558, L83
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Peacock, J. A., & Dodds, S. J. 1996, MNRAS, 280, L19
Peebles, J. 1980, The Large-Scale Structure of the Universe (Princeton)
Porciani, C., & Giavalisco, M. 2002, ApJ, 565, 24
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Robin, A., & Creze, M. 1986, A&A, 157, 71
Roche, N., Shanks, T., Metcalfe, N., & Fong, R. 1993, MNRAS, 263, 360
Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
Shapley, A. E., Steidel, C. C., Adelberger, K. L., et al. 2001, ApJ, 562, 95
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119
Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996a, AJ, 112, 352
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996b, ApJ, 462, L17
Steidel, C. C., & Hamilton, D. 1993, AJ, 105, 2017
Stevens, R., & Lacy, M. 2001, MNRAS, 325, 897
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485
Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, AJ, 117, 68
Webb, T. M., Eales, S., Foucaud, S., et al. 2003, ApJ, 582, 6
Wechsler, R. H., Somerville, R. S., Bullock, J. S., et al. 2001, ApJ, 554, 85