

Galaxy Zoo: The Environments of the Peas

Abstract

1 figures

2 Introduction

This memo contains Figures / Tables made in the GZ investigation into the environment surrounding green pea galaxies. Along the way, a search is made to update our catalog of 'green peas' and investigate the initial search parameters characterizing the hunt for Peas.

Low mass starforming galaxies are thought to be the building blocks of galaxies, playing an important role in early galaxy assembly and evolution (?). 'Green Peas' or 'Peas' were galaxies first discovered in the Galaxy Zoo Survey, due to their small and green appearance in the SDSS images. Followup studies have shown them to be examples of relatively lower-mass, highly starforming galaxies, perhaps analogous to SF episodes occurring in the early universe. The question we wish to investigate is do the peas have a 'typical' environment

Encoded in the large scale structure of the Universe, is a variety of cosmological parameters as well as key information about the physical processes which underpin the formation of cosmic structures. Hierarchical structure formation, our current paradigm of galaxy formation, places them a key role on gravitational evolution of dark matter clustering around initial peaks providing potential wells for gas halos and galaxies to form in.

Cardamone et al 2009 leveraged a sample of 251 colour-selected galaxies in SDSS DR7 using a series of color cuts at $0.112 < z < 0.360$, where the OIII line falls in the r-band filter, to define the original sample of 'pea' galaxies. A final cut produced 80 confirmed peas after fitting the spectra for confirmation of star formation driving the emission.

At that time, using counts of projected neighbors to the original sample.

Now return using

The clustering of a set of galaxies can provide statistical measure of the average environment in which they live. As galaxies

We adopt a λ CDM cosmology with $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\lambda = 0.7$.

Figure 1 Figure 3 from Cardamone et al. 2009 (made by Stephen Bamford). The cumulative distribution of neighbour counts for the Peas (solid line) is significantly different than a comparison galaxy sample (dotted line). No Peas live in the highest density environments.

3 Data

3.1 Green Pea Galaxies

The original sample of 80 confirmed peas was derived from DR7 after fitting the spectra for BPT Star Forming Galaxies from a sample within the color space (above) and $0.112 < z < 0.360$, where the OIII line falls in the r-band filter. The main goal of Cardamoneetal2009 was to select a uniform sample of objects that could be investigated to uncover their properties. The redshift range was selected so that the OIII line fell within the r-band filter (providing the 'green' appearance in the SDSS images) and a color selection region was defined to avoid sequences of galaxies and QSOs. There is no uniform selection for the objects throughout the SDSS footprint, as we look for Peas only in the spectroscopic sample of objects and keep them only in objects with classified as Starforming with spectral line amplitude over noise (AON) ≥ 3 in all 4 lines in the BPT diagnostic. In this paper, our goal is to quantify the environments in which these galaxies live. In order to study this we take advantage of the huge expansion of SDSS spectroscopy through DR13, and use this to expand our sample of Peas.

Starting with the simple color-selection region within the redshifts $0.112 < z < 0.360$, DR13's PhotoObjAll catalog provides a sample of 4155 galaxies, which could be nominally labeled as 'pea' colored. This sample also includes the requirement's of a PSFMAG_R ≥ 10 and PETRORAD_R ≤ 5 . However, DR13's PhotoObjAll catalog contains numerous interloping objects (from artifacts, small regions of larger galaxies, and other objects with bad photometry). We wish to again select a catalog of Star Forming Peas, compact and with OIII emission powered by star formation. To recover this sample, we require spectral line measurements and therefor use the catalog provided by the Portsmouth Group in DR12, which uses the same programs we used for the original sample (Penalized PiXel Fitting (pPXF, Cappellari and Emsellem 2004) and Gas and Absorption Line Fitting code (GANDALF v1.5; Sarzi et al. 2006)) to derive emission line properties (?). Looking at images of the galaxies, shows many interloping objects which

do not physically appear as peas. Of the objects in the color-redshift space, 2203 have spectral fits in the Portsmouth catalog. Many have non-starforming BPT diagnostics, including 213 Seyfert galaxies, 65 LINERa, and another (52+103+87=242) composite spectra. Selecting for starforming objects only, leaves a sample of 1592 objects. Requiring clean photometry (SDSS flag CLEAN=1, removing objects with deblending issues, those on plate edges, etc.), leaves 1454 objects. Of the 138 objects with the Clean flag =0, there are perhaps a dozen that look like clean peas on the image. If we wish to limit the sample to those with individual $AON > 3$ in each of the special line measurements, only 739 objects remain. For the initial purposes of the clustering measurements - we'll use the full sample.

Figure 2 Phot Flag 0 sample of objects, removed from the spectroscopic sample even though classified as SF. Of the 138 objects with the Clean flag = 0, there are perhaps a dozen that look like clean peas on the image.

3.2 Luminous Red Galaxies

Luminous Red Galaxies represent a sample of passive, massive early-type galaxies (Eisenstein et al. 2001) extending over a redshift range $0.16 \leq z \leq 0.7$ in SDSS much further than the main galaxy sample (Strauss et al. 2002). Samples from SDSS have been used for a variety of studies from baryon acoustic oscillations (Kazin et al. 2010) to clustering (Padamanabhan et al. 2009) and redshift space distortion studies (Hikage and Yamamoto 2013). Classic samples of LRG were targeted as part of the SDSS-I and SDSS-II programs (Eisenstein et al. 2001) and from SDSS-III as part of the Baryon Oscillation Spectroscopic Survey program (Dawson et al. 2013). Luminous Red Galaxies are well suited for tracing dark matter distribution and a useful comparison for cross-correlations due to their intrinsic brightness and relatively high spacial density (Eisenstein et al. 2001; Wake et al. 2006).

The clustering properties of LRGs have been used to constrain cosmological parameters (e.g., Tegmark et al. 2004) as well as measuring baryon acoustic oscillation scales (Eisenstein et al. 2005, Anderson et al. 2012). Halo Occupation Distribution models have been shown to be a good representation for the autocorrelation function of the Luminous Red Galaxies measured from SDSS (Zehavi et al. 2005; Reid & Spergel 2009; White et al. 2011). This modeling work shows

that LRGs reside in massive halos with $M \sim 10^{13} h^{-1} M_{\text{sun}}$ and parent dark matter halos with masses of $M \sim 10^{15} h^{-1} M_{\text{sun}}$ can host several LRGs (Masaki et al. 2013).

For our analysis we use the BOSS DR12 Galaxy Catalog (Alam et al. 2015, Eisenstein et al. 2011) LOWZ sample of Luminous Red Galaxies (?) and the associated randoms catalog.

4 Clustering

For our new-expanded sample of Peas, we can compute a three-dimensional correlation function. However, even with our order of magnitude larger expanded sample of Peas, small number statistics limit the accuracy with which we can determine correlation functions, particularly in the case of autocorrelation functions. Cross-correlation functions can utilize higher space density samples of objects in similar sky and redshift space as our Peas. Luminous Red Galaxies low- z sample is good for this. Cross correlating samples of lower-density objects with dense galaxy samples tracing the large scale structure in the field can lead to lower statistical errors than measuring the autocorrelation function directly. Additionally, using cross-correlation functions does not require a complete understanding of the selection function of the galaxy sample of interest, which may not be well understood. INstead, we can leverage a well determined selection function for the galaxy tracer sample.

To interpret a correlation function, a large scale bias measure can be computed for linear growth and biasing schemes - this will only be accurate for correlation function measurements on larger scales ($\sim 1 - 2 h^{-1}$ Mpc). In order to more accurately explore the correlation function's measurements, a Halo Occupation Distribution framework can be used to constrain how various galaxy samples are distributed among dark matter halos - in particular explore both their presence as centers and/or satellite galaxies in the dark matter halos. To compute a correlation function, a random sample must be generated for comparison to each dataset, accounting for any variations in detection limits. To generate the random catalog for the Peas, we drew redshifts from the Pea-redshift distribution and randomized their location in RA and Dec by adding or subtracting up to a degree from their positions. ξ was determined using the pair counter from CorrFunc (Manodeep et al.) in bins of distance perpendicular to the line of sight r_p and parallel to the line of sight π . For catalog's like the peas, it is simple to use a KDE gaussian kernel to smooth their z distribution along the line of sight. However, it is far more

complicated in terms of RA/DEC position. Not all objects in the photometric catalog that fit a pea's color, shape, etc. received redshift measurements. In fact, most 'peas' selected received an extra fiber available in the field and were thus marked as a 'serendipitous source'.

4.1 Two Point Correlation Function ξ

The two point correlation function (ξ) quantifies the excess probability (relative to a Poisson distribution) of finding pair of galaxies at a given distance (r) from each other in a given volume element dV . Peebles 1980 expresses this given a mean number density of galaxies n as:

$$dP = n[1 + \xi(r)]dV \quad (1)$$

As the RA and Dec of the objects are known to a great degree of accuracy the angular two-point correlation function ($w(\theta)$) can be computed to determine the conditional probability of finding a galaxy within a solid angle $d\Omega$ at an angle θ from another galaxy. With the additional information provided by redshifts, real-space or 3D correlation functions can be computed. However, projected distances (in RA/Dec) have greater accuracy than those along the line of sight because the redshifts will be distorted due peculiar motions (Fisher 1994). Therefore the spatial correlation function is typically split into a projected co-moving dimension along the line of sight π and perpendicular to it r_p . The projected correlation ($w(p)$) function along a line of site can be determined to integrate over individual redshift-space distortions along the line of sight (Davis & Peebles 1983).

$$w_p(r_p) = 2 \int_0^{\pi_{max}} \xi(r_p, \pi) d\pi \quad (2)$$

The correlation function can be represented as a power law with the 3D correlation slope γ and correlation length r_0 .

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad (3)$$

And the projected correlation function can be represented as $w(r_p) = \Gamma(1/2)\Gamma[(\gamma-1)/2]/\Gamma(\gamma/2)r_0^\gamma r_p^{1-\gamma}$ (Peebles 1980). At larger scales, typically $r_p \geq 1-2$ Mpc/h, the amplitude of the correlation function ($w_p(r_p)$) is due to the correlation between objects in distinct halos. This linear regime is well modeled by the power law fit

and the bias parameter represents the relationship between the large-scale clustering amplitude of the correlation function of the data and the 2-halo term for the dark matter distribution.

$$b_{2h} = \sqrt{\frac{w_{p,2h}(r_p)}{w_{DM,2h}(r_p)}} \quad (4)$$

Computing this requires determining a dark matter projected correlation function.

The bias can be related to the halo mass (van den Bosch 2002) Sheth et al. 2001 through a bias-mass relation $b(M_h, z)$. The clustering and number density of a given halo is determined by its mass. Analytically this is characterized by the amplitude of density fluctuations from which a halo of mass M_h forms at a given redshift ($\nu = \delta_c/\sigma(M_h, z)$). The halo bias can then be modeled a function of halo masses or a fixed redshift (see e.g., (?), (?)).

The mass of a halo is related to it's clustering and the number of hals <http://adsabs.harvard.edu/abs/2002>

The two-point correlation function is most determined using a comparison to a random sample (pure noise) to see the strength of the signal at any given distance. The Landy-Szalay estimator is commonly used today, as it has been shown to provide the smallest statistical variance (Fisher 1994) and to be more accurate at larger scales (PonsBorderia et al. 1999, Kerscher et al. 2000). Given the Peas sample (D_1), a comparison galaxy sample (D_2) and matched samples randomized in locations (R_1, R_2) the Landy-Szalay estimator is:

$$\xi(r) = \frac{D_1 D_2(r) - D_2 R_2(r) - R_1 D_2 + R_1 R_2(r)}{R_1 R_2(r)} \quad (5)$$

In this fuction, each pair corresponds to pairs of objects with a given separation (e.g., $D_1 D_2(r)$ is the number of pairs between the galaxy and pea catalogs with a distance r). However, this measurement of the correlation function requires the construction of an accurate estimate of the distribution function to create matched random catalog for each data set, reproducing the data's selection function and observational limitations. The estimate of $\xi(r)$ is strongly dependent on the accuracy from which selection effects are reproduced including flux limits and optical followup limitations (the mask effect). Given the non-uniform selection of the Peas, it is not possible to estimate an accurate detection map projected on the sky from which the random distribution can be drawn. Lacking the necessary projected detection distribution for deriving an accurate randoms catalog, Padamanabhan et al. 2009, used a the Peebles estimator estimator for a quasar function:

$$w_\theta(R) = \frac{D_1 D_2}{D_1 R_2} - 1 \quad (6)$$

(Davis and Peebles 1983). An accurate estimate of the distribution of random samples is necessary to compute a correlation function.

However, a better interpretation fits an empirical model relating the galaxies to their dark matter halo with a statistical model of the clustering of galaxies within a given dark matter halo. This transforms pair counts of galaxies into a more physically informative relationship between the galaxies and the dark matter halo in which they reside.

4.2 Halo Model

To interpret large-scale clustering data, the relationship between the galaxies themselves and their dark matter halo hosts must be modeled. The Halo Occupation Distribution is an empirical model that associates the galaxies in the data with dark matter halos as a function of the mass of the dark matter halos (Peacock & Smith 2000). This approach first models a distribution of halos from a given cosmological model and then populates the galaxies of interest into the dark matter halos resulting in a probability that a halo of mass M_h contains N objects of a particular type - $P(N|M_h)$ the halo occupation distribution function. To determine the HOD we can model the two-point projected correlation function as a sum of two terms, a 1-halo term (1h) where both objects occupy the same dark matter halo and a 2-halo term (2h), where each pair of objects occupy a different dark matter halo.

$$w_p(r_p) = w_{p,1h}(r_p) + w_{p,2h}(r_p) \quad (7)$$

Applying a predefined functional form to this relationship, allows To fit a measured correlation function, we can parameterize the Halo Occupation Distribution (as a sum of the number of central and satellite galaxies) and adopt a cosmology. Zheng et al. (2009) modeled a cross correlation function between luminous red galaxies and a set of normal L_* galaxies. The mean function of of a halo of mass M , ($\langle N \rangle (M)$), is typically a power law distribution with a cutoff mass M .

5 Analysis

We use `corrfunc` (?) to estimate the correlation function.

A reliable estimate of $w(r_p)$ depends on setting π_{max} so that it's not too large, adding noise to the estimate, or too small, failing to recover the signal. We computed the cross correlation function for a variety of π_{max} values and adopted 60 Mpc/h as the best compromise.

bootstrap resampling was used for uncertainties in the correlation function (norberg)

6 Analysis

Using the sample of the Peas taken from DR12, there aren't ample sources available for an autocorrelation, but we calculate one anyways to see what we would get.

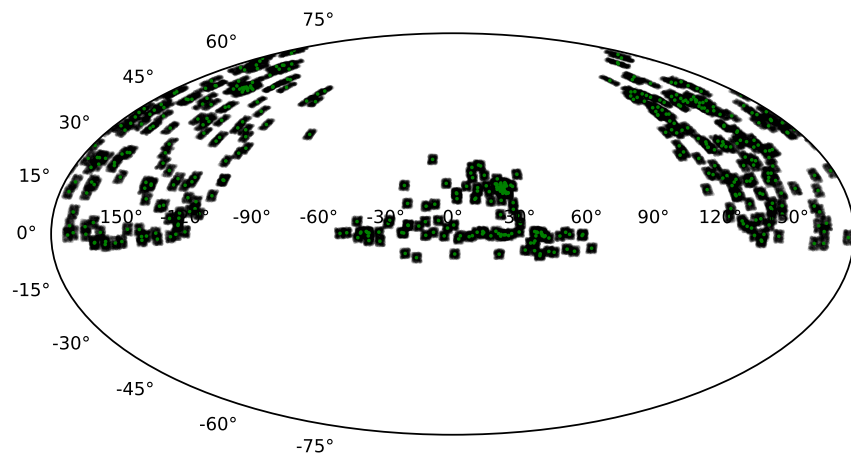


Figure 3 Looking into redshift distribution of new Peas from DR12. There are perhaps relatively more at lower $z < 0.2$

7 The original 80

Cardamone et al 2009 leveraged a sample of 251 colour-selected galaxies in SDSS DR7 between $0.112 < z < 0.360$ using a series of color cuts to define the original ‘pea’ galaxies.

$$u - r \leq 2.5 \quad (8)$$

$$r - i \leq -0.2 \quad (9)$$

$$r - z \leq 0.5 \quad (10)$$

$$g - r \geq r - i + 0.5 \quad (11)$$

$$u - r \geq 2.5(r - z) \quad (12)$$

Using the existing spectra, the sample was further refined to 80 objects that met a specific requirement of AON greater than 3 in each of the 4 line diagnostics in BPT diagram and falling into the star forming spectral line ratios. This came from a sample of 103 narrow line objects with $S/N > 3$ in each of the lines of the BPT diagram. One of the objects was a Broad-line AGN and 8 were Narrow Line Seyfert 1s, meaning that 112 out of the 251 galaxies had spectra that were classified in that original work. This limited sample was followed up on in the paper as it’s characteristics (including masses, metallicities and star formation rates were determined). With that original sample from DR7, an attempt to estimate their environment was made selecting a sample of random galaxies matched in luminosity and redshift and comparing projected neighbors within a radius of 1Mpc brighter than an iBand magnitude of -20.5 (At the redshift of the target). This initial test showed that the Peas inhabit significantly lower density regions than typical galaxies at the same i-band magnitude and a median environmental density surrounding them was less than 2/3rds that of normal galaxies. In this memo, we will refer to this final sample of 80 object as the original pea galaxies.

8 environments of the original peas

8.1 Searching for Neighbors

At $z = 0.275$, the median redshift of the original 80 peas, the angular scale distance is 871.7 Mpc, providing a scale of 4.226 kpc/” (Wright (2006, PASP, 118, 1711). Therefore, a galaxy of size 10 kpc would extend 2.37 arcseconds on an SDSS image at this redshift, and a group scale of 1 Mpc would extend 236” on

the sky, and a large cluster on the scale of 10 Mpc would be 2366'' or nearly 40' on the sky in diameter.

$$1'' \times 0.004226 \text{ Mpc} / '' \quad (13)$$

We started a simple search near the positions of each of the original 80 peas in the SDSS DR13 catalogs. Searching within ± 1 Mpc in RA and DEC of the 80 pea's positions, the DR 13 photoobj all catalog had 107298 objects and the spectroscopic catalog had 38854 objects (at any redshift!). All of the original 80 objects are returned in this search of the DR 13 catalog, although one near a saturated bright star has slightly different photometry. Because the main photometric catalog contains many objects, most of which will turn out to be at other redshifts, we search the photometric redshift catalog for each of these objects. It's also worth noting that fibre collisions on the same SDSS plate prevented fibers from being placed closer than 55 arcsec and can effect the density of objects with redshift measurements in any field. The photometric redshifts are calculated for all objects in the primary catalog tagged as galaxies using an empirical training set with a kd-tree nearest neighbor fit (Beck et al. in prep). There are 18,505 objects with photometric redshifts or 17% of the catalog. However, of those objects only 4125 have reliable photometric redshifts with calculated uncertainties (less than 4% of the sample).

Starting with this set of objects, we find the neighbors for each pea within a projected distance of 236 arcseconds on the image. The number of neighbors for these peas range from just over 200 to 4000. For the two peas in Stripe 82, with the deeper imaging, the neighbors are greater (8634 and 14742). However, this suggests that these peas can be found in a variety of environments.

What would be most helpful, would be to select for objects near the redshift of the pea. But numbers vary with depth of image (not necessarily the physical space density, as can be seen with the Stripe 82 data).

Figure 4 Looking into redshift distribution of new Peas from DR12. There are perhaps relatively more at lower $z < 0.2$

The original sample of 80 confirmed peas was derived from DR7 after fitting the spectra for BPT Star Forming Galaxies from a sample within the color space (above) and $0.112 < z < 0.360$, where the OIII line falls in the r-band filter. The main goal of Cardamone et al 2009 was to select a uniform sample of objects that could be investigated to uncover their properties. The redshift range was selected

so that the OIII line fell within the r-band filter (providing the 'green' appearance in the SDSS images) and a color selection region was defined to avoid sequences of galaxies and QSOs. There is no uniform selection for the objects throughout the SDSS footprint, as we look for Peas only in the spectroscopic sample of objects and keep them only in objects with classified as Starforming with spectral line amplitude over noise ($AON \geq 3$ in all 4 lines in the BPT diagnostic. In this paper, our goal is to quantify the environments in which these galaxies live. In order to study this we take advantage of the huge expansion of SDSS spectroscopy through DR13, and use this to expand our sample of Peas.

Starting with the simple color-selection region within the redshifts $0.112 < z < 0.360$, DR13's PhotoObjAll catalog provides a sample of 4155 galaxies, which could be nominally labeled as 'pea' colored. This sample also includes the requirement's of a $PSFMAG_R \geq 10$ and $PETRO_R \leq 5$. However, DR13's PhotoObjAll catalog contains numerous interloping objects (from artifacts, small regions of larger galaxies, and other objects with bad photometry). We wish to again select a catalog of Star Forming Peas, compact and with OIII emission powered by star formation. To recover this sample, we require spectral line measurements and therefor use the catalog provided by the Portsmouth Group in DR12, which uses the same programs we used for the original sample (Penalized PiXel Fitting (pPXF, Cappellari and Emsellem 2004) and Gas and Absorption Line Fitting code (GANDALF v1.5; Sarzi et al. 2006)) to derive emission line properties (Thomas et al. (2013)).

Looking at images of the galaxies, shows many interloping objects which do not physically appear as peas. Of the objects in the color-redshift space, 2203 have spectral fits in the Portsmouth catalog. Many have non-starforming BPT diagnostics, including 213 Seyfert galaxies, 65 LINERa, and another $(52+103+87=242)$ composite spectra. Selecting for starforming objects only, leaves a sample of 1592 objects. Requiring clean photometry (SDSS flag CLEAN=1, removing objects with deblending issues, those on plate edges, etc.), leaves 1454 objects. Of the 138 objects with the Clean flag =0, there are perhaps a dozen that look like clean peas on the image. If we wish to limit the sample to those with individual $AON > 3$ in each of the special line measurements, only 739 objects remain. For the initial purposes of the clustering measurements - we'll use the full sample.

Figure 5 Phot Flag 0 sample of objects, removed from the spectroscopic sample even-though classified as SF. Of the 138 objects with the Clean flag ne 1, there are perhaps a dozen that look like clean peas on the image.