The zCOSMOS Data Release 2: the "zCOSMOS-bright 10k-sample" and structure in the Universe out to redshifts of order unity

Simon Lilly (ETH Zurich, Switzerland)
and the zCOSMOS team*

The global COSMOS project is aimed at understanding the evolution of galaxies and active galactic nuclei, and in particular the role of the galactic environment in that evolution. It is built around observations of a single equatorial 1.7 deg^2 field corresponding to transverse dimensions of 100 x 100 Mpc^2 at high redshift. The COSMOS field is emerging as the premier extragalactic survey field, and is currently the object of study of large observational programmes on most of the major observational facilities around the world. The zCOSMOS programme on the VLT is securing spectroscopic redshifts for a very large number of galaxies and quasars over the whole redshift range $0 < z < 3.5$, using the VIMOS spectrograph. This allows the environment of galaxies to be characterised on all scales from that of the immediate group environment, 100 kpc, up to the 100 Mpc scales of the cosmic web. The second public data release, DR2, of approximately 10 000 zCOSMOS spectra, took place via the ESO Science Data Archive in October 2008. This article describes the current status of the project and in particular of this so-called "10k-sample", and our reconstruction of large-scale structure in the Universe out to $z \sim 1$.

* Katarina Kovac[1], Christian Knobel[1], Angela Iovino [2], Vincent Le Brun[3], Christian Maier[1], Vincenzo Mainieri[4], Marco Mignoli[5], Pawel Kampczyk [1], Marco Scodeggio[6], Gianni Zamorani [5], Olivier Ilbert [7], Mara Salvato[8], Pascal Oesch [1], Marcella Carollo[1], Thierry Contini[9], Jean-Paul Kneib [3], Olivier Le Fèvre [3], Alvio Renzini[10], Sandro Bardelli[5], Micol Bolzonella[5], Angela Bongiorno[11], Karina Caputi[1], Graziano Coppa[5], Olga Cucciati[2], Sylvain de la Torre[3], Loic de Ravel[3], Paolo Franzetti [6], Bianca Garilli [6], Fabrice Lamareille[9], Jean-Francois Le Borgne[9], Roser Pello [9], Yingjie Peng [1], Enrique Perez Montero [9], Elena Ricciardelli[10], John Silverman[1], Masayuki Tanaka [4], Lidia Tasca[3], Laurence Tresse[3], Daniela Vergani[5], Elena Zucca[5], Umi Abbas[3], Dario Bottini[6], Peter Capak [8], Alberto Cappi[5], Paolo Cassata [3], Andrea Cimatti[12], Martin Elvis[13], Marco Fumana[6], Luigi Guzzo[2], Gunther Hasinger [11], Alexei Leauthaud[14], Dario Maccagni[6], Henry McCracken [15], Pierdomenico Memeo[6], Baptiste Meneux[11], Cristiano Porciani [16], Lucia Pozzetti [5], David Sanders[7], Roberto Scaramella[17], Claudia Scarlata[8], Nick Scovile [8]

1 ETH Zurich, Switzerland
2 INAF Osservatorio Brera, Italy
3 LAM Marseille, France
4 ESO
5 INAF Osservatorio Bologna, Italy
6 INAF IASF Milano, Italy
7 Institute of Astronomy, University of Hawaii, Honolulu, USA
8 Caltech, California, USA
9 CNRS Université de Toulouse, France
10 Universita di Padova, Italy
11 Max-Planck-Institut für extraterrestrische Physik, Garching, Germany
12 Universita di Bologna, Italy
13 Harvard Smithsonian Center for Astrophysics, USA
14 University of California, Berkeley, USA
15 Institut d'Astrophysique de Paris, France
16 University of Bonn, Germany
17 INAF Osservatorio di Roma, Italy

There are many reasons to suspect that the environment plays a major role in driving the evolution of galaxies and active galactic nuclei (AGN). In the local, present-day Universe, clear trends are seen between many galaxy properties, such as their star formation rates and structural morphologies, and different measures of the environment of galaxies. These were first quantified over twenty years ago by Alan Dressler, and have now been seen clearly in the large-scale local surveys of the Sloan Digital Sky Survey (SDSS). Broadly speaking, galaxies in higher density environments are, today, less actively forming stars and more often have spheroidal 'early-type' structures rather than 'late-type' disc-dominated morphologies, than galaxies in lower density parts of the Universe. The actual physical cause of these effects is however far from clear, and the basic question of the relative importance of 'nature' and 'nurture' in controlling galactic development is far from settled.

There is no shortage of plausible mechanisms whereby the galactic environment could influence the evolution of a particular galaxy: on the one hand, the accumulation of material via the hierarchical assembly of dark matter haloes and the accretion of gas, whether through cooling of shock-heated gas or through cold mass flows out of the cosmic web, must depend on the immediate environment of the system in question; on the other hand, the effects of interactions with the surrounding intergalactic medium (IGM) through starburst-driven winds or from energy injection from AGN, may provide a feedback onto the properties of the galaxies. Major mergers of galaxies, which almost certainly play a very large role in the morphological transformation of galaxies, and possibly also in the control of their star formation rates, will occur at very different rates in different environments because of the different number densities and velocity dispersions of the galaxies. Even a purely internal rearrangement of material within galaxies through dynamical instabilities may have been triggered by nearby neighbours. In the richest environments, the intracluster medium may strip out material from galaxies, while close, high speed 'fly-bys' may also have transformational consequences, even in the absence of mergers. Finally, the overall time scale for the growth of large-scale structure in the Universe via gravitational instability is set by the amplitude of the density field on large scales. No doubt all of these processes play a role, but their relative importance is not known. Even in the present-day Universe, there is controversy as to the scale on which the environmental signature is present — solely on the scales below 1 Mpc characteristic of a single dark matter halo, or also on larger scales indicating a role also for the larger cosmic web of structure.

Looking to larger distances, and thus observing the Universe at earlier epochs directly, and establishing when and how these observed relations are established should go a long way to resolving these questions. For many of the physical processes listed above, it is the environment on the scale of galaxy groups that is suspected of being most relevant. In the standard ΛCDM (Lambda Cold Dark Matter) paradigm for cosmological structure formation, these environments have undergone strong evolution since redshifts $z \sim 2$ and environmentally-driven evolution may plausibly therefore be the cause of the dramatic decline in galactic activity over this same period.

Unfortunately, environmental information on distant galaxies has been very limited up until now. This is mostly because of limited sample sizes, and the small size and sparse sampling of the deep extra-galactic survey fields. The COSMOS survey (Scoville et al., 2007) was designed to remedy both of these, by bringing to bear on a single large field all of the techniques, that have been developed over the last decade or more, to study distant galaxies over a wide range of redshifts. The COSMOS field is about 600 times larger than the famous Hubble Deep Fields, and about thirty times larger than each of the two Great Observatories Origins Deep Survey (GOODS) fields. In addition to the initial imaging with the Hubble Space Telescope (HST), COSMOS is now also quite unique in the breadth and depth of the imaging data that have been assembled using large amounts of observing time on the X-ray satellite observatories XMM-Newton and Chandra, with the ultraviolet Galaxy Evolution Explorer (GALEX) and infrared Spitzer space telescope, and, on the ground, with the Subaru, the Canada France Hawaii Telescope (CFHT) and the UK Infrared Telescope (UKIRT) optical/infrared telescopes. At longer wavelengths, the Very Large Array (VLA) radio telescope and various millimetre-wave facilities have also observed the field. In the future, the COSMOS field will be the major focus of the very deep UltraVISTA infrared imaging survey at ESO.

zCOSMOS provides the crucial 'third-dimension' to COSMOS by measuring accurate redshifts for large numbers of galaxies in the COSMOS field.

zCOSMOS at the VLT

zCOSMOS is a major project (ESO Large Program 175.A-0839) that is using 600 hours of observing time with the VIMOS spectrograph on the VLT UT3, spread over five observing seasons 2005–2009. It consists of two parts (see Lilly et al., 2007, for details): The first, "zCOSMOS-bright", obtains spectra of about 20 000 galaxies selected to have $I_{AB} < 22.5$ across the full 1.7 deg^2 of the COSMOS field. zCOSMOS-bright was designed to yield a high and fairly uniform sampling rate (about 70%), with a high success rate in measuring redshifts (approaching 100% at $0.5 < z < 0.8$), and with sufficient velocity accuracy (about 100 km/s) to efficiently map the environments of galaxies down to the scale of galaxy groups out to redshifts $z \sim 1$. The second part, zCOSMOS-deep, will consist of about 10 000 spectra of higher redshift galaxies, colour-selected to have redshifts in the $1.4 < z < 3.0$ range, and lying in the central 1 deg^2 region of the COSMOS field.

After the first two zCOSMOS observing seasons in 2005 and 2006, about a half of the zCOSMOS-bright observations had been completed, yielding a total of over 10 500 spectra from which redshift measurements have been made, or attempted — the so-called "10k-sample" (Lilly et al., 2008). This sample was released, with the help of the ESO External Data Products Group, to the wider science community via the ESO Science Archive (*http://archive.eso.org/cms/eso-data/data-packages/zcosmos-data-release-dr2/*) on 1 October 2008. It is being used by the zCOSMOS team for a number of science investigations that are now at various stages of the publication process.

In the meantime, further observations have taken zCOSMOS-bright almost to completion, with only a handful of the 180 spectroscopic masks remaining to be observed at the start of next year. We therefore anticipate constructing the final "20k sample" after these observations are completed in 2009. Observations of zCOSMOS-deep were phased later in the programme, but are now over 50% complete. Hopefully, observations for this part of the survey should also be completed by the end of the 2009 observing season.

The zCOSMOS-bright 10k-sample

A great deal of effort by the team has gone into ensuring the high quality of the zCOSMOS data products. In particular, a redshift survey like zCOSMOS produces redshift identifications with a range of reliabilities, simply because of the faintness of the galaxies and because we are pushing the limits of what is possible.

The vast majority of redshifts are, of course, very secure, but some are unavoidably less reliable, and a few are little better than guesses. For some, we cannot offer even a tentative redshift identification. To get the best science out of such a large and well-defined sample, and to enable their use by others, it is essential to characterise the reliability of the redshifts, and to understand any biases present in the set of objects for which usable redshifts are secured — failures cannot be simply thrown away. To deal with this, every redshift measurement is assigned its own individual 'Confidence Class'. This is already the result of 'reconciling' two independent reductions of the observational data at two (of the six) zCOSMOS institutes. This duplication catches most of the potential problems in the reduction process. The Confidence Class scale varies from Class 0 (no redshift) up to Class 4 (very secure), with an additional Class 9 which designates 'one-line' redshifts, with various additional modifiers to reflect details such as whether the target is an AGN, or whether it was observed serendipitously in a slit targeted at another object (see Lilly et al., 2008, or the DR2 release notes for details). We then need to quantify the reliability of each of these classes. Our team has approached this in two ways.

First, repeat spectra for over 600 objects have been taken through a variety of different pathways. These repeat spectra are processed blind to the first reduction, providing an invaluable check as to whether the same redshift is found the second time around. With the simplifying assumption that the chance of getting the same wrong redshift twice is negligible, we can construct a simple probabilistic measure of the reliability. Our most secure Class 3 and 4 redshifts, which form the bulk of the sample, are indeed highly repeatable, > 99.8%. With the lower reliabilities, we find that generally we were conservative: Class 2, intended to be only 75% reliable is in fact confirmed 92% of the time, and even our Class 1 'guesses' are correct in 70% of cases. The repeat spectra also yield an empirical measure of the velocity accuracy of our redshifts, which comes out to be 110 km/s or $\Delta z = 0.00036\,(1 + z)$.

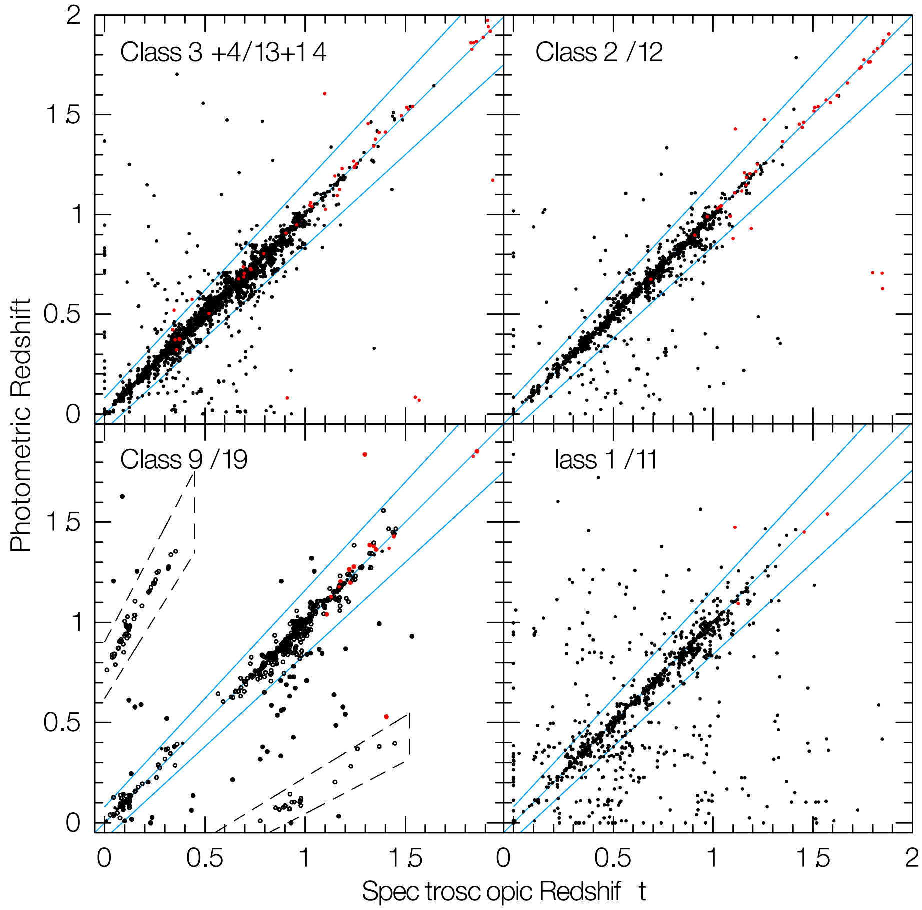


Figure 1. Comparison of photometric and spectroscopic redshifts of zCOSMOS-bright galaxies, as described in the text. Note that Class 9 redshifts are those based on a single emission line, which therefore have two alternative redshifts based on either a [O II] 3727 Å or H α 6563 Å identification. A “1” preceding the Class (e.g. 4 → 14) indicates a broad line AGN, indicated by red points on the figure.

Secondly, we also use independent redshifts that are estimated solely from the colours of the galaxies, (photo-*z*). These are based on the unparalleled richness of the photometric data available in the COSMOS field. We examine the consistency with our spectroscopically estimated redshifts, which is shown in Figure 1, and find that there is generally a very good consistency: Comparison with our most reliable redshifts shows that there is a floor of about 3% photo-*z* ‘failures’, but otherwise an impressively small statistical dispersion of about $\Delta z = 0.01\ (1 + z)$ in redshift (Ilbert et al., 2008). The failures can probably be traced to unavoidable problems with the photometry, such as close pairs of galaxies or areas of the sky close to bright stars. However, as we then look at our less reliable redshifts, we see an excellent correspondence between the spectroscopic confirmation rate described in the previous paragraph and the consistency with the photo-*z* to within $\Delta z = 0.08\ (1 + z)$. As well as confirming the empirical calibration of our spectroscopic reliability, this suggests three things: first, we can use the photo-*z* to indicate which of our less reliable redshifts are very probably right and which are probably wrong — information that we capture as a further decimal place modifier to the Confidence Class of individual redshifts; second, we can use the photo-*z* themselves to estimate the redshifts of those galaxies for which we have failed to measure the redshift, thereby quantifying the biases in the final spectroscopic sample — this is shown in Figure 2; finally, we could, though we haven’t yet done this, go back to the spectra of these failures and search again for a new spectroscopic redshift identification in the narrow redshift range indicated by the photo-*z*.

This powerful two-way complementarity between spectroscopic and photometric redshifts is a unique feature of COSMOS and zCOSMOS, and allows us to produce a spectroscopic catalogue with high accuracy redshifts and very high fidelity. Figure 3 shows the distribution of redshifts in the 10k-sample. The richness of the redshift structure in the distant Universe is immediately apparent, with structures spanning the range from about 100 km/s, up to the 20 000 km/s ($\Delta z \sim 0.05$) for the largest voids and overdensities.

Reconstruction of the density field to $z \sim 1$

We further exploit the complementarity of spectroscopic and photometric redshifts when we construct the overall galaxy density field in the COSMOS field out to redshifts $z \sim 1$. This is central to our scientific goal to determine the role of the environment in driving galaxy evolution and is already being used as the basis of several papers in an advanced state of preparation.

The density field, presumably a continuous distribution of dark matter, must be reconstructed using the discrete locations of relatively sparse galaxies, which inevitably introduces a smoothing into any reconstruction of the density field. The minimum physical scale of this smoothing depends on the number density of the tracer galaxies. There are of course many more galaxies with only a photo-*z* measurement than galaxies with a much more accurate spectroscopic redshift. Currently only one galaxy in three with $I_{AB} < 22.5$ has been observed in zCOSMOS-bright, and this will not get much above 60% even when the full survey is finished. Recognising that these photo-*z* galaxies nevertheless have very accurately determined locations in the (x,y) plane of the sky, we have developed a new algorithm (ZADE, see Kovac et al., 2008, for details) which seeks to combine the redshift accuracy of the spectroscopic redshifts with the increased numbers, leading to greater spatial 'resolution' in the density field, of a photo-z sample. This is done by modifying the probabilistic redshift distribution P(z) for each of the individual photo-z using the spectroscopic redshifts of nearby galaxies, making use of the fact that galaxies are clustered in the Universe.

One advantage of the new approach is that the complex selection function of the 10k sample, especially the currently, highly non-uniform, spatial sampling on the sky, is automatically taken into account. Extensive tests of the algorithm have been made against mock catalogues that have been generated from large-scale cosmological simulations (Kitzbichler et al., 2007), and on which we impose the different zCOSMOS selection functions. These tests have shown that the ZADE approach allows a much better reconstruction of the density field than traditional weighting methods, with very little systematic bias and with improved statistical accuracy. By using this new approach of combining spectroscopic and photometric redshifts, plus the use of a smoothing scale that is adapted to the local density of galaxies, we can achieve a spatial resolution in the density field that extends down to 1 h^{-1}Mpc, or below, even at the highest redshifts.

Figure 4 shows the resulting density field throughout the COSMOS cone out to a comoving distance of 2 400 h^{-1}Mpc (corresponding to $z = 1$), giving us a unique view of large-scale structure in the distant Universe. The density field has been normalised to the average density of galaxies to show the over-density δ, and the different cones in Figure 4 show iso-density surfaces for four different values of δ. Again a rich hierarchy of structure is revealed, showing variations on scales up to at least $\Delta z = 0.05$.

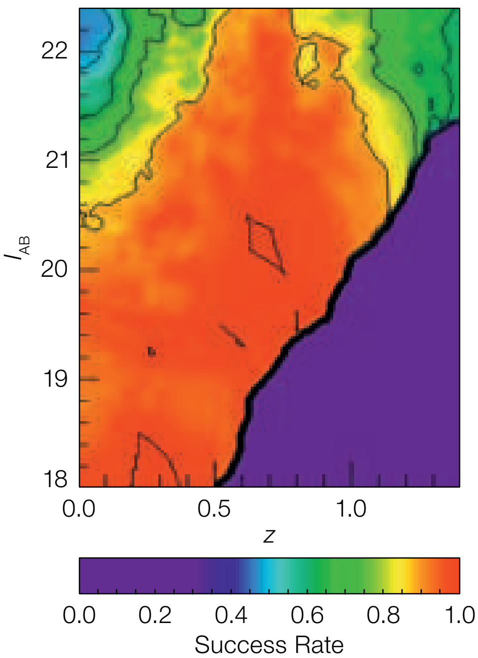
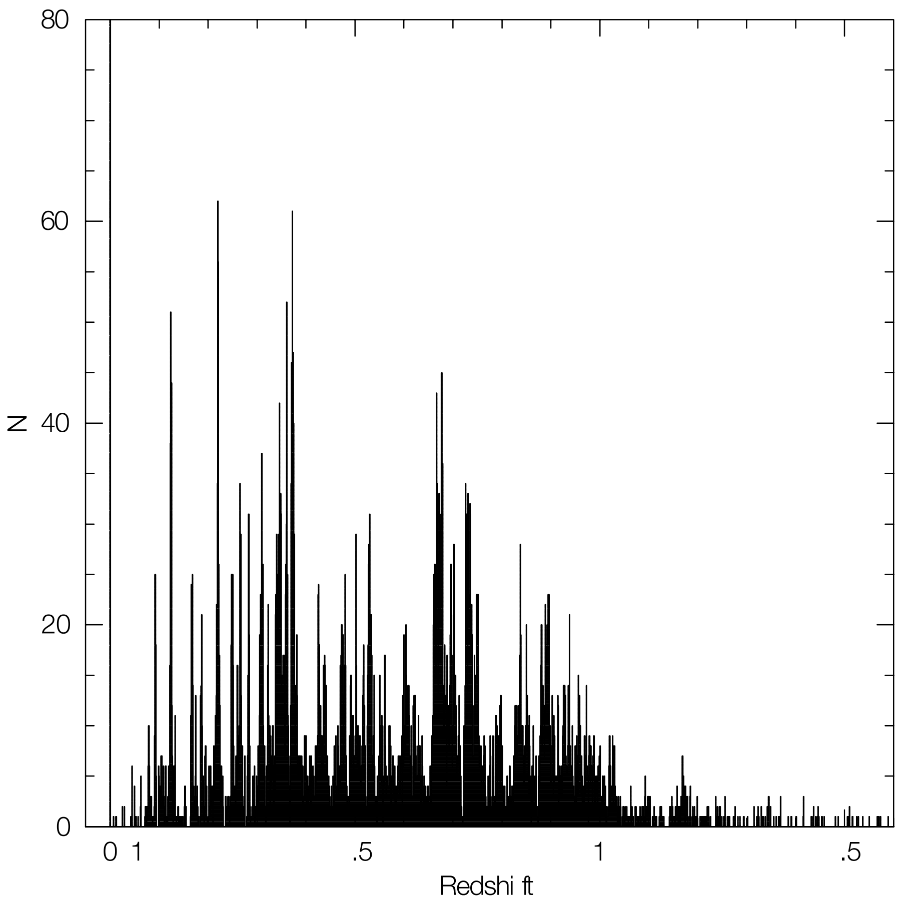


Figure 2. (Left) Fraction of spectra yielding a successful spectroscopic redshift measurement as a function of redshift (derived from the photo-z for the remainder) and I_{AB} magnitude. Notice how the success rate stays high in the key redshift range $0.5 < z < 0.8$ all the way to the limiting depth of the survey.

Figure 3. (Below) Redshift distribution of extragalactic objects in the zCOSMOS-bright 10k sample with secure redshifts, binned in intervals $\Delta z = 0.001$, which is larger than the redshift uncertainty by a factor of three at $z = 0$ and of two at $z \sim 1$. Despite the large transverse dimension of the survey, the redshift distribution shows structure on all scales from the velocity resolution up to $\Delta z \sim 0.05$.



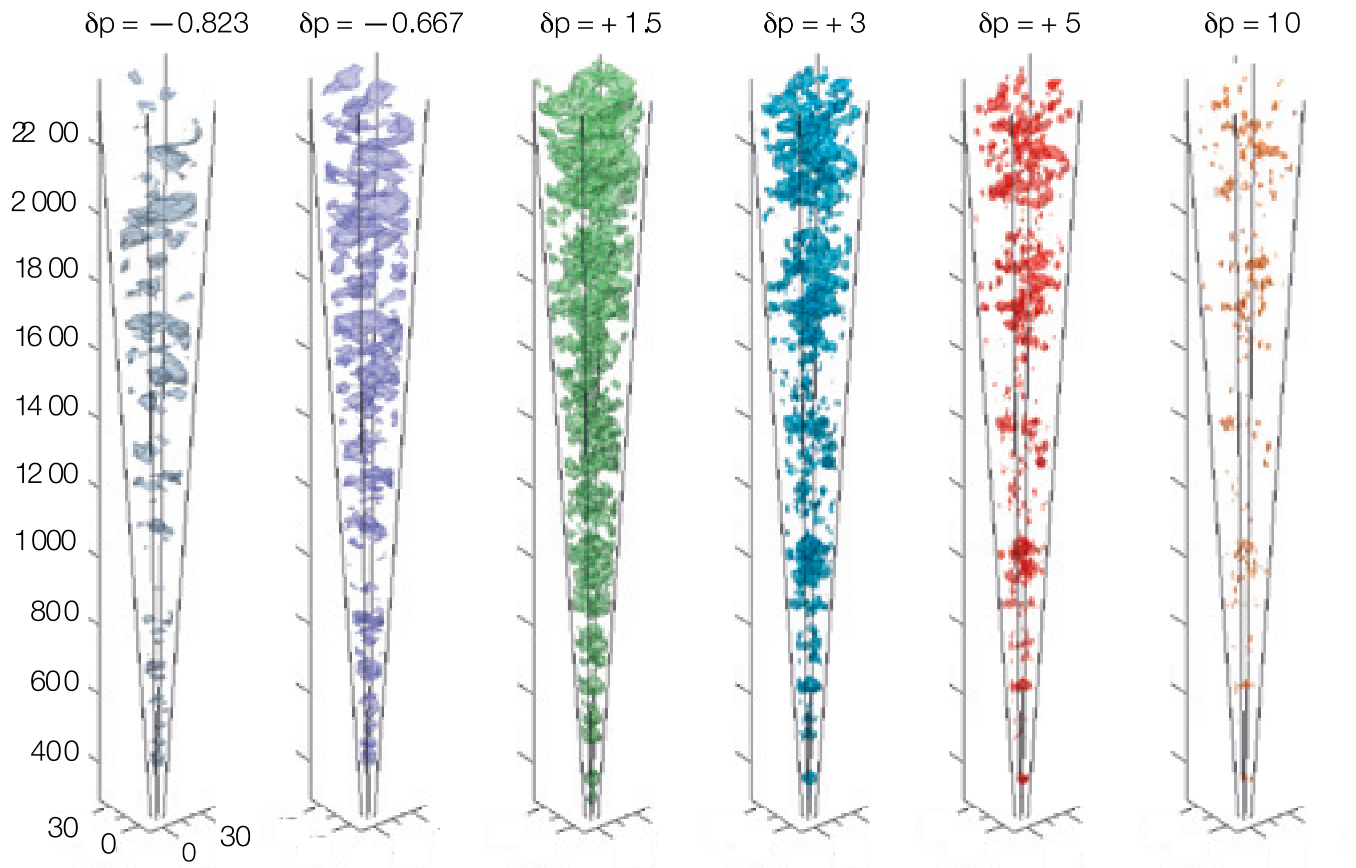


Figure 4. The galaxy density field reconstructed from the zCOSMOS 10k sample using the ZADE algorithm to include about 30 000 galaxies with photometrically estimated redshifts. The maximum comoving distance of 2400 Mpc corresponds to a redshift $z = 1$. The cones show, from left to right, iso-density surfaces corresponding to underdensities of δp = -0.823, δp = −0.667 and overdensities of δp = +1.5, 3, 5 and 10 respectively. Because the density field is locally projected to avoid the effects of peculiar motions, the equivalent physical overdensities are significantly higher, approximately $\delta \sim 3$, 7, 13 and 35 for the four rightmost cones.

There are many detailed choices for exactly how to construct a density field, and different ones may be best suited to some particular science applications. Accordingly, we have generated many such density reconstructions, each with different choices of tracer galaxies (flux- or volume-limited samples of galaxies), of smoothing kernels of different geometries (cylindrical or spherical) and scales (fixed size or adaptive), and of how the tracer galaxies are weighted (by straight number, by luminosity or by stellar mass). These are all described in detail in Kovac et al. (2008) and will soon be released at *http://www.exp-astro.ethz.ch/COSMOS.*

A large catalogue of galaxy groups to z ~ 1

The density field described above is inevitably on rather large scales, above one comoving Mpc. We may also be interested in the smaller-scale structure of individual galaxy groups — which we define to be galaxies moving within the gravitational potential well of a single virialised dark matter halo. zCOSMOS is well suited to this, having a relatively high sampling rate compared with other surveys (this will be especially true of the final 20k sample).

Numerous algorithms have been developed in the literature to identify galaxy groups in redshift surveys. We have extensively tested these against the zCOSMOS mock catalogues, which reproduce the complex selection function of the actual survey, and for which we know the host dark matter halo for each and every galaxy. By optimising against these very realistic mock catalogues, we have a very good idea of the statistical properties of our group catalogue, in terms of 'purity' and 'completeness' — the probabilities that our detected groups are real and that a given real group is detected, respectively — and the analogous 'interloper fraction' and 'completeness' for individual galaxy members of the groups. Compared with previous practices in the literature, we find that we can improve the fidelity of the group catalogue by introducing a multi-pass scheme in which we progressively alter the group-finding parameters to optimally find smaller and smaller groups, and by comparing and combining two

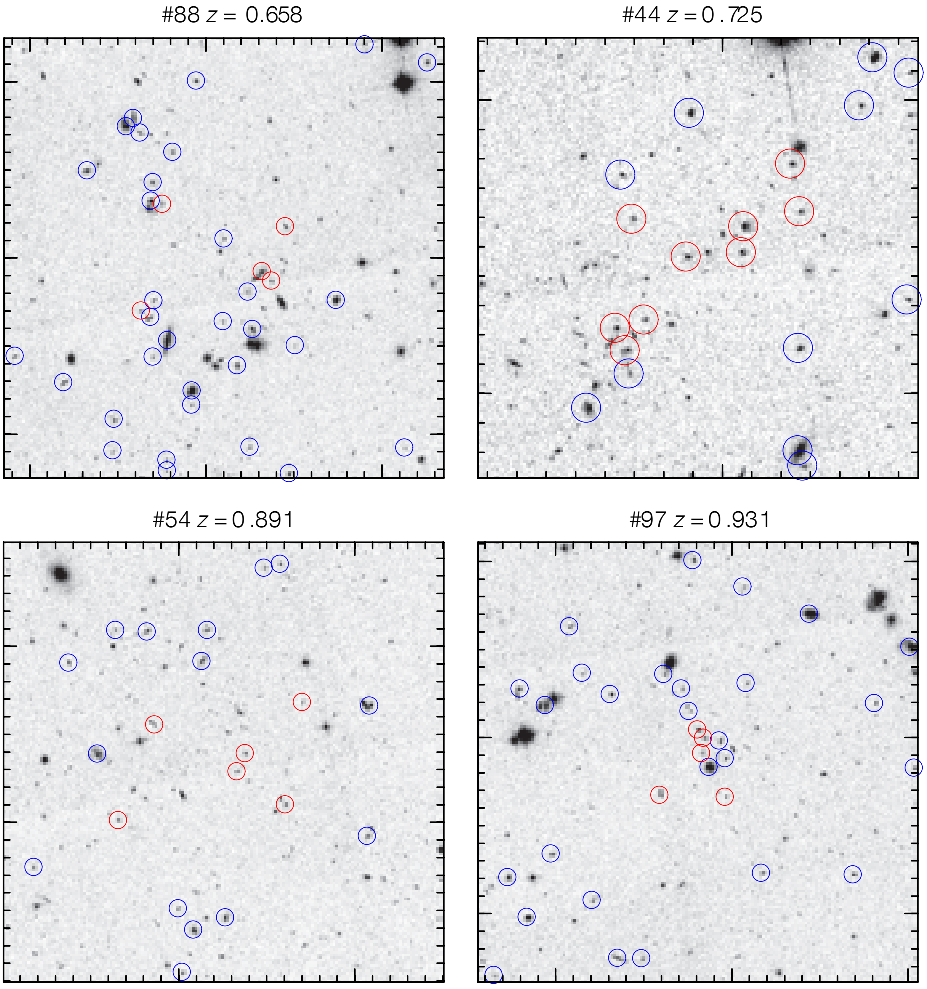


Figure 4. Four representative zCOSMOS groups at $0.6 < z < 1$. In the complete 10k sample there are 151 groups with four or more members and another 649 with two or three members. Group members are circled in red. Other galaxies with spectroscopic redshifts that are not in the group are circled in blue.

different approaches — the linking length based "friends-of-friends" method and a Voronoi–Delauney tessellation approach.

Already with the 10k sample, we have been able, with these improvements, to achieve, despite the currently quite inhomogeneous sampling, an impressively high fidelity in our group catalogue — significantly better than others in the literature at these redshifts. The group catalogue will continue to improve with the doubling of the number of spectroscopic redshifts that will be in the future 20k sample. Already we have identified 151 groups with four or more spectroscopically confirmed members and a further 649 groups with two or three member 'groups'. The zCOSMOS group catalogue is already one of the largest, and certainly the best defined, catalogues of galaxy groups at high redshift.

At low redshifts, $z \sim 0.3$, about a third of the galaxies in the 10k sample can be assigned to a group. This falls to about 15% at redshift $z \sim 0.8$, partly because the higher redshift galaxies are brighter, and therefore only intrinsically richer groups will be detected as a group, and also because there are fewer groups even at a fixed richness, because of the hierarchical growth of structure on these scales. Figure 5 shows a selection of a few of these groups at redshift $z > 0.5$. The zCOSMOS group catalogue is described in detail in Knobel et al. (2008) and will soon be released (see *http://www.exp-astro.ethz.ch/zCOSMOS*).

References

Ilbert, O. et al. 2008, arXiv:0809.2101, to appear in ApJ
Kitzbichler, M. G. & White, S. D. M. 2007, MNRAS, 376, 2
Knobel, C. et al. 2008, ApJ, submitted
Kovac, K. et al. 2008, ApJ, submitted
Lilly, S. J. et al. 2007, ApJS, 172, 70
Lilly, S. J. et al. 2008, ApJS, submitted
Scoville, N. Z. et al. 2007, ApS, 12,