Scales of Dark Matter Clustering in High Redshift Galaxies

Scott Underriner

Abstract

I analyzed images of high redshift galaxy clusters in order to determine the distribution of redshifts in them so that a more accurate cluster mass measurement could be made. This information can be used in other weak gravitational lensing research to improve the accuracy of their results.

Introduction

Statement of Purpose

Gravitational lensing is the process by which the dark matter is detected and measured indirectly by observing its gravitational effect on incident light. The degree to which this light is affected is proportional to the amount of matter in the gravitational lens. By measuring the amount of bending, or the shear, we are able to measure the mass distribution of the lens. Lensing research uses a reference field for the redshift distribution and assumes a redshift distribution based on that reference field. This assumption leads to an increase in error in the results. The biggest source of error due to this assumption is the inclusion of foreground galaxies in the matter distribution of the gravitational lens. Galaxies in the image that are between the observer and the lens are not affected by the lens, but because of the assumption about the objects' redshift they are treated as being part of the source so that their light is not bent by the lens but their mass contributes to the matter distribution of the background or lens.

This research is therefore concerned with developing a method that accurately measures the redshift distribution of the source, lens and foreground objects so that the mass measurements are more accurate and can be normalized, and the magnitude of the error due to the foreground galaxies can be ascertained.

Background Information

In talking about the search for dark matter, it is important to discuss what dark matter is. The exact nature of dark matter is unknown, but that fact that it constitutes ~26% of the mass-energy of universe means that it plays an important role in the interactions of the cosmos. Dark matter does not interact electromagnetically. This means that it does not absorb or emit electromagnetic radiation, making it impossible to see by typical methods of observation such as infrared or visible astronomy. Dark matter is known to interact gravitationally, however. Because dark matter interacts gravitationally but not electromagnetically, it has the tendency to clump together into clusters. Baryonic matter has the ability to radiate, creating a pressure that acts against the attractive force of gravity. Without this force to counterbalance gravitational attraction, dark matter must contract and clump together. These clusters cause large gravitational effects, allowing the dark matter to be indirectly observed. It is by its gravitational effects that it was first observed, and how we measure its quantity and distribution today.

One method used to observe dark matter is gravitational lensing. According to Einstein's theory of relativity, mass bends the space-time around it, causing the paths of objects to be changed. Thus, light that passes by a large gravitational source will have its path affected. An example of this can be seen by examining the change in the line element, the separation between two points in Minkowski space-time. Without a massive object, the line element in natural units is:

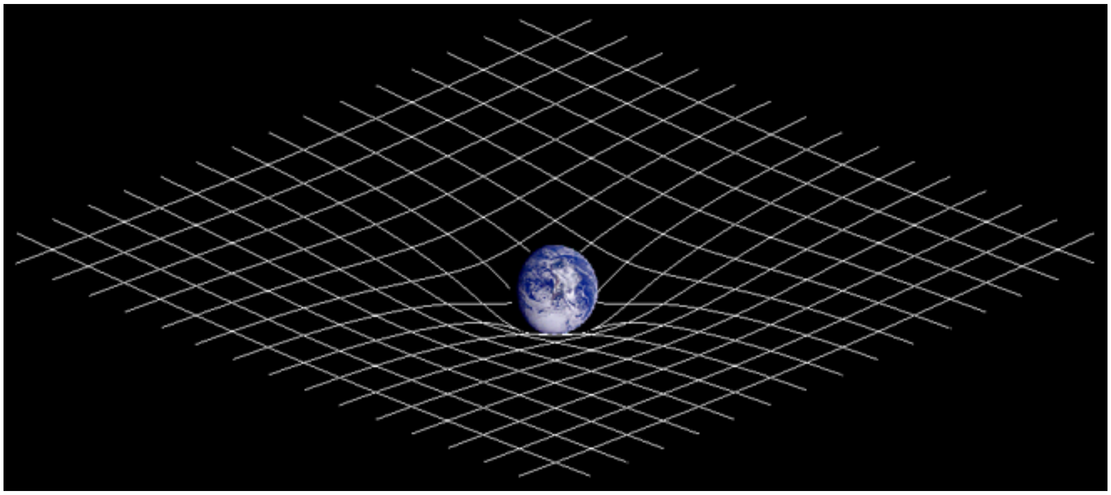
$$ds^2 = dt^2 - dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2$$

With the introduction of a massive object, the line element changes to the Schwarzschild metric:

$$ds^2 = \left(1 - \frac{r_s}{r}\right) dt^2 - \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta \, d\phi^2$$

The Schwarzschild metric describes the gravitational field around a spherical object with a few assumptions: the electric charge of the sphere, its angular momentum, and the cosmological constant are all zero.

In both equations, *ds* is the differential line element, *dt* the differential in time, *dr* is the radial differential, *dθ* the polar angle differential, *dϕ* the azimuthal angle differential, and r_s the Schwarzschild radius, which is equal to *2GM*. *G* is Newton's gravitational constant and *M* is the mass of the body or lens.



An image illustrating the effect a massive body has on the space-time around it Source: http://upload.wikimedia.org/wikipedia/commons/2/22/Spacetime_curvature.png

The angle of the deflection of the incident light caused by the gravitational lens directly corresponds to the mass of the object, reflected in the equation (*c* is the speed of light in a vacuum):

$$\alpha = \frac{4GM}{rc^2}$$

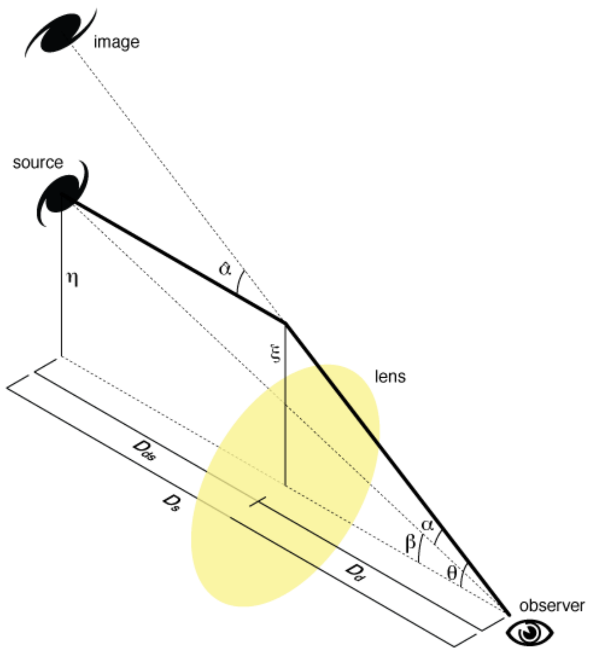
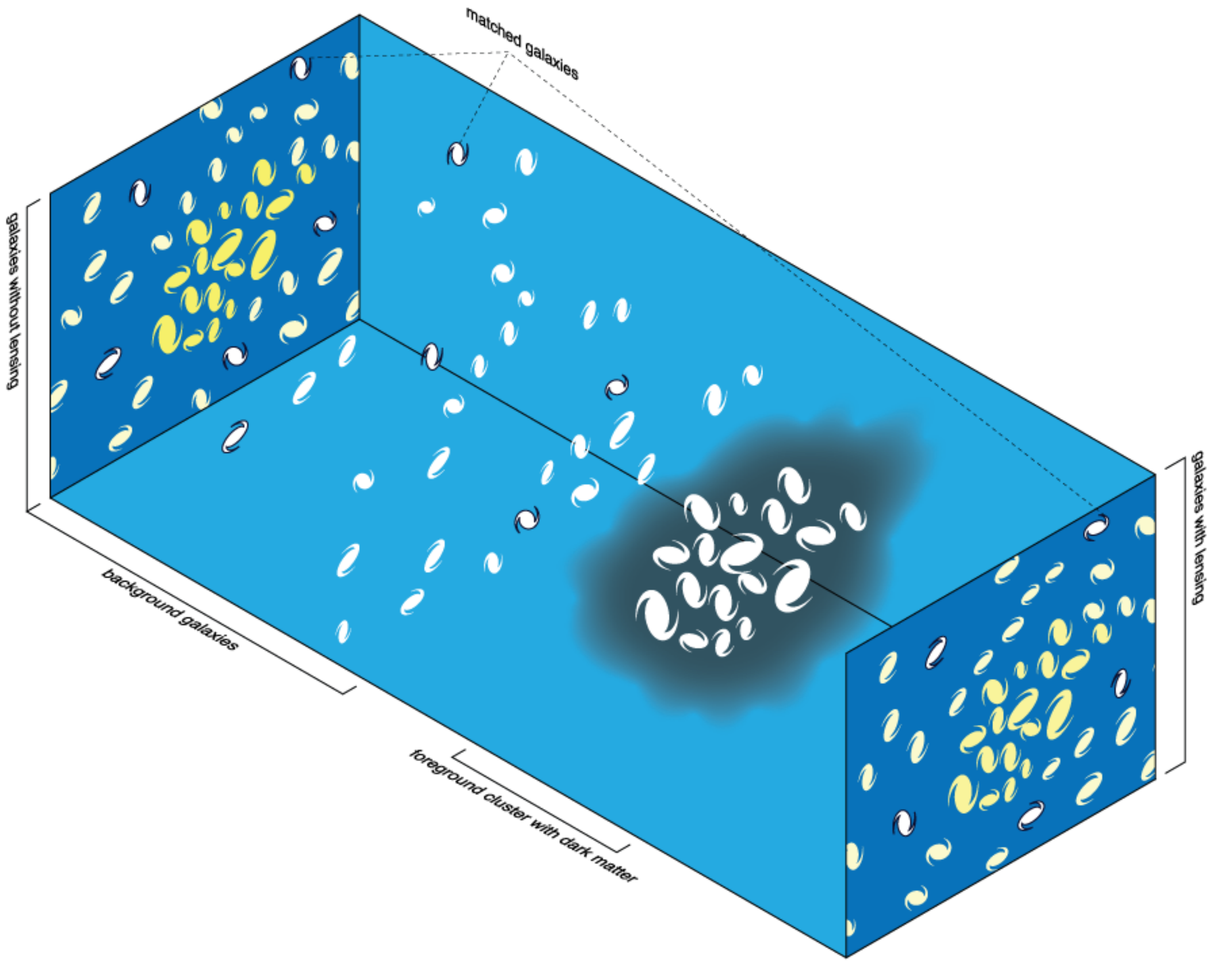


Image to the left side shows depiction of gravitational lensing and corresponding angular values. Source: http://upload.wikimedia.org/wikipedia/commons/e/e4/Gravitational-lensing-angles.png

There are different types of gravitational lensing, depending on the magnitude of the lensing, and thus the amount of matter in the lens. Strong gravitational lensing occurs when gravitational lenses bend light to create entirely new images, like Einstein rings, that are similar to mirages. Much of the work done with gravitational lensing, including the work of Professor Dell'Antonio's group, is done involving weak gravitational lensing. Weak gravitational lensing is involved in lenses of a smaller mass density than strong lenses. The resulting image distortion is therefore on a much smaller scale. Instead of the creation of entirely new images, the objects in the image are stretched and distorted slightly, with a statistical preference of direction perpendicular towards the center of mass of the lens.



An example of the effects of a weak gravitational lens. The image shows the background galaxies, the lens, and the image of the background galaxies after having their light affected by the lens. Source: http://upload.wikimedia.org/wikipedia/commons/b/b9/Gravitational-lensing-3d.png

Weak lensing occurs at the ends of a galaxy cluster, and strong lensing occurs at the cluster's center, where the matter distribution is denser.

To measure the gravitational effect on the lens on the background galaxies, the inherent ellipticities of the background galaxies must be measured so as to construct an estimate of the their alignment. The alignment of the galaxies should be random, so any systematic alignment between multiple galaxies can be assumed to be caused by the gravitational effect of the lens. In order to get a statistically sound result, many background galaxies must be measured. This preferential stretching effect is known as the objects' shear. With a known shear, the mass distribution of the lens can be constructed. Knowing the entire mass of the lens, baryonic matter and dark matter, the dark matter can be found by measuring the baryonic matter and subtracting that from the total matter measured by the shear. The remaining matter must therefore be dark matter.

An important phenomenon in determining the distance to the galaxies is the photometric redshift. A redshift is the increase in wavelength in the spectrum of radiation, and is given by the equation (where z is the redshift and λ is the wavelength):

$$z = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}}$$

Similar to the Doppler Effect, an astronomical redshift is not caused by the movement of the observer or source but by the expansion of space between the observer and the source. The universe is constantly expanding, which means that every object on a large enough scale (typically that of galaxy clusters) is redshifted to some degree with respect to us. According to Hubble's law:

$$v = H_0 D$$

Hubble's law says that the recessional velocity v of an object due to the expansion of space is equal to the Hubble constant H_0 times the distance D away from the observer. With Hubble's law, we see that the distance between us and an object is directly related to the recessional velocity, which can be determined with the redshift with the equation

$$z = \frac{v}{c}$$

Thus, redshift can be used as an indirect measurement of the distance.

In this research, the photometric redshift is used as the distance measurement. The photometric redshift uses the objects' photometry – the brightness of the object through various filters. The objects' fluxes are measured in filters such as the red light filter and other color filters, the ultra violet filter, and the infrared filter. By taking measurements of multiple filters, the readings of the objects' radiation are more accurate, and thus when observing the redshifting of this light the corresponding distance measurement is more accurate. In order to obtain the probability distribution of an object's redshift, its spectroscopic redshift is needed as well. The spectroscopic redshift refers to a change in the absorption and emission lines of electromagnetic radiation. According to quantum mechanics, every element and molecule has specified and quantized energy levels, permitting the absorption and emission of only particular amounts of energy. The energy of the emission is related to its wavelength by the formula (*h* is Planck's constant):

$$E = hf = \frac{hc}{\lambda}$$

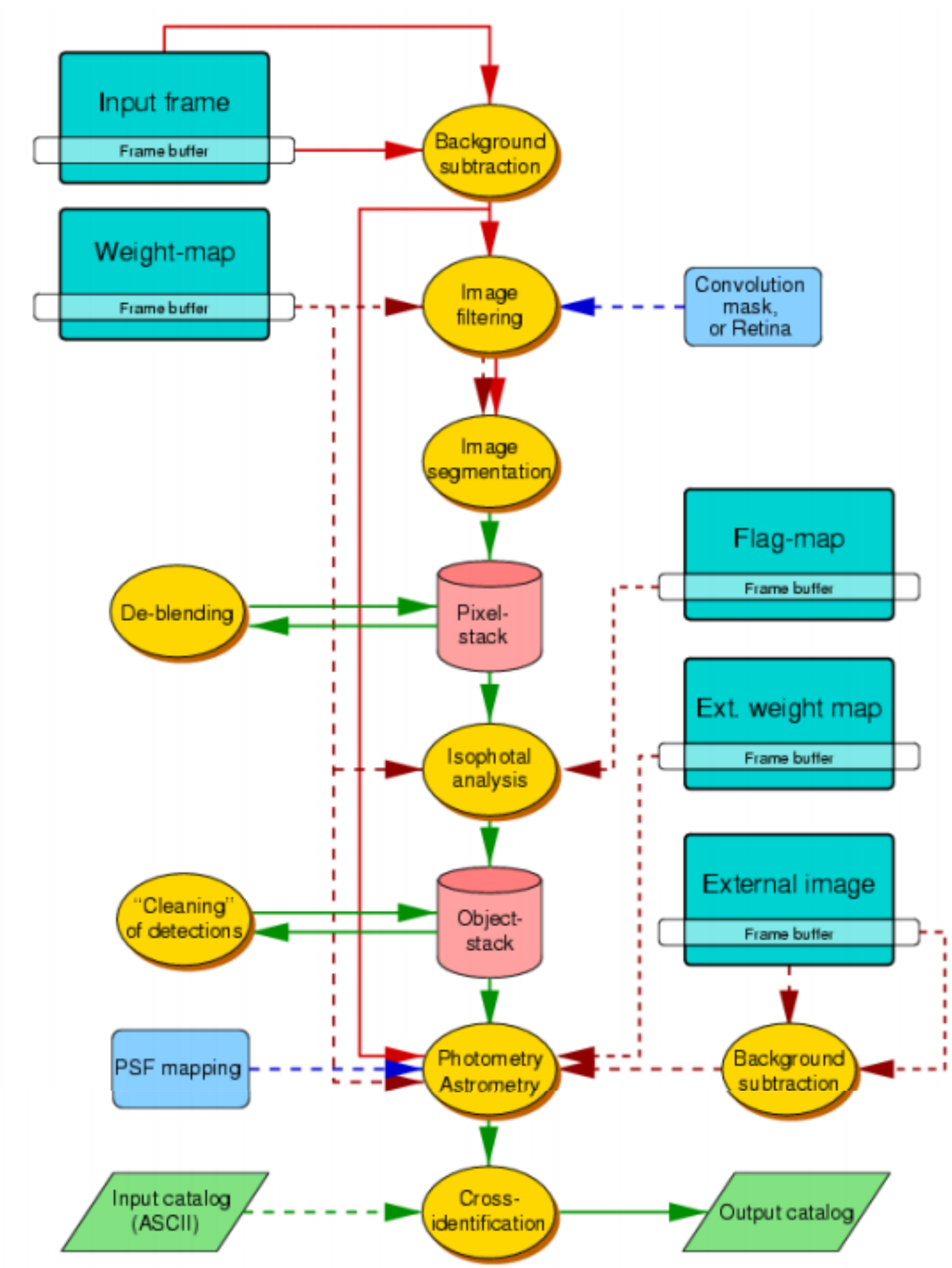
When measuring the spectra of the observed light, the absorption lights will be red or blueshifted by an amount determined by Hubble's Law. The combination of the spectroscopic and photometric redshift allow for the creation of a probability distribution, *P(z)*, that shows the probability that an object in the image is located at the specified redshift. This probability distribution will be created by fitting various spectral energy distributions (SED) until ones that match are found. An SED is a plot of flux versus wavelength that shows where the majority of the light emitted from an object falls on the electromagnetic spectrum. SEDs are drawn from models of real galaxies at low redshifts. One of the programs used in this research has a library of SEDs which it uses in order to calculate the probability distribution for the redshift. By fitting SEDs, an estimate for the type or types of galaxies for each object can be made.

The reason that the result of the photometric analysis is a probability distribution is because of the uncertainty in the nature of the object itself. The spectra of a galaxy varies with its redshift as its emitted radiation is redshifted, thus for a known object it can be easy to determine its redshift by seeing how much its spectrum has shifted. For an unknown galaxy, it is not as easy to determine its redshift because its spectra could be the result of different objects at different redshifts with the same spectra. For example, if there is a high energy and thus low wavelength galaxy that is very redshifted, its spectra might appear to be the same as a low energy and high wavelength galaxy at a much lower redshift. Thus, the probability distribution can have multiple peaks representing galaxies of different spectra and different redshifts that have similar observed spectra.

Data, Tools, and Procedure

As mentioned previously, images were of high redshift galaxies. The images were trained off the COSMOS field. The COSMOS field, also known as the Cosmic Evolution Survey, is a survey taken by the Hubble Space Telescope (HST) over a two square degree field with the Advanced Camera for Surveys. It is the largest survey HST has done and makes for an ideal training set given the number of objects in the field of view. The purpose of training using the COSMOS field data was to measure the photometric redshift to observe any shift or offset in the data from the filters. A program used in this research that calculates the redshift of the galaxies can take a photometric redshift offset as an input parameter, allowing the other images to be more accurate. All images of the COSMOS field and other clusters were taken from the National Optical Astronomy Observatory's (NOAO) database (http://portal-nvo.noao.edu/search/query). Images were collected in multiple filters. The filters used were the R, G, B, V, I, and Z bands. Each of the images was run through Source Extractor. Source Extractor is a program used for automatic detection and photometry of objects within the Flexible Image Transport System image (FITS image).

Source Extractor works by measuring the photometric flux of each pixel. Under a certain flux threshold, pixels are labelled as background pixels and are subtracted from the image. The remaining pixels are taken to be parts of objects in the image. If there are two distinct peaks in the light distribution of an object, it is taken to be two separate objects and is split up. Source Extractor then measures the shapes and positions of the objects and then reconsiders its original detections. It then performs the photometry, classifies the objects as either a star-like or galaxy-like object, and outputs a catalog with a list of parameters for the image set by the input files. In this instance, the parameters are the object's identification number, its photometric flux, the error in its photometric flux, and its spectroscopic redshift.



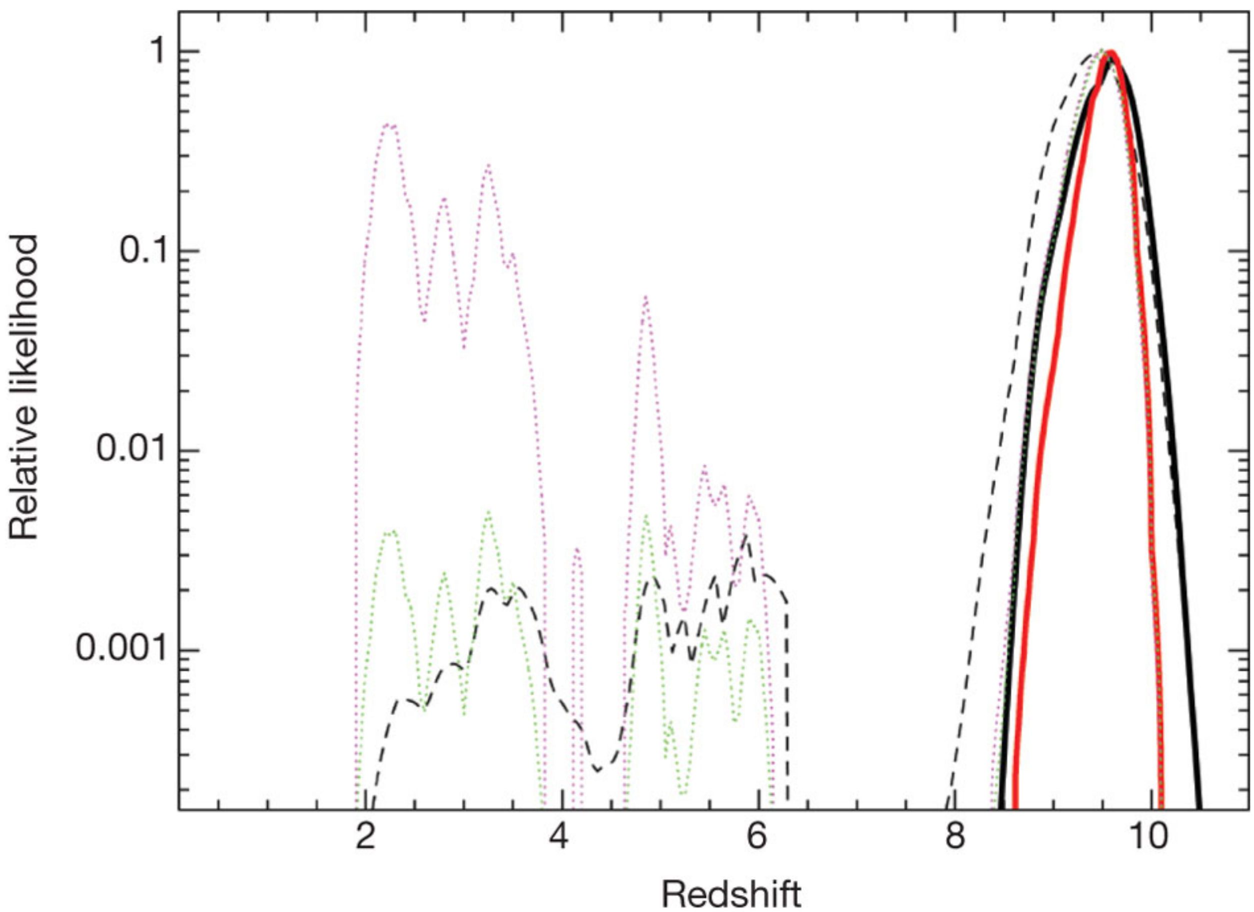
The above image is a diagram of the process Source Extractor takes given in the second manual. Image Source:

http://astroa.physics.metu.edu.tr/MANUALS/sextractor/Guide2source_extractor.pdf

```
# fiat 1.0
# converted from sextractor format Mon Apr 6 15:34:42 EDT 2015
# via the command: /usr/dist/dls/bin/sex2fiat test.cat
# TTYPE1 = NUMBER / Running object number
# TTYPE4 = MAG_ISO / Isophotal magnitude
# TTYPE5 = MAGERR_ISO / RMS error for isophotal magnitude
1 22.1128 0.0612
2 22.0076 0.0523
3 22.3621 0.0604
4 21.9984 0.0492
5 21.6383 0.0391
6 21.3925 0.0451
7 21.1146 0.0384
8 22.6762 0.0711
9 22.5625 0.0698
10 21.9041 0.0505
11 22.4848 0.0624
```

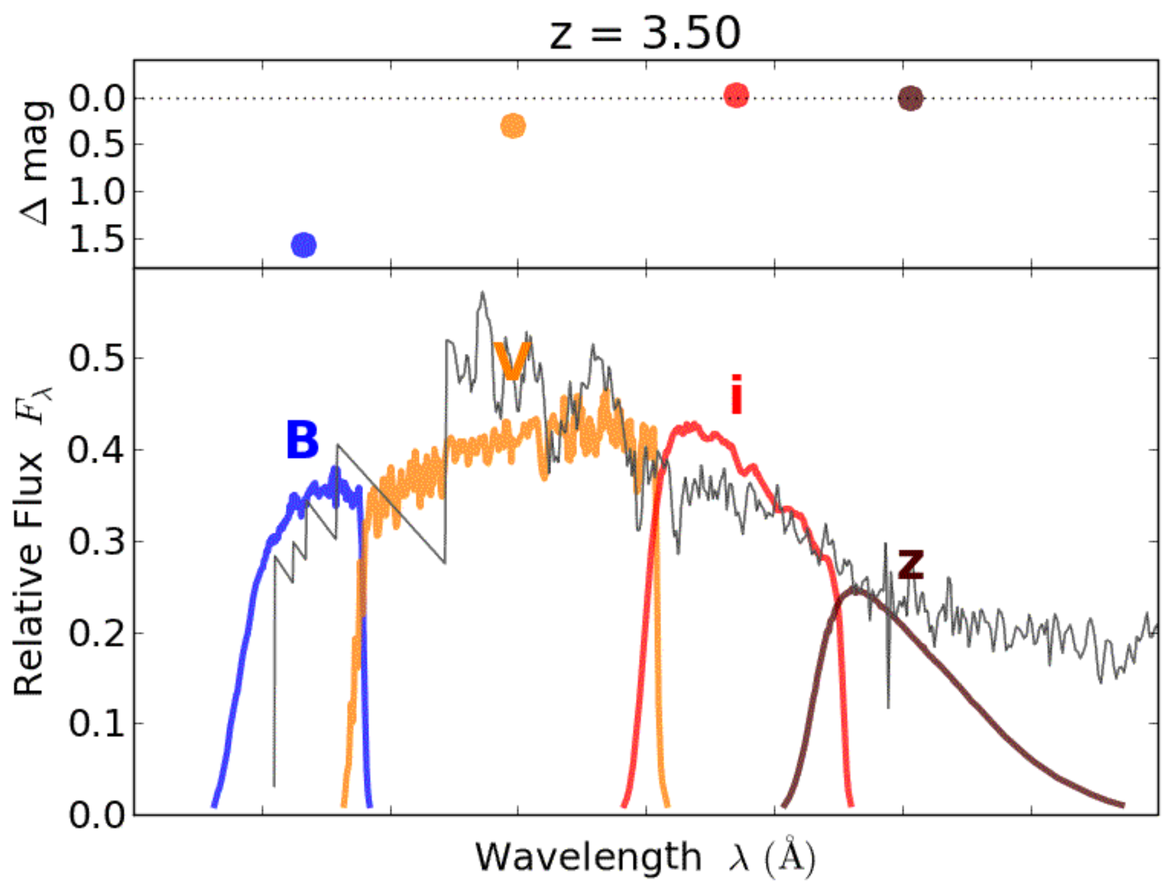
Image above is an example catalog taken from the COSMOS field after being pruned for relevant information

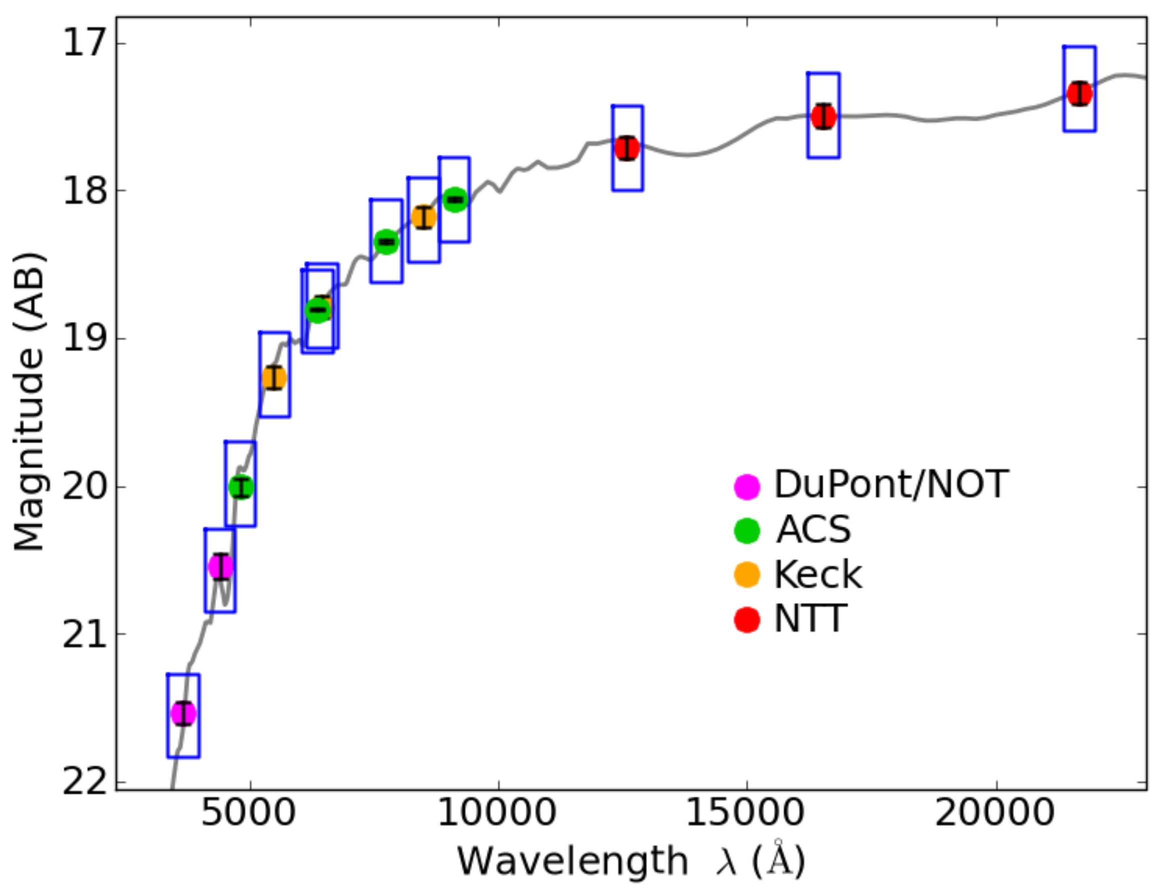
The output catalogs contained the magnitude and magnitude error for each of the objects detected in the image for that filter. A python script was written to take all of the information from the individual filter catalogs and to combine them into a single catalog with the objects identification number, its magnitude and magnitude error in each band, as well as its spectroscopic redshift. The spectroscopic redshift was taken from the Infrared Processing and Analysis Center (IPAC) at Caltech. This all bands inclusive catalog was run through Bayesian Photometric Redshift (BPZ). BPZ is a program that takes images with multiband photometry catalogs for its objects (like the catalogs produced by Source Extractor) along with a spectroscopic redshift (taken from IPAC – this measurement is also a probability distribution *P(z)*, however, the distribution is very small) and estimates a redshift probability distribution for the objects in the catalog. BPZ works by fitting the input photometry with spectral energy distributions redshifted by different amounts. Multiple grid of fit choices are attempted, varying the redshift and spectral type, which gives the likelihood distribution $P(z,t)$. BPZ uses Bayesian interference and priors (in this case, $P(z,t \mid m)$) to estimate the photometric redshift. With the prior, BPZ produces a result of $P(z)$ for the objects. This $P(z)$ is a probability distribution for each object's redshift. It is non-Gaussian, and shows the probability the galaxy happens to be within a range of redshifts.



Green and red lines in image show examples of *P(z)* curves generated by BPZ. Source:

http://www.nature.com/nature/journal/v489/n7416/fig_tab/nature11446_F3.html



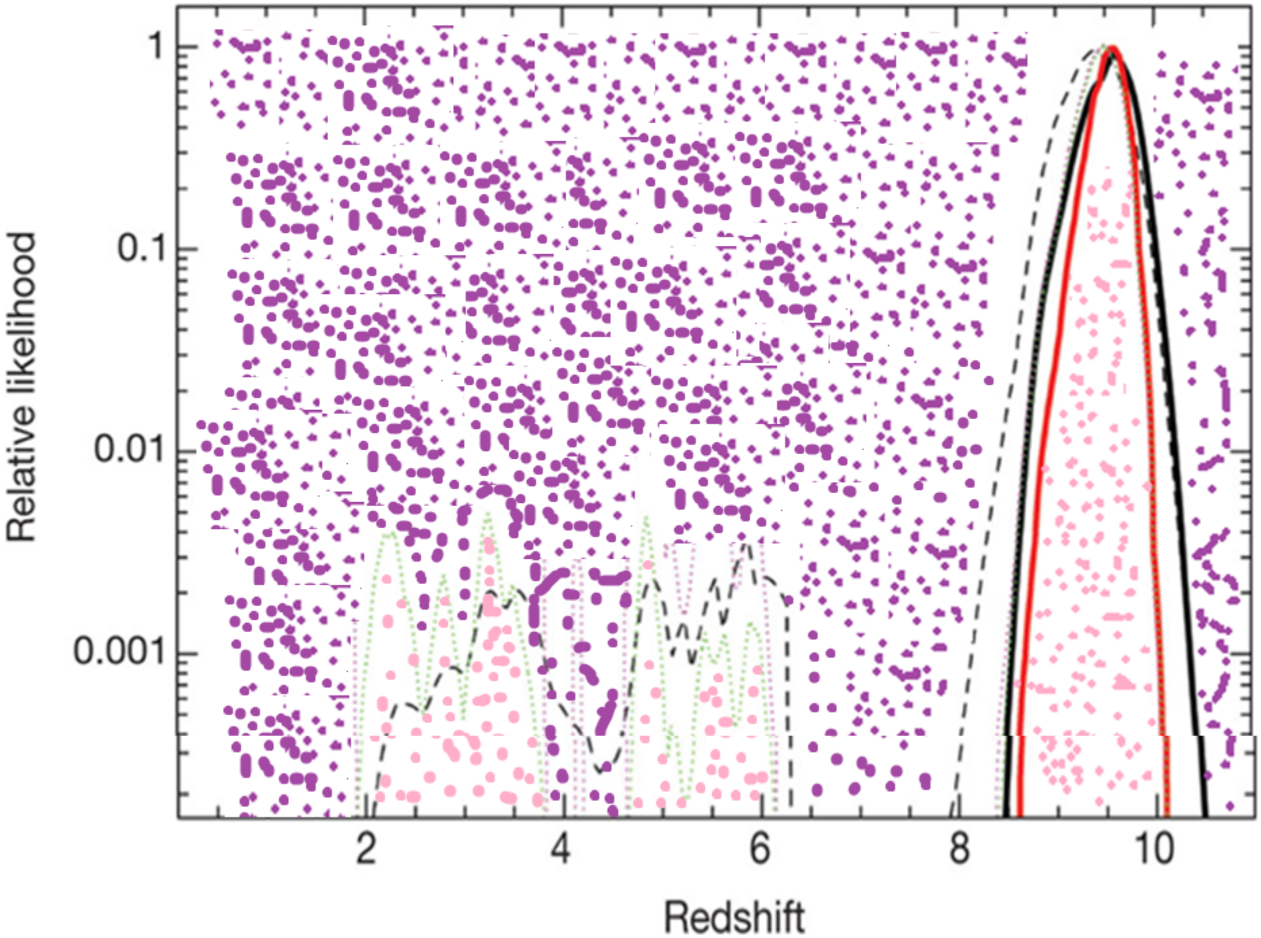


An example of SED fitting taken from: http://www.stsci.edu/~dcoe/BPZ/intro.html , and an example result of running BPZ on a catalog taken from http://arxiv.org/abs/1005.0398 The first image shows the process of fitting for an SED as the redshift varies (originally a gif, the picture is a freeze frame at z = 3.50). The second image shows and SED fit for a z = .18 elliptical. The colors of the circles correspond to different telescopes, and the letters to the band that they were measured in.

Once an image has all of its objects assigned a *P(z)*, the next step is to take these probabilities and turn them into an estimated distance. In order to do this, a Monte Carlo method is used. A Monte Carlo method is a method that solves a problem by generating a large amount of random numbers that fall within the domain of acceptable values, and observing the results to see if the numbers obey a property, law, or preference. In this particular problem, the redshift probability distribution for each galaxy in an image *P(z)* is used, and a large number of redshifts are used to sample the distribution within the range of redshift values that BPZ computes. From there, observing the peaks with the highest number of random inputs gives the most likely redshift values for that object.

However, as it was mentioned before there can be multiple peaks in the redshift distribution for on object, owing to its uncertain composition and location. Therefore, multiple probable redshift estimations will be taken from each Monte Carlo simulation so that different version of the lensing distribution map can be created. This has the advantage of adding the real uncertainty into the distribution maps that are currently lacked due to the assumption that all of the galaxies are at the same redshift. It also provides multiple maps so that a best fit map can be found, as opposed to just picking the most probable redshift for each object and assuming that it must be the one. This creation of multiple maps by the Monte Carlo method allows for mock observations. By using the different redshift measurements, the effect of assigning each redshift to each object can be known, allowing for the selection of values that best fit the models.

For an example of the Monte Carlo simulation being run on a *P(z)* distribution, we take the previous probability distribution and add the random sampling dots for the red curve.



The pink dots represent points that sample inside of the probable redshift value, and the purple dots are points that were outside. With sufficient sampling across the entire graph, the samples can be computed to return which redshift values have the highest number of dots and are therefore the most probable redshift values for that object.

Results and Discussion

The research did not reach fruition. No comments can be made on the effectiveness of this method and the resulting changes on the matter distribution of gravitational lenses and galaxy clusters, or on the total error induced by the incorrect inclusion of foreground galaxies. That being said, there can still be discussion on the potential ramifications of the results, as well as discussion of how the research can help others in their own research.

Jacqueline McCleary, a graduate student in Professor Dell'Antonio's lab, researches the substructure of nearby galaxy clusters. With this method described in this paper, her research will benefit in two ways. The first way is the rejection of foreground galaxies. As mentioned previously, the assumption that all galaxies in the image are at the same redshift induces error by way of the inclusion of galaxies that are in front of the lens and the background sources. The Monte Carlo multiple matter distribution maps method results in a more accurate mapping that reduces the inclusion of foreground galaxies in the matter profile, in turn leading to a more accurate result on the scale and shape of the substructure. This redshift technique will also help in finding a more accurate redshift measurement for the source galaxies in the images her work uses. In addition to the increased accuracy of her work, the ability to find new cluster candidates behind the nearby clusters by using a joint spatial and redshift probability distribution, with the redshift part of the probability distribution being produced from this work.

Paul Huwe, a former graduate student, worked on measuring the masses of substructure clumps of dark matter in HST images. His work would benefit from the results from the foreground rejection as well. Many of the clusters that his work examines are at very high redshifts, which means that the number of foreground clusters is much higher than for the nearby clusters. The increased number of foreground clusters means that the level of error added by their inclusion in the matter distribution in the source is higher than closer clusters.

Finally, Ryan Michney, another graduate student in the lab, performs work on blind searches for clusters. These searches require redshift estimate and as the mass estimates of the clusters require a full *P(z)* distribution to properly account for mass errors due to uncertainty in cluster galaxy selection. The method in this research will help to accurately identify which galaxies are a part of the cluster, and which are not, which leads to more accurate measurement of the mass and therefore or a more complete matter distribution profile.

In addition to my on work on this new method of redshift estimation, there has been some research in this subject. The majority of the work has been concerned with low redshift objects whose distances are easier to measure and whose information was more easily accessible. In 1995 there was research done (http://adsabs.harvard.edu/abs/1995PASP..107..945F) by Japanese scientists Fukugita, M.; Shimasaku, K; and Ichikawa, T. on closer galaxies with known redshifts, as observed by Sloan Digital Sky Survey (SDSS) with Johnson and Cousins filters. This method used galaxy colors to figure out the type of galaxy, either elliptical or spiral, via their spectroscopic energy distribution and a spectrophotometric synthesis technique. However, this method was unable to work with faint galaxies, which they were unable to accurately measure under their method, leading to a dead end. This research's method produces a result that is capable of working at high redshifts and does not have some of the limitations of the research based on galaxy color.

The Weighing the Giants project (http://kipac.stanford.edu/collab/research/highlights/tidbits2012/pzc) has worked on a redshift estimation method similar to the method described in this paper. The key difference between their work and mine is that the Giants team does not use the Monte Carlo method. The Giants team uses the probabilistic result of their photometry to find the probability weighted distance as opposed to the most probable distance. As distances are anisotropic, the most probable distance is not the most probable redshift. In addition, another major difference between the two is that the Giants group uses a singular probabilistic estimate, whereas this research does a many realizations estimate with the probability distribution given by BPZ.