

SPECTROSCOPIC REDSHIFTS TO 1300 LOW-MASS GALAXIES WITH $0.3 \leq z \leq 0.4$ IN THE COSMOS FIELD

B. LORENZ¹ AND D. D. KELSON²

The Observatories, The Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101

Draft version August 26, 2017

ABSTRACT

Recent theoretical models and hydrodynamic simulations of galaxy evolution give predictions for the star-forming properties and sizes of low-mass galaxies. Due to their faintness, there exist few unbiased samples of low-mass galaxies to test these predictions. However, the Hubble Space Telescope's deep observations of the COSMOS field allow for the creation of an unbiased collection of galaxies with stellar masses $10^9 M_{\odot} \leq M \leq 10^{10} M_{\odot}$ in the redshift range $0.3 \leq z \leq 0.4$. In particular, the UltraVISTA catalog (Muzzin et al. 2013) identifies roughly 250,000 objects to $K_{s,tot} = 23.4$ mag (AB) in COSMOS and includes estimates of photometric redshifts for the galaxies. To study the star formation and morphological properties of an unbiased sample of galaxies, we targeted 1131 unique galaxies in the above redshift range with the IMACS spectrograph on the Magellan Baade 6.5m telescope. Using real domain, inverse-variance weighted cross-correlation with galaxy templates from the Sloan Digital Sky Survey, we measure accurate spectroscopic redshifts for 86% of the targeted galaxies. Our work provides a population of low-mass galaxies for which we can measure star formation rates, furthering our understanding of low-mass galaxy evolution.

Keywords: galaxies - evolution

1. INTRODUCTION

Galaxy evolution is a well-studied topic in recent years, though questions remain as to how galactic properties can affect growth and star formation. In particular, low-mass galaxies can be difficult to observe at distant redshifts, so much of our understanding of distant galaxies depends on simulation, models, and nearby observations. New theory and simulations have made observable predictions regarding star formation histories, and, with the latest generation of space telescopes, we can measure significantly faint distant galaxies to test these ideas.

The stochastic processes approach to galaxy evolution of Kelson (2014) and Kelson, Benson, & Abramson (2016) predict variance in the star formation rates (SFR) of low-mass galaxies. Their results are based solely on first principles, independent of knowledge of particular astrophysical processes. Instead, they assume that star formation is a nonnegative stochastic process to derive predictions for observable scatter in SFR at a fixed mass, among other observables.

Hydrodynamical Simulations with FIRE (Hopkins et al. 2014) show that low-mass galaxies have “bursty” star formation between 9 Gyr and present times, while medium sized galaxies have relatively constant SFR (El-Badry et al. 2016). Additionally, they find that SFR strongly correlates with galaxy size. However, galaxy samples to test these predictions are difficult to obtain. Low-mass galaxies are intrinsically less luminous than massive ones, and our local low-mass galaxies are neither numerous enough nor a representative sample. To avoid biases, galaxies must be selected from a deep survey to ensure a uniform stellar mass limit that is inde-

pendent of star formation history and redshift. Using recent Hubble Space Telescope observations, we can now test these simulations with a sample of low-mass galaxies with $0.3 \leq z \leq 0.4$, which also provides a convenient population to examine the aforementioned predictions of Kelson (2014).

The UltraVISTA photometric survey (Muzzin et al. 2013) of the COSMOS field forms the basis of our unbiased sample. COSMOS is a 2 deg² field of the sky that has been studied extensively, particularly by the Hubble Space Telescope (Scoville et al. 2007). Sections of the field were investigated in greater detail by Muzzin et al. (2013), who creates a catalog of 250,000 objects in a 1.62 deg² area down to $K_{s,tot} = 23.4$ mag (AB). They identify galaxies, then calculate photometric redshift measurements for each galaxy using EAZY (Brammer, van Dokkum, & Coppi 2008). EAZY matches the overall spectral energy distribution of the galaxy with templates in order to estimate its redshift.

The UltraVISTA Catalog contains roughly 3,000 galaxies with redshift $0.3 \leq z \leq 0.4$ and masses $10^9 M_{\odot} \leq M \leq 10^{10} M_{\odot}$. We select 1131 galaxies in this range to form our sample of low-mass galaxies. These data are required for accurate measurement of redshift and Balmer line fluxes, which are then used to compute star formation rates. In this work, we present our methods of processing the data from our sample of galaxies and the results of our redshift measurements.

2. OBSERVATIONS

Data were taken using the IMACS spectrograph (Dressler et al. 2011) on the Baade 6.5m Telescope, one of the Magellan Telescopes at Las Campanas Observatory. IMACS is a multi-object spectrograph, allowing for the acquisition of data for roughly 150 COSMOS galaxies at a time. Data were taken on two observing runs, with eight observations in February 2016 (1066 galaxy spec-

¹ Pomona College Physics and Astronomy, 610 N. College Avenue, Claremont, CA 91711

² The Observatories, The Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101

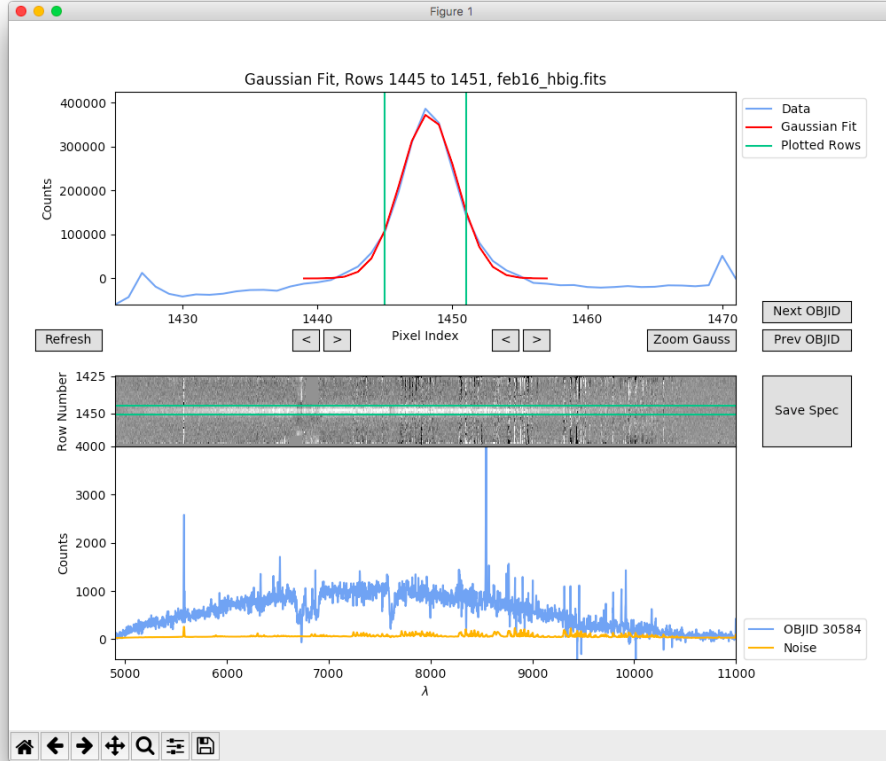


Figure 1. Example of spectrum extraction. The top panel shows a Gaussian fit to the distribution of light along the slit. The middle contains the 2-D spectrum. Green lines over the Gaussian and 2-D file indicate the extraction region, which gives the 1-D spectrum shown in blue in the lower panel. Noise is added in quadrature, appearing as orange in the 1-D spectrum. The lines are placed by default at 1.4 standard deviations from the mean, but can be changed by the user if they wish to fit over a different range for particular objects.

tra) and two observations in February 2017 (259 galaxy spectra). Spectral masks were created for each of the observations, containing slits for galaxies, a few hot stars to be used in background subtraction, and boxes around alignment stars. Some galaxies were observed more than once since the observations in 2017 revisited objects from 2016 with low signal-to-noise, as well as new targets.

3. DATA PROCESSING

The spectra were processed as described in (Kelson et al. 2000) and (Kelson 2003), including sky subtraction and rectification of the spectra. For our analysis, we first extract only the portions of the data that contain galaxies, correct them for atmospheric absorption, and finally correlate them with spectral templates to determine their redshift.

3.1. Extraction of Spectra

First, a one-dimensional spectrum of each object is extracted from the two-dimensional spectral files while maximizing signal-to-noise. All of the spectra are collapsed along the dispersion axis, summing over wavelength, to produce an object profile (Figure 1). This ideally forms a Gaussian distribution with the object’s signal in the center and noise dominating in the tails. In practice, slits may have multiple objects in them, be contaminated with cosmic rays, or have a range of other problems. However, most of the data can be approximated well by fitting a Gaussian curve, returning a mean

and standard deviation for a fit of each object. The optimal signal-to-noise ratio is obtained at 1.4 standard deviations from the mean, found by assuming a Gaussian signal and dividing that by \sqrt{n} noise, where n is the number of rows. The row number at 1.4 standard deviations from the mean is computed, rounding to the nearest row, then rows in that range are summed to create a one-dimensional spectrum. The noise is added in quadrature over the same region.

3.2. Telluric Correction

The atmosphere is less transparent in the infra-red, causing it to absorb a fraction of the signal emitted by each object. In a given observation, this telluric absorption should be nearly identical for all objects, since the atmosphere does not change significantly over such a small solid angle. We identify a hot star in each observation to correct for this absorption. Hot stars have few absorption features due to their ionized atmospheres, so the absorption from the Earth’s atmosphere can be easily seen and corrected.

Our detector is more sensitive in the central region and less sensitive at the ends. To isolate the effects of the atmosphere, we must correct for this sensitivity function. Each spectrum is first smoothed by setting the value of each pixel to the median of the pixels nearby, ignoring clear emission and absorption. Then, a polynomial is fit to the smoothed spectrum using single value decomposition. Dividing the star’s spectrum by this polynomial

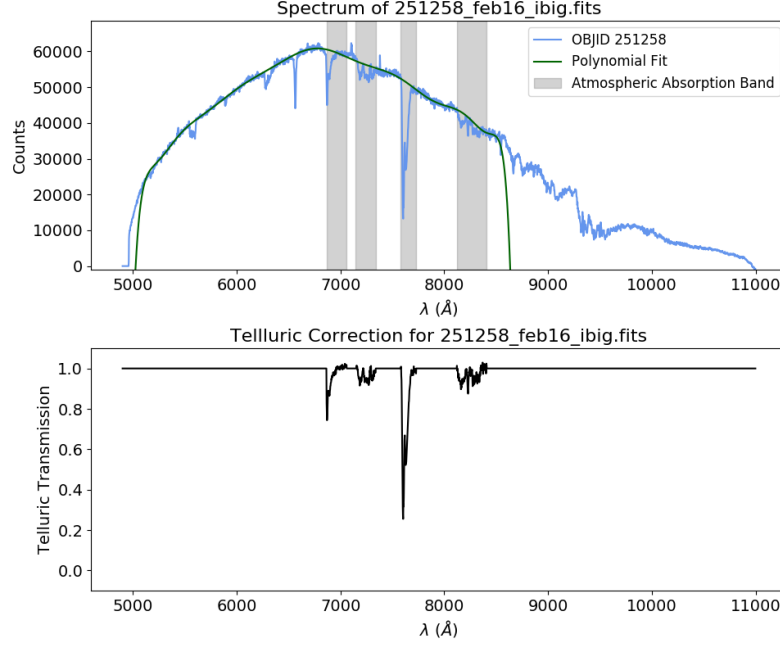


Figure 2. Generation of a telluric correction file from a hot star spectrum. In the top panel, the star’s spectrum (blue) is fit by a polynomial (green) to remove our detector’s sensitivity function. Then the spectrum is divided by the polynomial in the regions with significant atmospheric absorption (shaded gray). This produces the telluric correction file shown in the lower panel. Regions outside of these bands are set to unity in the correction file so that the star spectrum’s features do not affect our galaxy spectra.

corrects for our sensitivity function, since the general shape is removed, leaving only the deviations from that shape (i.e. emission and absorption). This division is performed in the regions known to have atmospheric absorption, generating a telluric correction file (Figure 2). Next, we apply the correction file to each object in the observation.

Each object’s spectrum may be shifted slightly due to how the data was processed. To align each object with its telluric correction file, we apply a simplified version of the cross-correlation routine that is used to find redshifts. A polynomial is fit to each galaxy in the same manner as the hot stars, smoothing the spectrum and clipping the features. The galaxy is subtracted by its polynomial to correct for sensitivity, then multiplied pixel-by-pixel with the correction file. Summing over all pixels in this product gives the value for the correlation coefficient, $C(x)$, at $x = 0$. Then we shift the correction file by one pixel to $x = 1$ and repeat, continuing to repeat this process until we have calculated the correlation coefficient at each pixel between $x = -20$ and $x = 20$. The maximum correlation coefficient corresponds to the pixel shift that best aligns the spectrum and the correction file. We find the maximum $C(x)$, shift the correction by the necessary amount, then divide the original spectrum file by the correction file (Figure 3).

3.3. Cross-Correlation

Here we employ a real domain, inverse-variance weighted cross-correlation routine to match our galaxy spectra with Sloan Digital Sky Survey (SDSS) templates from data release 2. We use five templates from SDSS, ranging from a young, blue, star-forming galaxy (template 27) to an old, red, quiescent galaxy (template 23). SDSS compiled these templates to model an ideal galaxy spectrum for a galaxy of a given age. Since we are

working with observational data, our spectra are a superposition of the “real” spectrum and noise from the instrument, atmosphere, and contamination from other objects. To measure an accurate redshift, we can cross-correlate these templates with our galaxy spectra, constructively interfering the real features while canceling out the noise.

The advantages of operating in the real domain instead of the Fourier are significant. While the Fourier domain is faster, it is limited in that each pixel is weighted evenly. The real domain allows for the introduction of a mask to remove bad pixels (e.g. atmospheric absorption lines, gaps in our detector, cosmic rays) while also weighting our correlation by the inverse of the variance in each pixel. This puts more weight on data that is dominated by signal, and less on data in regions with larger noise. More explicitly, we compute the correlation coefficient at a given redshift, $C(z)$, as

$$C(z) = \sum_{i=1}^n \frac{M_i}{\epsilon_i^2} L(T)_i L(G)_i \quad (1)$$

where the correlation is over n pixels, M is a mask that is equal to zero or one for each pixel, ϵ_i is the noise in pixel i , and $L(T)$ and $L(G)$ are the continuum-subtracted template and galaxy spectrum, respectively (Kelson, Martini, & Mulchaey 2005).

Another advantage of real domain cross-correlation is that the error in redshift is easy to calculate from the shape of $C(z)$ vs z . For more information on the calculation, see (Kelson, Martini, & Mulchaey 2005). Here, we will simply note that the one-sigma error in redshift is given by

$$C(z \pm \sigma) = C(z_{peak}) - \frac{1}{2} \quad (2)$$

To continuum-subtract the template and galaxy spec-

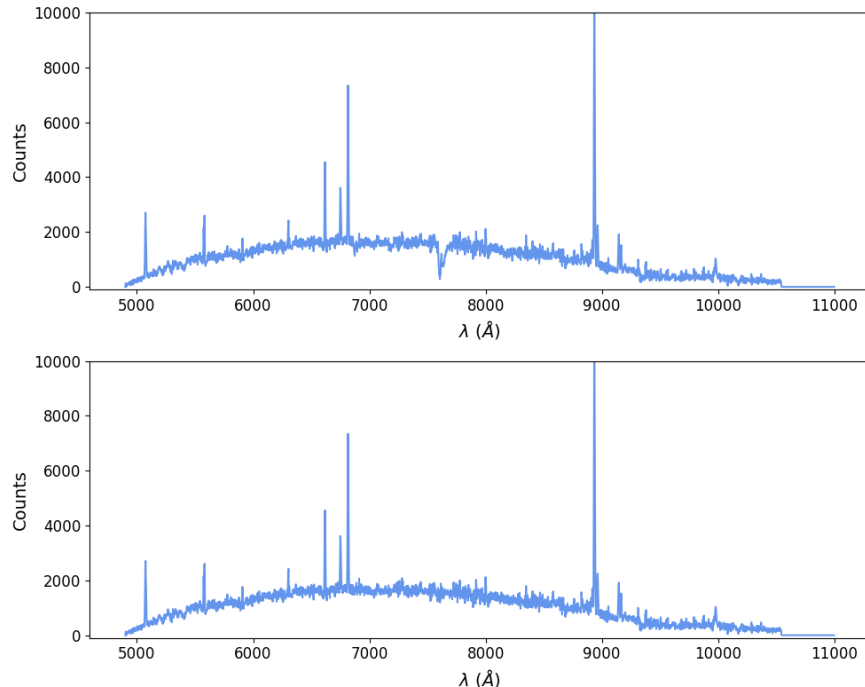


Figure 3. The atmospheric absorption present at 7650 Å in the top panel is clearly corrected in the bottom panel. The bottom panel is the result of correlating the galaxy with the telluric correction file, then dividing the galaxy by the telluric correction spectrum.

tra, we perform a similar procedure as described earlier. The galaxy is smoothed and clipped of features, and a mask is applied over regions that should be completely ignored. This process is automated at first, but can be modified by the user. The galaxy is fit with a polynomial, then subtracted by the polynomial, leaving a spectrum that is independent of the sensitivity function of our detector. It is then multiplied by $\frac{M}{\epsilon}$, where M is the mask and ϵ is the noise, which weights the spectrum towards better data. We repeat this process with a template shifted to a range of redshifts $0.2 \leq z \leq 0.5$, fitting a polynomial to each, subtracting it out, then multiplying by the same $\frac{M}{\epsilon}$ so the templates and galaxy are weighted equally. We compute $C(z)$ at each of the redshifts by simply multiplying these described galaxy and template spectra together, then summing over wavelength. This is equivalent to equation (1). Finally, we repeat this process with each of the five templates.

We have developed a graphic interface to quickly analyze a large set of redshifts using the cross-correlation technique (Figure 4). It first performs this correlation with the emission lines left in, then displays the result to the user. The template is overlaid on the spectrum at the maximally correlating redshift, with zoomed-in regions on the Ca H and K absorption lines and OIII and H β emission lines. The user may flip between templates to view their correlation and measured redshift, and can recalculate the correlation for a given template over a smaller range if necessary. They can also change the mask applied to the galaxy if they notice bad regions that were not removed automatically, and can enlarge or reduce the wavelength range over which to correlate. They can toggle between keeping or clipping emission lines, which should be clipped for absorption-dominated galaxies. If they are not, the emission lines on the tem-

plate may correlate with noise in the galaxy. Finally, the user can also flag each object as a “Star” (we are not interested in the stars and must filter them out), “Bad” (something went wrong in data processing), “Unsure” (if they are not confident in the measurement), “Revisit” (if this has been examined twice and is still unsure), “Other” (outside of the redshift range or something else noteworthy), and “Unusable” (the data is so bad that it cannot be easily recovered).

We run this interface over each of the 1414 stars and galaxies, measuring redshifts and sorting them according to the flags above.

4. RESULTS

While there were 1131 galaxy targets, we took 1325 galaxy spectra, repeating 194 of the targets for better signal-to-noise and to test the effectiveness of our measurements. Of the 1325 galaxy spectra, we were able to measure confident redshifts for 1138 of them (86%). Only 44 redshift measurements fall outside our target range of $0.3 \leq z \leq 0.4$ (Figure 5).

Muzzin et al. (2013) claimed that roughly 20% of galaxies in the ULTRAVista catalog with $10^9 M_\odot \leq M \leq 10^{10} M_\odot$ should be quiescent. As shown in Figure 5, our measurements have 21% of galaxies correlating most strongly with SDSS template 23, the quiescent template.

4.1. Measurement Confidence

We assess the confidence of our redshift measurements by examining the galaxies that were observed multiple times. Each spectrum was analyzed independently, meaning that we did not use the first redshift measurement of a galaxy to influence the measurement of the second. When plotting first redshift measurement vs second redshift measurement for the 194 repeated galaxies (Figure 6), we find that every galaxy for which we were

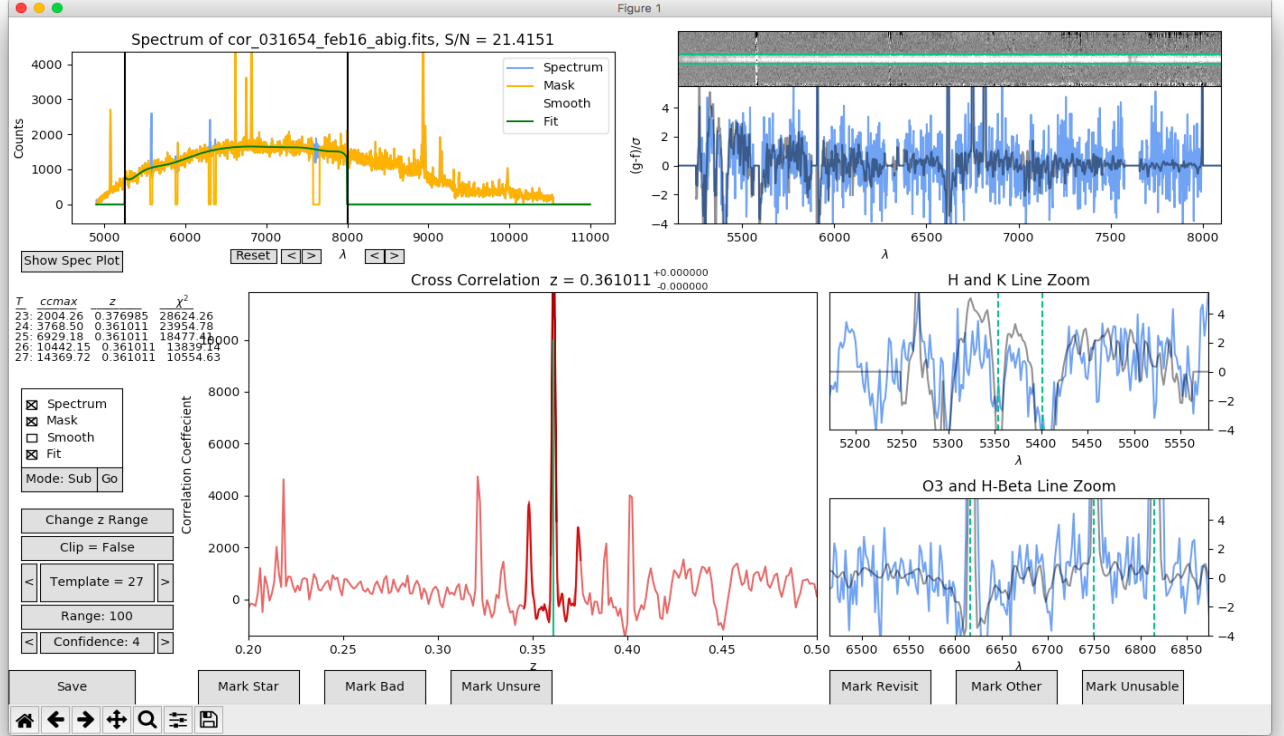


Figure 4. Cross-correlation for an object with strong emission. Our graphic interface allows the user to easily check if they are confident in the redshifts with the plots on the right, as they overlay the template and galaxy spectra. The cross correlation plot allows the user to determine how strongly this correlation is favored compared to other possible redshifts. The buttons at the bottom allow the user to assign a confidence to their fit and flag interesting data.

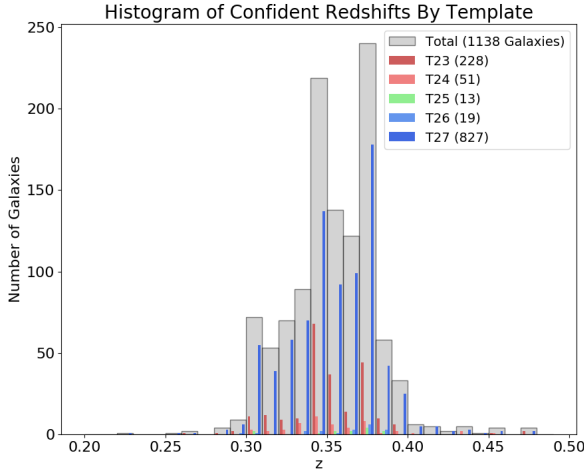


Figure 5. The gray bars are a histogram of redshift for galaxies that are not marked as “unsure,” and the colors break down which template gives the strongest correlation coefficient for each galaxy. These range from the star-forming T23 (blue) to the quiescent T27 (red). In the legend, we display the number of galaxies that best correlated with each template. 96% of the galaxies are in our target redshift range of $0.3 \leq z \leq 0.4$. Also, 21% of galaxies correlate best with the quiescent template (T23), in agreement with the roughly 20% in the UltraVISTA catalog (Muzzin et al. 2013). Finally, most galaxies are either star-forming (T27) or quiescent (T23), with few in the regions between. This agrees with the observation that galaxies quickly transition from star forming to quiescent.

confident in both measurements (157 galaxies) has the

same redshift both times. In other words, the only time that the measurements did not agree was for galaxies we marked as having “unsure” measurements. While we cannot be fully confident that every measurement is accurate, this result gives us very high confidence in all of the galaxies that are not marked as “unsure.”

4.2. Uncertainty Analysis

Each redshift measurement has an associated one-sigma error from the cross-correlation mentioned in section 3.3, equation (2). Note that these errors are only accurate for galaxies measured using absorption (364 galaxies). However, these errors are based on the sigma maps created in processing the data, which may have overestimated or underestimated the true errors.

We examine the distribution of reduced χ^2 , which should be equal to unity for each measurement if errors are appropriately accounted for (Figure 7). Instead, our distribution is skewed right with median at reduced $\chi^2 = 2.27$. This suggests that our errors are underestimated by a typical factor of $\sqrt{2.27} = 1.51$. The median of our errors in redshift for absorption galaxies is 35 km/s, so our true median uncertainty is around 50 km/s.

To see if this is reasonable, we can examine the difference in errors for the 47 objects that were observed twice with absorption. Their standard deviation is roughly 80 km/s, giving a slightly greater uncertainty than our correlation error estimation. However, it is clear that our uncertainties are mostly below 100 km/s, particularly with the inclusion of emission line galaxies, which

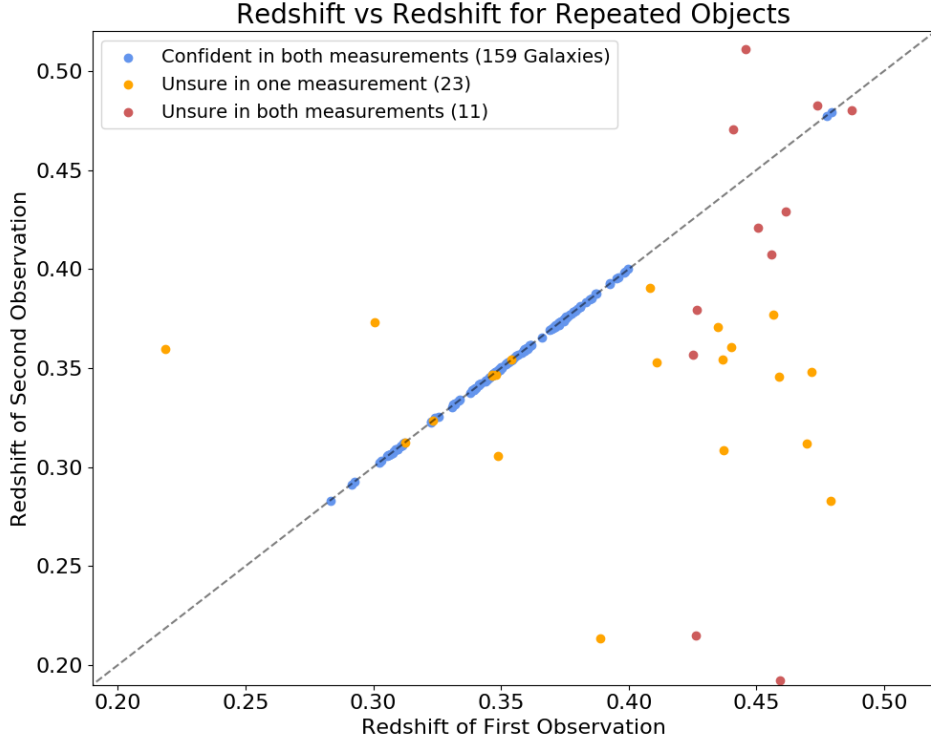


Figure 6. We assess the confidence in our methods by comparing our measured redshifts for objects that were observed multiple times. Objects not marked “unsure” had the same measured redshift both times, suggesting that measurements not marked as “unsure” can likely be trusted.

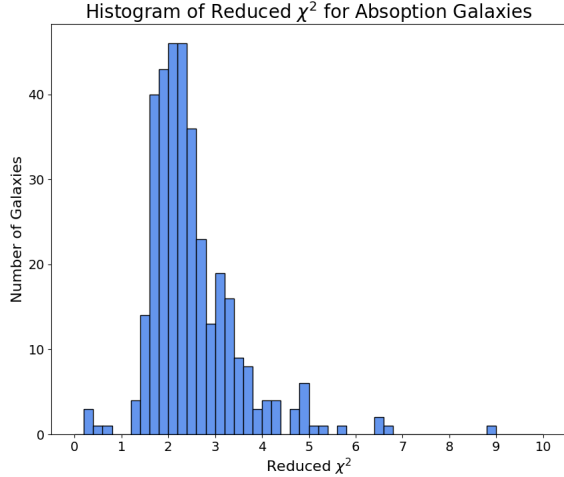


Figure 7. The distribution of reduced χ^2 for absorption galaxies is skewed right with a median at 2.27, implying that our errors are likely underestimated by a typical factor of $\sqrt{2.27} = 1.51$

drops the standard deviation to 20 km/s. This level of redshift precision allows for analysis of where the galaxies lie in space, giving insight into the large-scale structure of this region of the universe.

4.3. “Unsure” Galaxies

While we did measure redshifts for 86% of the sampled galaxies, the remaining 150 objects are heavily biased towards quiescent galaxies. We see this in examining the specific star formation rate (SSFR) vs stellar mass plot, as well as in the UVJ diagrams (Figures 8 & 9).

SSFR is (star formation rate) / (stellar mass), so it is a measure of how quickly a galaxy is forming stars that is independent of mass. Our “unsure” galaxies mostly fall in the region of this plot with low SSFR and low stellar mass. This is unsurprising since quiescent galaxies are fainter due to less emission from young stars, giving a lower mass-to-light ratio. Lower mass galaxies are also intrinsically dim. Low-mass, quiescent galaxies therefore have the lowest signal to noise, making it much more difficult to obtain a reliable redshift for them.

We can also visualize the population that we are missing in UVJ diagrams with a mass cut. UVJ diagrams are color-color diagrams, plotting V-J vs U-V. Muzzin et al. (2013) propose a boundary region to cut the quiescent population from the star forming population. We have measurements for almost all of the higher mass galaxies ($9.5 \leq \log(M) \leq 10$ solar masses), with measurements distributed roughly equally across the sample. The lower mass galaxies are far more biased, lacking accurate redshifts in the quiescent region.

To mitigate this bias, we would like to obtain confident redshifts to as many of the galaxies as possible. Dan Kelson revisited the “unsure” galaxies in two of the ten observations, identifying spectral features by eye. His measurements recovered the 14 galaxies shown in UVJ space in Figure 10. These recovered galaxies are mostly quiescent, lower-mass galaxies, which is the region that we are missing the most galaxies from. At this rate, we expect to find redshifts to 60 more objects, which will reduce the bias towards star-forming galaxies.

5. CONCLUSION

We intend to measure the star formation rates of low-mass galaxies with $0.3 \leq z \leq 0.4$ in the COSMOS field.

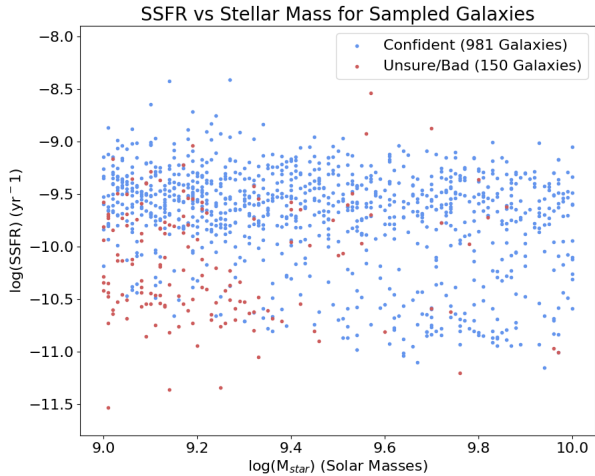


Figure 8. Specific Star Formation Rate (SFR/Stellar Mass) vs Stellar Mass, showing where our “unsure” galaxies fall. They are clustered in the region of low SSFR and low mass, as these galaxies have the lowest signal-to-noise, making them difficult to measure. This introduces a measurement bias into our sample.

Our sample of 1131 galaxies was chosen randomly from galaxies in our mass and redshift range from the ULTRAVista catalog (Muzzin et al. 2013). After extracting the spectra and correcting for atmospheric absorption, we have confidently measured spectroscopic redshifts to 981 of the 1131 galaxies from our sample using cross-correlation. Our uncertainties appear to be on the order of 80 km/s, giving enough detail to map out the structure of the universe in this region. The galaxies that still have

“unsure” redshifts are heavily biased towards low-mass, quiescent galaxies.

In future work, we intend to recover as many of the low-mass, quiescent galaxies as possible to remove our measurement bias. Then, we will measure star formation rates for each galaxy using Balmer emission, and compare these results to the predictions of (Kelson 2014), (Kelson, Benson, & Abramson 2016), and (El-Badry et al. 2016).

6. ACKNOWLEDGEMENTS

The author thanks the Rose Hills Foundation and Carnegie Observatories for funding, Dan Kelson for overseeing the project and invaluable advice and guidance, Shannon Patel for suggestions during data processing, and Gwen Rudie for organizing a fantastic summer program at Carnegie Observatories.

REFERENCES

- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Dressler, A., Bigelow, B., & Hare, T. et al. 2011, *PASP*, 123, 288
- El-Badry K., Wetzel A., Geha M. et al. 2016, *ApJ*, 820, 131
- Hopkins, P. F., Keres, D., Onorbe, J., et al. 2014, *MNRAS*, 445, 581
- Kelson D. D., Illingworth G. D., van Dokkum P. G. & Franx M. 2000, *ApJ*, 531, 159
- Kelson D. D. 2003 *PASP*, 115, 688
- Kelson, D. D., Martini, P., & Mulchaey, J. S. 2005, <http://code.obs.carnegiescience.edu/Algorithms/realcc>
- Kelson, D. D. 2014, *arXiv:1406.5191*
- Kelson D. D., Benson A. J. and Abramson L. E. 2016 *arXiv:1610.06566*
- Muzzin, A. et al. 2013, *ApJS*, 206, 8
- Scoville N., Abraham R. G., Aussel H. et al. 2007, *ApJS*, 172, 38

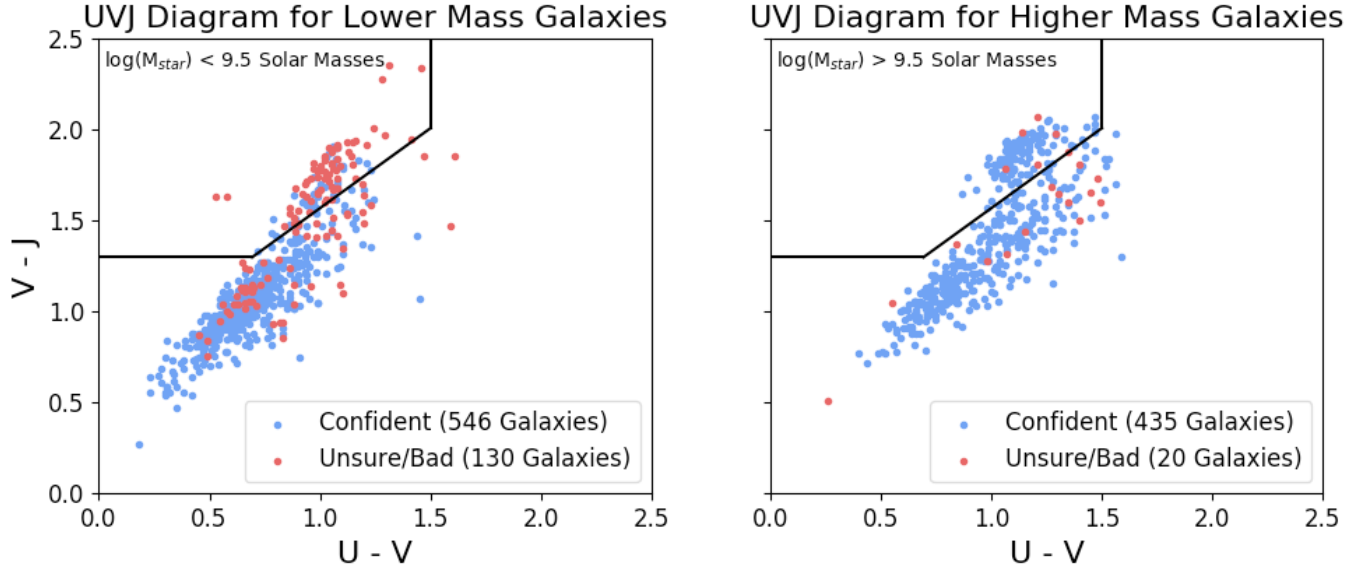


Figure 9. Another view of the population of galaxies that are “unsure,” as in Figure 8. The black lines divide the galaxies into quiescent (upper left) and star-forming (below) (Muzzin et al. 2013). When $\log(M_{\text{star}}) > 9.5$ solar masses (right panel), we measure a fairly representative fraction of the galaxies. However, at lower mass (left panel), we see that the majority of “unsure” galaxies are biased towards the quiescent region.

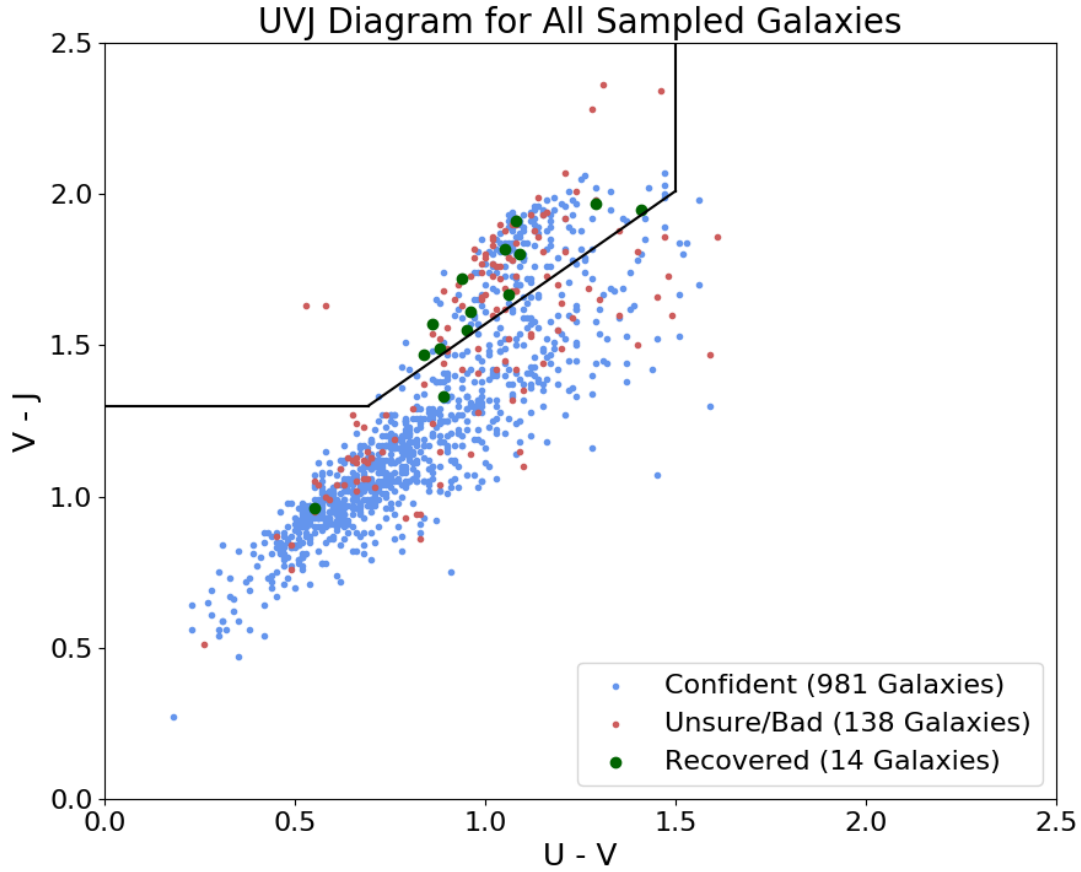


Figure 10. UVJ diagram for all objects, highlighting the 14 that were recovered by closer inspection of spectral features. These were mostly low-mass, quiescent galaxies, helping to remove some bias towards the brighter star-forming galaxies. We expect to recover roughly 60 more objects when going through the remaining eight observations.