Draft version December 13, 2022
Typeset using LaTeX default style in AASTeX631

# Expanding Photometric Redshift Calculations for High Redshift EELGs

Kelcey Davis, Rachel Lee

## ABSTRACT

We will expand an existing search in James Webb Space Telescope (JWST) data for Extreme Emission Line Galaxies (EELGs) at high redshift, 4 < z < 12, by taking a closer look at SED fits and their associated photometric redshifts. This will include investigating the accuracy of existing redshifts by considering the flux effects per filter from emission line presence.

This project will estimate photometric redshifts for EELGs in the Cosmic Evolution Early Release Science Survey (CEERS) field, including a sample of the most probable redshifts for candidates with multiple possible solutions. This research is incredibly important as it allows for a robust statistical analysis of high-redshift EELGs, which have the potential to inform us of the properties of the early universe.

## 1. SCIENTIFIC BACKGROUND

The James Webb Space Telescope (JWST) introduces an exciting new pathway to exploring the early universe. This is because JWST is designed to observe the universe with a broader wavelength region than ever before. JWST has the capability to observe light to almost the far-infrared regime, whereas older telescopes, such as Hubble, only go to near-infrared. This is incredibly important when looking further and further into the universe as infrared wavelengths can penetrate through dense gas and dust, allowing us to see more than we have before.

With these new capabilities, not only can we observe more objects, but we can observe objects that are further away, and therefore have higher redshifts. These objects are much further and much older than anything we have observed before, allowing us to peer into the earlier stages of our universe. An exciting example of some of these objects are Extreme Emission Line Galaxies (EELGs) at high redshifts. What makes these galaxies different is their extreme emission features, which are consistent with potential star formation. With that being said, these galaxies are thought to be high star forming regions, giving us insight into early galaxy and star formation. With JWST, observing high redshift EELGs will be much easier than it was previously.

When determining the redshift of an object, there are two primary methods. The first is using spectroscopy and the second is using photometry. Spectroscopic redshifts are based on changes in spectroscopic observations of objects. We know the wavelengths of absorption and emission features of different elements, and when we notice shifts or broadening of these features, we can determine the redshift of the object. Since this method requires a large amount of data points, it is more computationally expensive and time consuming. Photometric redshift, on the other hand, looks at Spectral Energy Distribution (SED) plots, which measure the energy of light in broad wavelength regions. These redshift values are then determined by $\chi^2$ squared fitting of real and template SED plots. We are interested in investigating the photometic redshift of EELGs because JWST does not have high confidence spectroscopic data of these objects yet so obtaining accurate photometric redshift values is important to investigating EELGs.

## 2. RESEARCH OBJECTIVES

We are interested in investigating how we can change photometric redshift calculations for high redshift sources as a way to improve accuracy and confidence in estimates. The current methodology is to look at SED plots of sources artificially placed at z = 0 and then shift the source value to create multiple redshift templates. These templates are then compared to the real SED plot and the redshift is determined by the best fit. This method has been sufficient when utilizing older telescopes, but JWST does not provide high confidence fitting with higher redshift sources.

Our methodology will be to utilize existing SED templates from Astropy's Eazy library in addition to templates published in Larson et al. (2022). We will then expand these templates to contain values of 0 < z < 12 to include the redshift values we are interested in. With these templates, we will create a grid of redshift values in which we can compare real and generated SED plots for $\chi^2$ squared values. These $\chi^2$ squared values will tell us how our template is doing in comparison to the published redshift values of our sources.

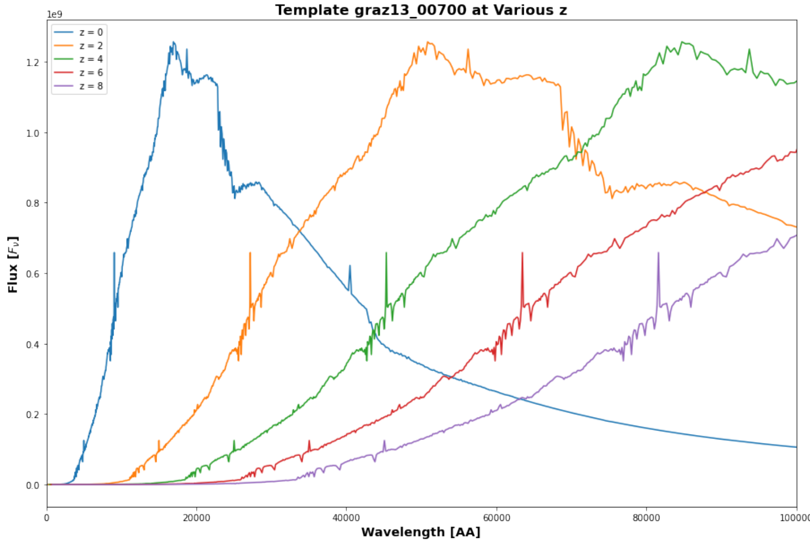


**Figure 1.** Template spectra redshifted to a few different z values.

We plan to specifically investigate EELGs to see how SED fitting works for high redshift sources in comparison to objects with lower redshift values. Once confidence in our fitting method is established, this method can then be applied to other types of sources and sources of a multitude of redshifts ranges.

## 3. COMPUTATIONAL METHODOLOGY

We begin our computational work by first answering a basic question: What is an EELG? EELGs have been studied by other telescopes at lower redshifts. They have previously been studied up to a redshift of z = 1.7 (van der Wel et al. 2011). With the capabilities of James Webb, we ambitiously attempt to pull these interesting galaxies out of deep field data, targeting a redshift of 5 < z < 9. EELGs by their very nature have extreme emission features. We guess that, at these higher redshifts, the emission features for these galaxies would be most prominent in the three reddest JWST filters.

We then selected our potential EELGs first by using a color-color diagram. We selected them to be extreme in the F356 - F410 color or the F410-F444 color. Sources that were extreme in either of these were then selected for equivalent width. Equivalent width (EW) is a way to measure how peaked a particular feature is. These can be calculated in the rest frame, which requires an accurate redshift calculation, or the observed frame, which does not. We then select the sources to have an observed frame equivalent width greater than 5,000 Å in one of the three reddest filters. We take this remaining population of about 600 sources and further cut them down to 182 ideal targets. These cuts are further detailed in the jupyter notebook.

Once the target sources are identified, it is simple to read in the filter data for each source as an SED. We can then begin to consider how best to estimate the redshfit of our target sources. Most SED photometric redshift fitting code follows the same basic format. We take template spectra which have been arbitrarily developed with no flux units at a redshift of 0. This is a look at what a spectrum of some type of source might look like at no redshift. This allows us to scale the spectra to reflect what that type of source looks like at another redshift. This is a fairly simple calculation. We multiply the flux values in the template by (1+z) where z is the value to which we would like to redshift the spectrum.

We used templates from Astropy's Eazy photometric redshift fitting library. This is a popular photometric redshift fitting code from Brammer et al. (2010). We also utilized newly published templates intended to work better on z > 7 galaxies (Larson et al. 2022). We then created a grid of these templates, redshifted in increments of z = 0.1, from a redshift of z = 0 to z = 10. After finalizing the code, we later identified several sources who were converging at z = 9.9 and hypothesized that expanding the grid could help mitigate this. After expanding the upper limit to z = 12, all sources seems to converge to a redshift value and none were reporting values of 11.9 or 12.

From here, we needed to find a way to determine how well each template fits our observed SEDs. We utilize a $\chi^2$ fit. Our first task was to think about how our error values would affect the $\chi^2$ fit. We want to ensure that our

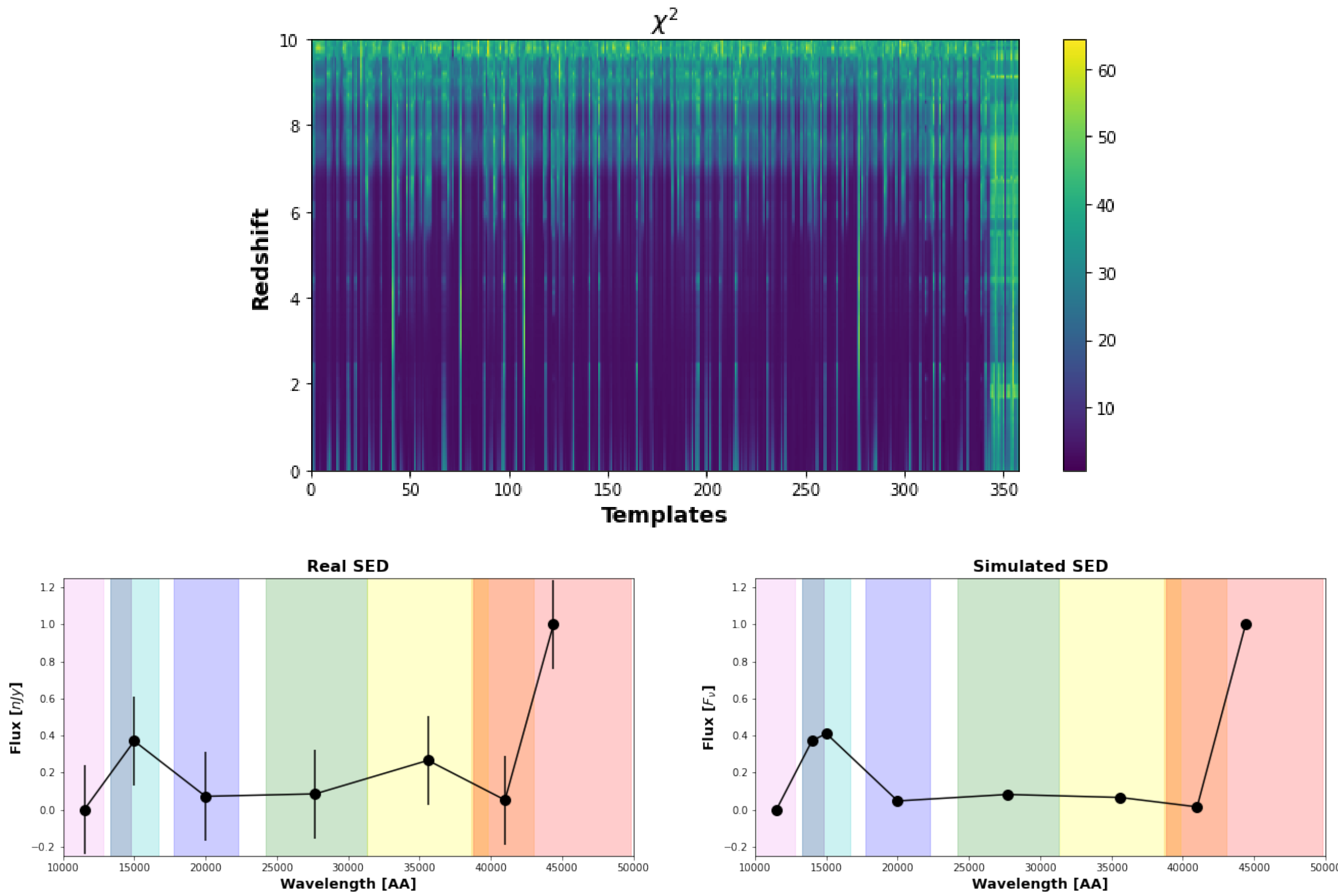


**Figure 2.** Top panel: $\chi^2$ values for each of the SED templates in the matrix. Bottom left panel: Real SED with adjusted error floors. Bottom right panel: Best fit simulated SED from Grid

error values are not under-reported because this would make the fit appear to be better than it really is. We fixed an arbitrary error floor at 10 nJy and any errors that were reported under this value were increased to it. We believe this may be the source of some of the potential over-fitting or the reason some sources have difficulty converging. A good solution would be to introduce a dynamic error floor, potentially something that scales with the reported brightness of a source. This is an area to explore in future work.

To calculate the $\chi^2$ values, we loop through every redshifted spectra in our grid of spectra. We then apply our JWST filters on the spectra to see what it would look like as an SED plot. We create a second grid containing these artificial SED plots. We then also need to deal with the flux units of the templates and the observed sources. Because the templates have arbitrary units, we can best compare two sources if we simply scale them both to have the same flux range. We scale both the template SEDs and the observed source SED between 0 and 1. We then compare each of the individual templates in the grid of SEDs to a given source by calculating it's $\chi^2$ value as:

$$\chi^2_{z,i} = \sum_{j=1}^{N_{filt}} \frac{(T_{zij} - F_j)^2}{\delta F_j^2} \tag{1}$$

Where $N_{filt}$ is the number of filters, T is Synthetic flux of template i in filter j for redshift z, $F_j$ is observed flux in filter, and $\delta F_j$ is observed error.

We can find the best fit where our value of $\chi^2$ is minimized. We can also explore how well these calculations converge. For the several sample sources we looked at, we considered their 10 best fit $\chi^2$ values and their associated redshifts. This helped us understand how well they converged, and more detail about this can be found in our jupyter notebook.

The sample source illustrated in Figure 2 converged to a redshift of z = 11.1. It's 10 most likely redshifts were all associated with the same template, however it was not a template from our 2022 sample. The lowest $\chi^2$ value for this

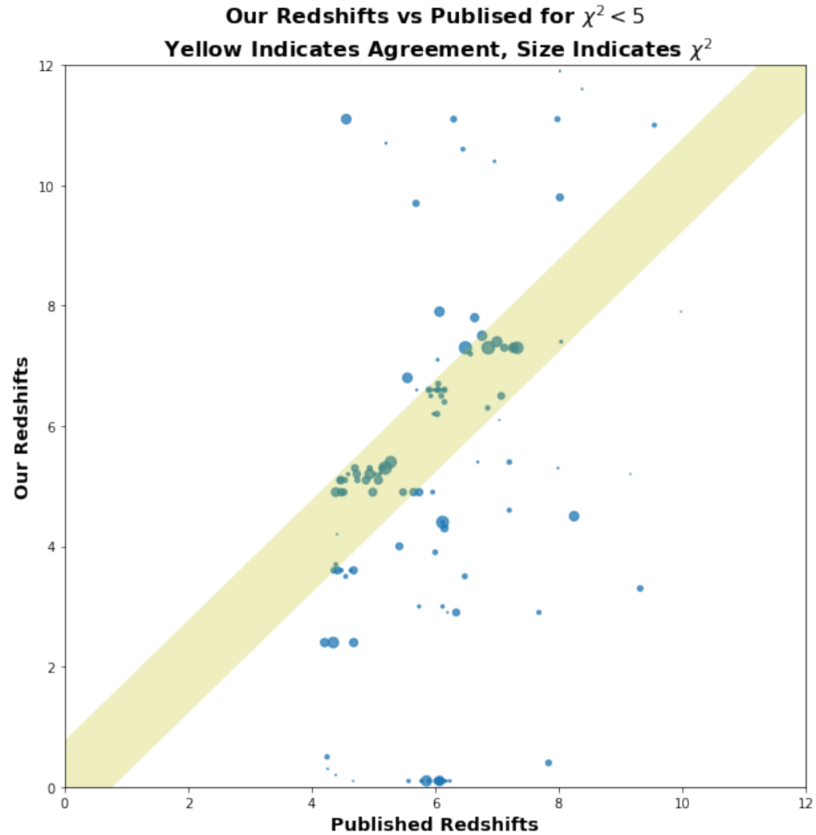


**Figure 3.** Published Redshift values versus our reported redshift values.

source was approximately 0.7. With some preliminary results, we can compare the results of this code run for the full 182 sources to their published redshift values.

One important observation is that the published values do not report redshifts above a value of z = 10. Consistently, extremely high redshift sources have been missed in JWST because it has been assumed that redshfits don't make sense to report at a certain range, and this could be a case of the published values having an arbitrary ceiling at z= 10. This is interesting because many of our z > 10 candidates have relatively small $\chi^2$ values. The sources also seem to cluster along the yellow line, indicating strong agreement to published results at $4 < z < 8$.

## 4. PLAN OF ACHIEVEMENT

We ambitiously originally proposed to run this for the 40,000 sources in the CEERS point of view, but this was ultimately too much for this project. Instead, we paralyzed the code and ran it for the full 182 potential EELG sources. We ran this on a personal computer rather than the BRIDGES-2 cluster. The problem was computationally intensive enough to require parallelization to execute in full, but not so computationally intensive that we needed to use the super-computing cluster. We ran the sources on 16 CPU threads. To accomplish this, we ran the code on 182 sources in parallel across the 16 CPU threads.

## 5. PERFORMANCE

We first ran the code for a single source, increasing the number of threads used, illustrated in Figure 4. We can see that the computation time steadily increases as the number of CPU threads increases, which makes sense. A single source is not very intensive to compute, and spreading it out across multiple CPU threads is not an efficient task. We instead need to run the code simultaneously for all of our 182 sources. Doing this, we see in Figure 5, that the computation time is initially high with the lower number of sources and then decreases with the higher number of sources running in parallel.

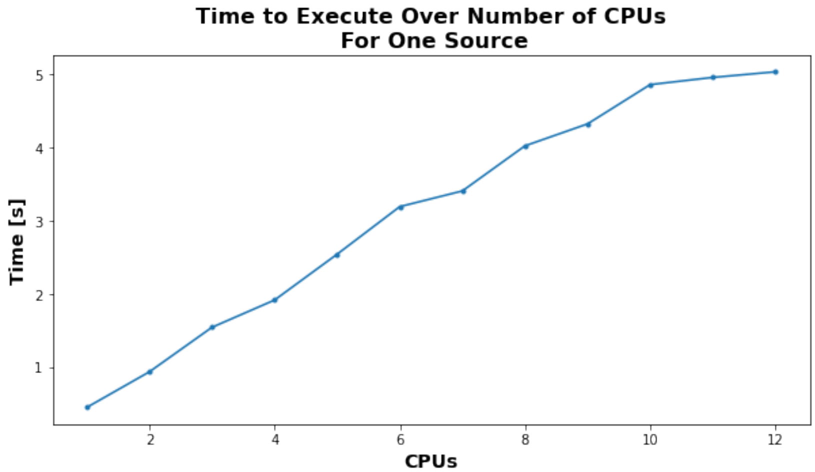


**Figure 4.** Computation time for a single source over an increasing number of CPU threads.

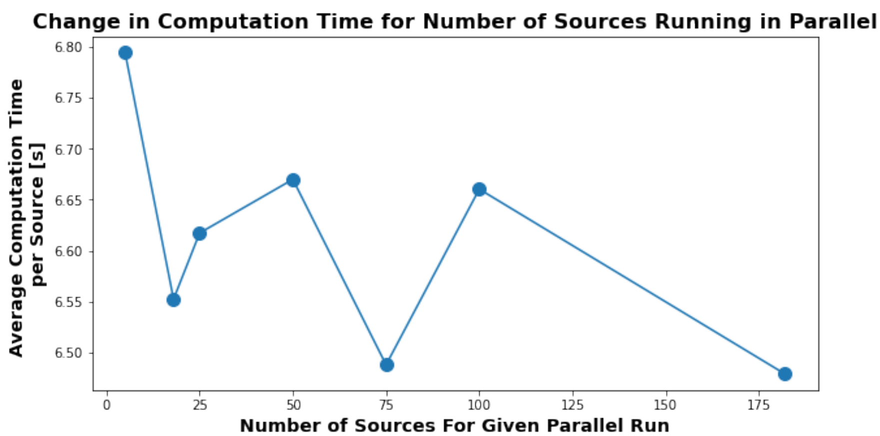


**Figure 5.** Average computation time per source for an increasing number of sources running parallel on 16 CPU threads.

## REFERENCES


Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2010, EAZY: A Fast, Public Photometric Redshift Code, Astrophysics Source Code Library, record ascl:1010.052. http://ascl.net/1010.052

Larson, R. L., Hutchison, T. A., Bagley, M., et al. 2022, arXiv e-prints, arXiv:2211.10035. https://arxiv.org/abs/2211.10035

van der Wel, A., Straughn, A. N., Rix, H. W., et al. 2011, ApJ, 742, 111, doi: 10.1088/0004-637X/742/2/111