DC4 / interface issues

E. Gaztanaga & W.Percival (for the LSS WG)*

April 23, 2009

Abstract

This document discusses a few of the issues to do with DES DC4 that have been arisen within the LSS working group. In terms of interface with the database, a SQL based front end, similar to CASJOBS for the SDSS would be good. LSS SQL queries could then be posted and exchanged on the wiki leaving a "paper-trail" for analyses.

1 Catalogue download

Early attempts to download the catalogue were hindered by several problems:

- lack of information (ie we need query examples that work).
- lack of options in the search and in the output
- having to download small areas at a time
- having to deal with many spurious objects and many parameters (artificially large a data set).

Current queries of the object catalog only allow selection based on (ra,dec) and do not give other options of which data to download. The output is large and has many artifacts which are not matched to any object in the input catalog. Ideally one would like to be able to query by magnitude, type or combinations, as in the SDSS SQL. We have had some trouble with some of the few functionalities but have worked this out directly by contacting the DESDM team. It seems that a good fraction of the objects in the catalog are on the boundaries of the CCDs, see Fig.1, which are probably caused by some boundary problem with Sextractor, or glowing edges that need to be masked.

After matching only true galaxies, we find that the resulting galaxy distribution clearly shows the structure of the co-add tiles (virtual squares patches of 0.75×0.75 deg used to co-add images). The structure appears because there is an overlap between tiles and there are repeated objects in the overlaps. There should be a flag in the DES-DM object catalog (eg primary object: yes/no) to be able to remove this repeats in an easy way (ie without need to match the catalog again). There is also a higher density of galaxies at ra$>$ 335 we do not understand right now (but could be related to the Pho-z catalog matching rather than the actual DC4 catalog). These can be seen in Fig.2. As expected (see Fig.4) this region contains fainter mean magnitudes. Contrary to expectations, this region also seems to contain galaxies with lower redshifts, (see bottom panel in the figure). The catalog only seems more or less homogeneous in density for $i < 20$.

*Key DC4 analysis performed by: Anna Cabré, Pablo Fosalba, Fernando de Simoni and Molly Swanson

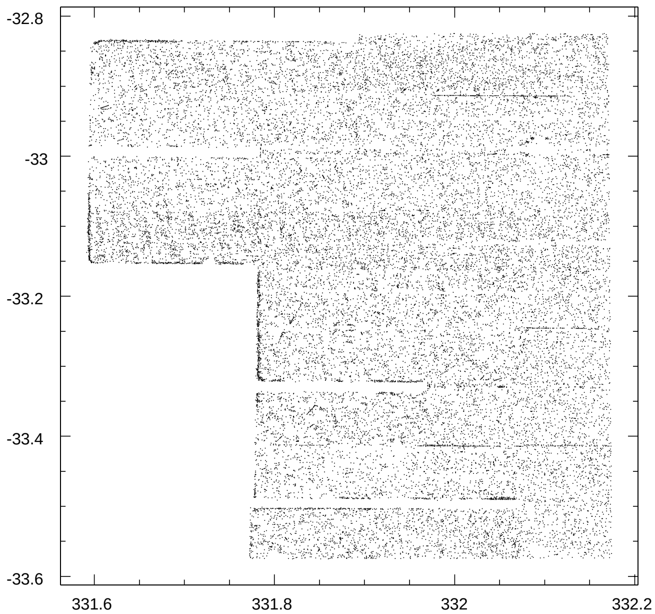


Figure 1: The distribution of objects on a piece of DC4. Edges of the CCDs can be clearly seem with more objects, indicating some problem with the boundaries.

2 Photo-z Catalogue

Estimated photometric redshifts were made available by Huan Lin on 7/4/2009. These were later made available through the DC4 portal, but we have not been able to check this directly. Only the maximum likelihood of the photometric redshift probability density function (PDF) was provided. We need the full PDF in order to optimally analyse the LSS.

The N(z) distribution is shown in Fig.3. There seems to be some major problem with the current photo-z values. The observed redshift distribution cannot be explained by cosmic variance.

3 Pointings & Mask

Previous lists of pointings sent to and used by the LSS group were based on the survey strategy plan. Actually, for the DC4 data, a different simple set of pointings was used, which only became apparent recently. These details (tiling pattern, chip positions, and tiling centres) should come from database queries. It would be useful to have these queries publicly available and executable, perhaps by means of an SQL query (see above) that can be checked and rerun.

A detailed explanation of some work by Molly Swanson on creating a mask using MANGLE (with help from Joe, Jim and DM team) can be found here:

`http://www.homepages.ucl.ac.uk/~ucapmes/des_mask.pdf`

Anna Cabre made an approximate mask based on the galaxy density HEALPIX, which is shown in Fig4.

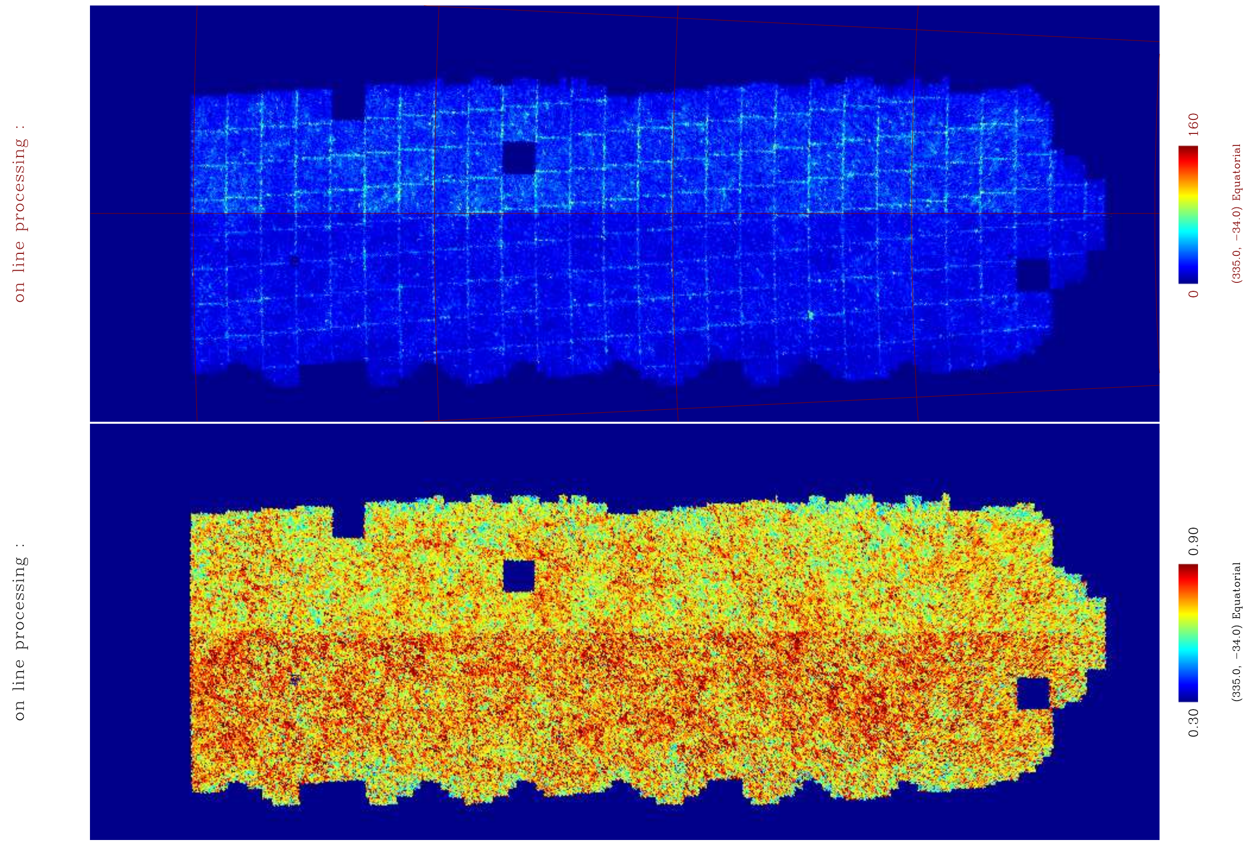


Figure 2: TOP: HEALPIX (N=2048, with1.7 arcmin pixels) galaxy density maps from co-adds in DC4. BOTTOM: Mean true redshift per pixel. Red corresponds to higher true redshift. Note how the lattice in the density map, apparent in top panel, has now disappear, indicating no correlation of redshift with the lattice. Also note, how the correlation of mean redshift with mean magnitude in Fig.4 is opposite to expectation.

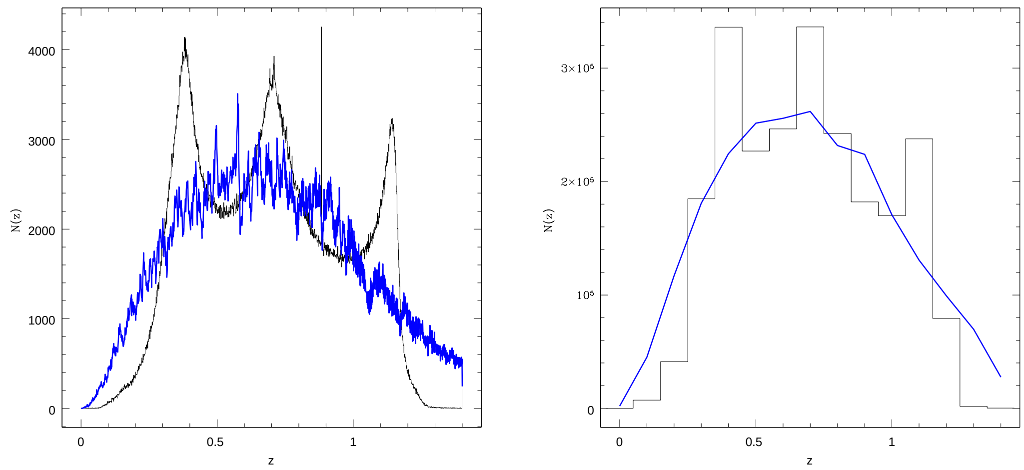


Figure 3: Number of galaxies N(z) as a function of true redshift (in blue) compare to N(z) as a function of photo-z (black histogram). There is some very obvious problem with the photo-z. In the right panel we use a larger bin of $\Delta z = 0.1$, comparable to the goal photo-z accuracy. The error in the distribution is less dramatic with such binning.

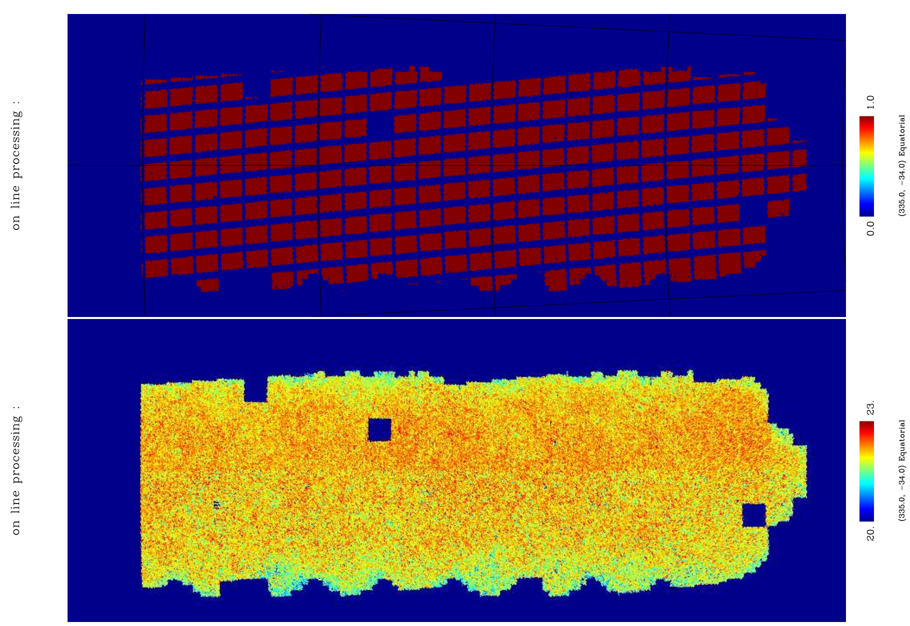


Figure 4: TOP: Approximate Healpix (N=2048, with1.7 arcmin pixels) angular mask based on galaxy density maps from co-adds in DC4. BOTTOM: Mean magnitudes per pixel. Red corresponds to fainter magnitude. Note how the lattice in the density map, apparent in Fig.2, has now disappear, indicating no correlation of magnitudes with the lattice.

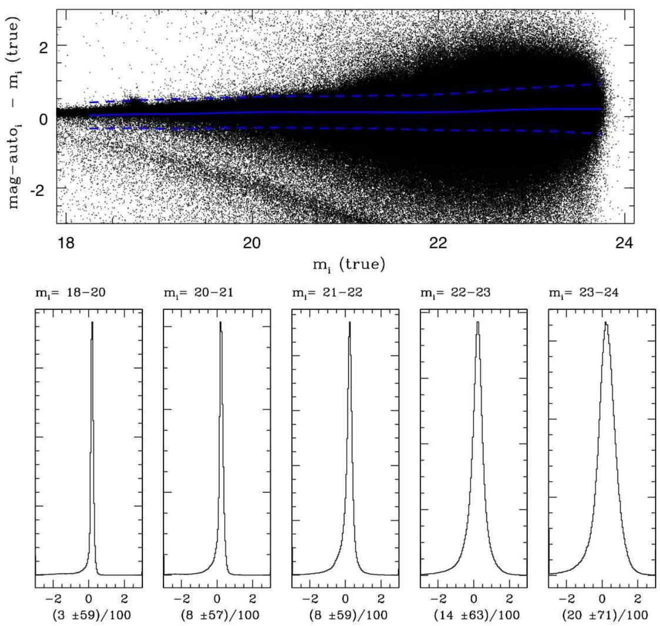


Figure 5: TOP: Comparison of magnitude differences Δm_i (measured auto-magnitude in the i-band minus true magnitudes in the input i-band galaxy catalogue). The blue continuous and dashed lines show the mean and rms in the scatter (for $|\Delta m_i| < 3$). The bottom panels shows histograms in different true i-band magnitude bins m_i, as labeled in each figure. Errors seem much larger than the nominal 10-sigma detection limits (ie 0.1 magnitude scatter to the depth of the survey?).

4 Magnitudes in the DC4 catalogue

Botton panel in Fig.4 shows the mean magnitude per pixel. The distribution of (measured-true) magnitude differences for galaxies has a larger dispersion than for stars (see Fig.5). It's not clear what is causing this. Most probably there is some problem in matching the true galaxies to objects in DC4. It seems that only 10% of the galaxies in DC4 are true galaxies and the rest is dominated by spurious faint objects that can, when match by proximity alone (as done in this case), be confused with real objects. Alternatives are that it is something to do with apertures used for the magnitude calculation, the estimate of the magnitude uncertainties, the light profile of galaxies in the simulations, or the simulation of the galaxy images.

5 Co-add Weights

The current strategy to stack images uses median combine. Will change to weighted mean (with cosmic ray masking) for later data challenges. This shouldn't make a big difference for current data. As part of the data analysis process, output weight maps are created (these give the inverse variance of the noise in each pixel). It would be good to be able to use these as the basis from which to calculate the mask, as we will then be sure that we match the co-add images. But the weight maps are not usefully output currently.

6 Zero points

These are stored in the header of each co-add image. They are not in the single images, which would be useful. Zero points per CCD and pointing have been obtained directly from the DESDM team. These were used with $1/\sigma^2$ weights to estimate the co-add zero point. Given the differences in approaches, we need to check that these magnitudes agree with galaxy magnitudes

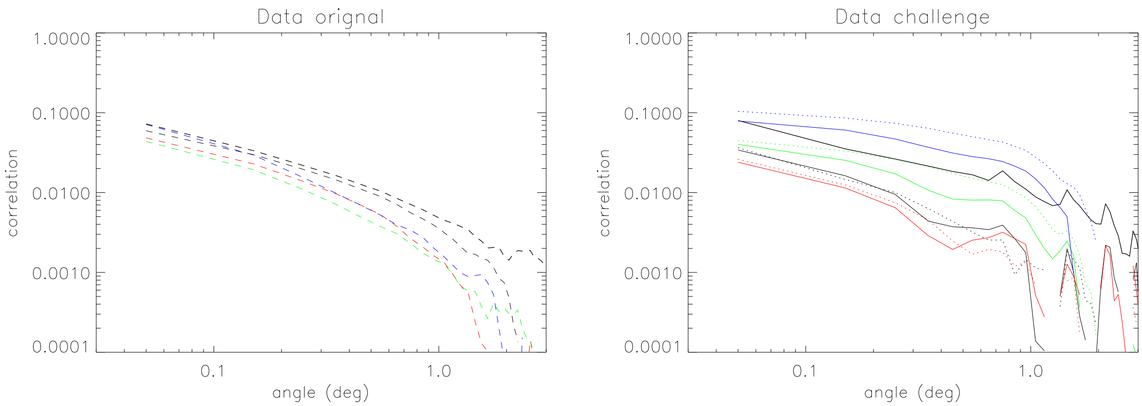


Figure 6: Comparison of the galaxy angular 2-point correlation function in the input catalog (left panel) and the DC4 data (right), using the mask in Fig.4. In both cases colors correspond to different (true) i-band magnitudes: i=23-24 (blue), 22-23 (green), 21-22 (red), 20-21 (black) and $i < 20$ (thick black). On the right panel, solid (dotted) lines correspond to $RA < 335$ ($RA > 335$).

in the database. The best procedure to get the final zero points is to work from the weight maps to ensure consistency between mask and galaxy magnitudes. These weight maps are a byproduct of the data analysis pipeline and contain the 1 sigma pixel noise and we know the ADU to magnitude conversion. The only issues left are the effects of stars, discontinuities in noise that are generic to point and shoot, etc

7 Clustering

Fig.6 shows the galaxy clustering in the full input catalog (on a 573 sqr deg area) as compared to galaxies in the DC4 catalog (a 75 sqr. deg subset) selected using the same (true) magnitude cuts in i-bands (different colors in the fiigure). Results are in reasonable agreement, specially in the light of the different artifacts that we found in the DC4 data. Note that in the input catalog the brighter magnitudes produce larger amplitudes of clustering. This is expected because brighter galaxies are closer to the observer and projection effects are smaller. This trend seems inverted in the DC4 data clustering, but it is not clear to us if this is due to sampling variance (the DC4 area is smaller and the brighter galaxies sample a smaller volume) or is related to the tendency that we found with mean redshift in Fig.2.

We have tried to address this question Fig.7 where we comapre both results. Ideally we would like to compare DC4 to the same region in the input catalog. But we do not know which part of the input catalog was used to make DC4. So what we show in the right panel of Fig.7 is a comparison of the full input catalog to the results in a subset which has identical area and mask to DC4 (but is probably not the DC4 region). As expected, there is some variance and a tendency to smaller correlation amplitudes for the smaller area (because of the integral constrain), but note that the order with magnitude slices is preserved. This indicates that the magnitude inversion that is apparent in the DC4 data is probably not due to sampling but rather to some problem with the measured magnitudes.

Fig.8 shows results for an complementary analysis of the same data. In this cases the mask uses all 3.4 arcmin pixels (n=1024) that contains galaxies. The angular clustering is computed

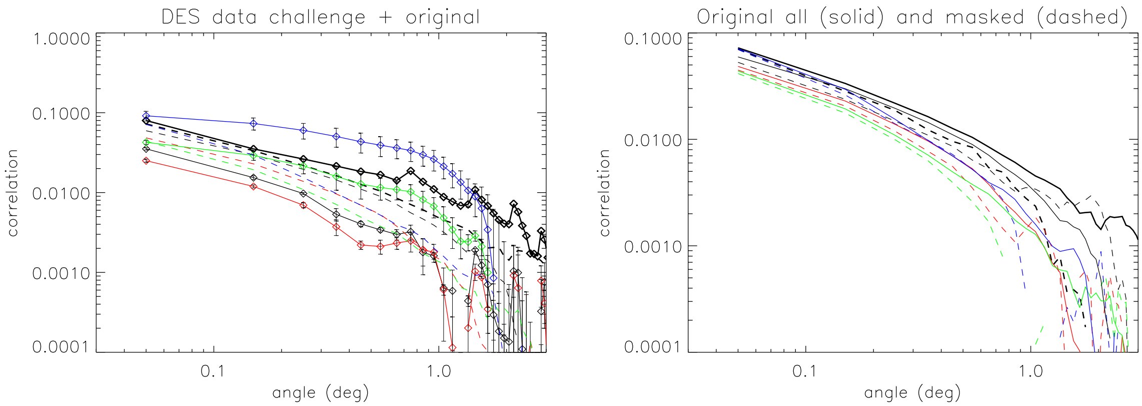


Figure 7: Left panel compares the clustering in the input catalog (dashed lines), ie same as left panel in Fig.6), to the mean correlation in the DC4 galaxies. Here symbols is the mean of $RA < 335$ and $RA > 335$ and errorbars correspond to the difference, indicating the level of sampling variance. Right panel shows the effect of sampling a smaller masked sub-area (dashed lines) within the input catalog (ie no DC4 objects) as compare to the full input galaxy catalog (solid lines), which is about 8 times larger.

using the Spice code and the results are similar to the ones in Fig.6.

The clustering amplitude in real and harmonic space can be roughly reproduced with linear theory (E&H power spectrum and cosmoloy as used by the input simulations) and a simple bias factor (b=2.3 for z 1, and b=1.2 for z 0.5), shown by dashed line in plots.

One thing that stands out is the fact that the z 0.5 sample shows an equality scale (through in correlation function) at the same angular scale than the z 1 sample, at about 3 deg, contrary to expectations which predict a shift to larger angles (4.5 deg)...this might be caused by the "step-in-density" artifact (or any other mask/systematic effect) that does not vary with z. Given that the mean redshift is anti-correlated with the magnitude in DC4 (ie Fig4), it is not surprising that the clustering does not quite follow expectations...

This is also seen in the harmonic Cls (right panel), where there is a peak in power at l 70 for the z 0.5 galaxies. The wiggles displayed by the DC4 angular power spectrum are probably not real, but they roughly follow those of linear theory (hardly visible given the width of the slices, see dashed line). The excess in amplitude at high l's should be due to non-linear effects.

Fig.9 shows the evolution of the slope in the correlation function (infer from the measured slope in the angular clustering) as a function of the true redshift as compared to the photo-z redshift. In order to avoid the border issue we have executed the analysis on the following angular limits: 334 to 338 degrees for the right ascension, and 42 to 25 for the declination. The area in this analysis is approximately 50 square degrees. For both photometric and real (truthtable) redshift, we have estimated the correlation length and the slope of the real space correlation function and its evolution with the redshift using the Limber approximation. The errors on the angular correlation function was done with the jackknife resampling method with 9 subareas.

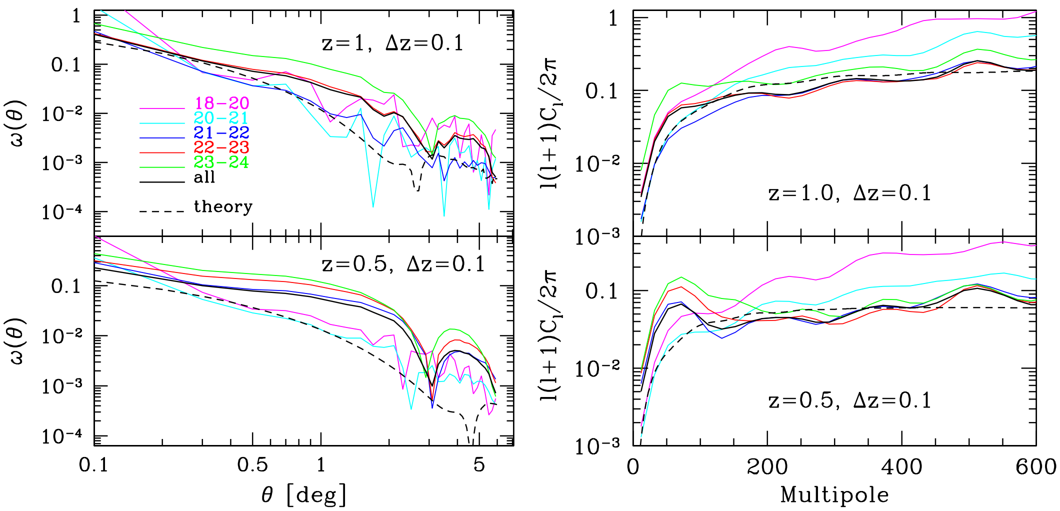


Figure 8: Left (right) panel compares the clustering in the the 2-point correlation (angular power spectrum) for two different redshift bins and different magnitude bins ("All" means $18 < i < 24$).

8 Future Plan

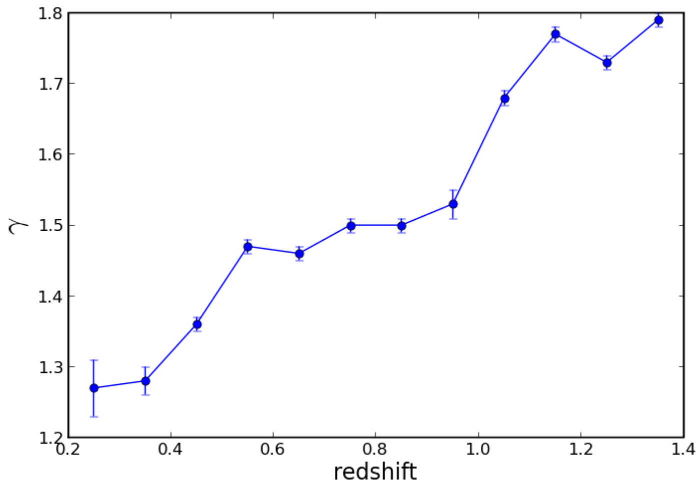
One of the activities within the LSS WG has been the "Value-added" catalog creation based on DC4. Details can be found under DES Brazil portal:

`http://testing.des-brazil.org/`

It is clear that we should have LSS working group code in the DESDM pipeline to help to or to determine the mask. This has been added as one of the goals of DC5. The exact mechanism for this is still uncertain. The aim will be to provide a procedure for quickly calculating magnitude limits for any ra,dec, matched to the magnitudes given in the catalogs. This might be fully integrated into the database and accessed using an SQL query, or via separate code using tables in the database. Options include:

1. A table containing magnitude limits for a list of regions and MANGLE parameters of the boundary polygons
2. A table of HEALPIX pixels for a pixelized version of the mask.
3. A SQL query that uses the weight maps and zero points stored in a table that would calculate magnitude limits "on-the-fly" for a given ra,dec, given this information.

There's a lot to be figured out, but there is time before DC5.



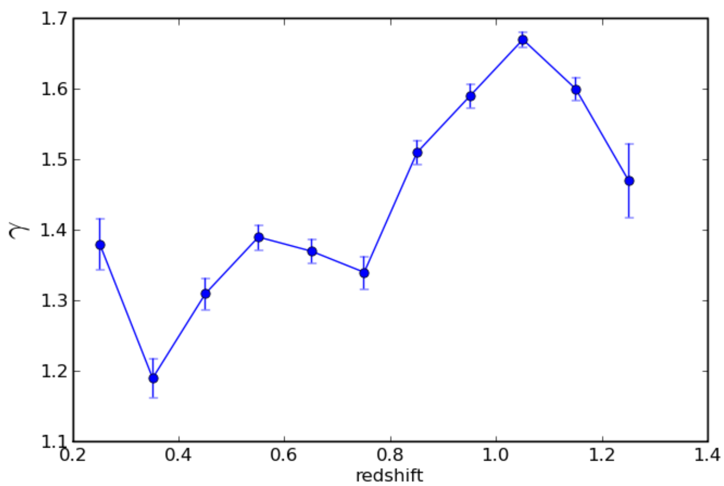


Figure 9: The slope in the 2-point correlation as a function of true redshift (left panel) as compare to the the photo-z (right panel).