Identifying Active Galactic Nuclei from Emission-Line Spectroscopy

Tyler Barna` Anna Acuna-Gillard Yasmine Brahmia Mohamed Elashaky

May 04, 2020

1 Introduction

From studying emission lines from light sources, researchers can collect a vast amount of data. This principle is applicable over all distances,and is especially important in astronomy. As light passes through gas, atoms are ionized by oncoming photons. The ionization level observed from the cloud reveals several physical and chemical properties, as well as the source that ionized it . From the data provided, active galactic nuclei were isolated from stars ionizing gas clouds. This study was broken into two parts.

The first part of this investigation involved studying emission line ratios. The insight from the referenced paper (Baldwin et al. [1981]) provided classification when plotting ratios [O III] λ5007/Hβ vs. [N II] λ6583/Hα, also referred to as a BPT diagram. From representing that data this way, the data was then classifiable by what ionized the gas cloud.

The second part saw the analysis of line widths in [N II] and Hα emissions. Broad velocity components of Hα lines are a result of gas deep in a super-massive black hole's potential well are moving at high velocities ($\Delta v \geq 1000\frac{km}{s}$). This phenomena is unique to this case, and can be used to identify active galactic nuclei. Emission lines from brighter stars in the galaxy are subtracted from the data in order to determine which galaxies demonstrate broad velocities. Existence of broad emission lines is finally used to find correlations between width and luminosity.

2 Methodology

2.1 Emission Line Ratios

For the first part of this project a full sample of 418 galaxies was analyzed and the goal was to identify candidate Active Galactic Nuclei, hereafter referred to as AGN, on the basis of emission line ratios. Out of the 418 sample of galaxies, 179 galaxies had substantial amounts of uncertainty in their flux, so galaxies above a 30% uncertainty were omitted. After narrowing down the sample size, the BPT diagram for [O III] λ5007/Hβ vs. [N II] λ6583/Hα was created. The sample did not include the values for the ratio of [O III] λ5007/Hβ, but it did include the ratio of Hβ/ Hα. Dividing the flux ratio present for [O III] λ5007/Hα by the ratio of Hβ/Hα yields the value for [O III] λ5007/Hβ so that it could be used. From there, the ratios [O III] λ5007/Hβ and [N II] λ6583/Hα were plotted on a log scale to get the BPT diagram seen in figure 1.

2.1.1 Galaxy Classification

For the galaxy classification, areas on the BPT diagram correspond to different types of galaxies and AGN. In order to better understand what was being looked at, a defining the boundary line between normal HII regions, gas ionized by normal galactic O and stars was plotted. According to Baldwin and Phillips, you can fit a simple HII region boundary line onto the BPT diagram by using equation 1

$$(\lambda 5007/\lambda 4861) = log 4.2 - 9.4 I([NII]\lambda 6583)/(IH\alpha\lambda 6563)] \quad (1)$$

The values below this equation are galaxies being ionized by young stars and galaxies below are likely to be AGN. For plotting, Excel was used in order to produce the BPT diagram in figure 1 along with the HII boundary line

described above. Excel was also used in order to sort the galaxies and identify the ones above and below the HII boundary line. The combination of the location of the galaxy on the diagram and its location above or below the line was used in order to determine which galaxies had gas mainly being ionized by young stars, those whose gas is mainly being ionized by AGN, and those whose gas is being ionized by a combination of both.

2.1.2 Correlation To Other Values

After identifying which galaxies had gas mainly being ionized by different sources, other relationships based on the data were investigated. The type of galaxy was compared to the values [OIII]λ5007/[OI]λ6300 emission line ratio, which is a measure of ionization state; the Hα/Hβ emission line ratio, which is a measure of dust content; the Hα luminosity; the Hα equivalent width (i.e., line-to-continuum ratio); the [NII]λ6583 velocity width. This was done by graphing the value investigated on the y axis and seeing how the type of galaxies were distributed amongst those values. The graphs seen in Figures 1 through 6 were produced using this process.

2.2 Emission Line Widths

Emission line spectroscopy in the optical wavelength range has been an oft-studied discipline in astronomy for more than a century. There is an abundance of information to be gleaned from spectra, especially because they can be used to study phenomena that can't be resolved. Active galactic nuclei represent one such case – they're relatively small, much less than a parsec in radius, so they can't be spatially resolved (Stirpe [1990]). However, the inordinate luminosity generated by active galactic nuclei makes their presence apparent in spectroscopic analysis of a galaxy, most notably in the form of a broad velocity Hα emission line. As a result, observation of a broad velocity component in the Hα emission line strongly suggests the presence of an active galactic nucleus. This project focused on ten pairs of galaxies and sought to identify correlations between the width of broad velocity Hα components and their luminosity.

Ten pairs of spectra of emission line galaxies have been analysed alongside the spectra of their partnered absorption line galaxies. The goal was to scale said absorption line spectra to match the emission line spectra; subtracting the scaled absorption line spectra produced a star-free emission line spectrum for each galaxy. The flux density of these star-free emission line spectra can be decomposed into Gaussian components and analysed to find various properties of the galaxies, such as whether a galaxy might have an active galactic nucleus.

2.2.1 Data Standardisation

As originally provided, the data was not conducive to immediate analysis. This was a result of two attributes – most of the paired spectra had wavelength ranges that varied between them, and the wavelengths at which they were measured was not consistent. To standardise the data, it was trimmed and interpolated so the paired spectra were sampled at the same wavelengths. This standardisation was done so that the absorption and emission line spectra could be more directly compared, allowing for the absorption line scaling to be more easily determined.

Each of spectra were imported from their original `.txt` files into unique `pandas` data frames using `python 3.8`[1]. A generalized function was written to compare the wavelength ranges of paired absorption and emission line spectra. It identified the minimum and maximum integer wavelengths at which the absorption and emission line spectra both had data and, using `pandas.query()`, filtered values outside of this range before returning these 'trimmed' data frames. The results of this can be seen in Figures 7 through 16, where the dashed lines represent the trimmed wavelength ranges. While many of the spectra pairs had roughly the same wavelengths, resulting in few data points being lost, a handful, such as Figure 8 and Figure 13, had upwards of 20 percent of data points lost, though they still had several hundred data points[2].

To standardise the data even further, a function was written to interpolate the spectra so the flux density was measured at the same points for both the emission and absorption line spectra of each galaxy pair. In the process of the correction to their rest wavelength, most of the data points shifted in a way so that the interval between each measured wavelength was not consistent. The function uses the wavelength range found when trimming the spectra with the `numpy.linspace()` function to generate a list of integer wavelengths to interpolate the spectra to. Interpolation is a method of estimation that uses local values of a data set to approximate the value of data at a

[1] The code for this portion of the project is maintained in a GitHub repository located at this link.

[2] Ultimately, these edge wavelengths are not relevant to the analysis and don't represent the loss of any particularly useful data.

point it was not originally measured at. `numpy` provides a method of interpolation in the `numpy.interpolate()` function, which is compatible with `pandas` data frames. Applying `numpy.interpolate()` to the spectra with the integer wavelength range as the new index, the absorption and emission line spectra of each galaxy pair are effectively re-sampled so the flux densities of the spectra can be directly compared for a given wavelength. The results of this can be seen in Figures 17 through 26. Looking at these Figures, one can see that the data wasn't altered in any meaningful sense by the interpolation. This is because, while they were not actually sampled at integer rest wavelengths, the measured wavelength values were usually fairly close to an integer, often only a tenth of an angstrom or so away. Because the measured wavelengths were so close to the integer wavelengths, the interpolation was able to estimate a fairly accurate flux density for each point[3].

2.2.2 Absorption Line Scaling

The next step in the process was to determine a scale factor to apply to the absorption line so that it matched the emission line as closely as possible. After subtracting a scaled absorption line spectrum from the associated emission line spectrum, the stellar contributions to the emission line spectrum should be removed. While the data standardisation outlined above made finding the appropriate scale factor for each galaxy pair more straightforward, there were still several considerations to be made so the most accurate star-free emission lines could be found.

Given that the absorption and emission lines are sampled for the same values, an effective method of determining the accuracy of an absorption line scale factor is the use of a penalty function.

$$\sigma(x,\lambda) = f(\lambda) - x \cdot g(\lambda) \tag{2}$$

Equation 2 is the penalty function used to find the difference, $\sigma(x,\lambda)$, between the emission line spectrum, $f(\lambda)$, and the scaled absorption line spectrum, $g(\lambda)$, multiplied by scale factor x. For a given value of λ, setting x equal to a more accurate value for the scale factor will result in $\sigma(x,\lambda)$ evaluating to closer to zero than less accurate values.

In order to find the best scale factor for the absorption line, the penalty function needs to be minimised. There are multiple ways to go about this, but this project used `scipy`. The `scipy.optimize` function `least_squares()` solves a bounded non-linear least squares problem. Least squares optimisation finds a solution such that the sum of the squares of the values of the function in a given range are minimised.

$$\sigma(x) = \sum_{\lambda_0}^{\lambda_n} (\sigma(x,\lambda))^2 \tag{3}$$

In this case, `least_squares()` would find a value for the scale factor x that minimises the sum of squares of the penalty function across the wavelength range for a given pair of spectra. For a normal least squares optimisation, `least_squares()` would attempt to find the value of x for which Equation 3 is minimised, where λ_0 and λ_n are the minimum and maximum values for the wavelength, respectively.

However, spectra have one quality that makes finding an accurate scale factor using a standard least squares implementation difficult in some cases. For spectra, emission lines are characterized by peaks in the data with large magnitudes but comparatively narrow widths. This feature is problematic when trying to find an optimal scale factor using the least squares method, as the peaks in the data can have a disproportionate effect on the sum of squares and bias the algorithm towards a value that minimises just the peaks rather than the data points with magnitudes closer to the median. Fortunately, mathematicians have developed ways to adjust least squares algorithms to account for outliers in the data and `least_squares()` allows for the usage of several of these by including optional arguments.

$$\rho(z) = z \tag{4}$$

$$\rho(z) = 2(\sqrt{1+z} - 1) \tag{5}$$

For this project, the 'robust' least squares implementation uses the `soft` least squares argument for the loss function, which is an approximation of the absolute value loss. Equation 5 is the loss function used when `soft` is used

[3] It's also worth noting many of the spectra were essentially super-sampled: there were more data points in the original set than in the interpolated set so `numpy.interpolate()` was able to re-sample the data more accurately. Another option would have been to identify the spectrum with less data points and then interpolate the other spectrum to match rather than interpolating both spectra

while Equation 4 is the loss function used for the standard least squares implementation[4]. Figures 27 through 36 demonstrate both the standard and robust least squares optimisations for each galaxy pair. In a handful of cases, such as Figures 27, 28, 32, and 35, the difference between the standard and robust least square methods was minimal. However, for many of the galaxy pairs, most notably Figures 29, 31, and 33, the difference is pronounced and the robust least square method finds a much more effective scale factor. The details on the difference between all of the fits are recorded in Tables 2, 3, and 4.

2.2.3 Gaussian Decomposition

Once the star-free emission line spectra were found, the discrete emission lines became much more apparent. Figures 37 through 46 show the spectra isolated to the wavelength range relevant to this project, with dotted and dashed lines showing where $[NII]$ and Hα are present; the star-free emission lines all show clear peaks at or close to these values[5]. One should also note the disparity in the magnitudes of peaks between different spectra; the highest peak in 38 is almost two orders of magnitude greater than the peak value of Figure 45. The consequences of this become apparent when decomposing the spectra.

An effective tool for analysis of spectra is Gaussian decomposition. Gaussian decomposition is a process in which a spectrum is fitted as a sum of multiple Gaussian curves.

$$G(x) = a \cdot e^{-\frac{1}{2}(\frac{x-b}{c})^2} \tag{6}$$

Equation 6 is the general form of a Gaussian function, where a, b, and c are arbitrary constants[6]. Gaussian functions are useful for modeling spectra because a well-fitted sum of Gaussian components can model how the spectra would appear without noise muddying the data (Lindner et al. [2015]). However, obtaining an accurate model for spectra can be difficult given the variety of potential ways in which to model a spectrum; there isn't a definitive answer for the best way of obtaining these results. While there are a number of packages that offer the ability to fit data to Gaussian components, they aren't necessarily designed for analyzing spectra. Fortunately, several packages have been written by astronomers for the express purpose of applying Gaussian decomposition to spectra. One such package is `GaussPy` (Lindner et al. [2015]), which automatically decomposes a given spectrum into Gaussian components based on a few parameters provided as arguments[7]. When `GaussPy` performs a Gaussian decomposition, it returns an array with numeric values for each of the fitted components that can be used to construct the Gaussian curve.

$$H(\lambda) = A \cdot e^{-4\ln 2(\frac{\lambda-\mu}{w})^2} \tag{7}$$

Equation 7 is the Gaussian function used to find the value of a Gaussian component as a function of the wavelength λ. The three constants are the quantities returned when `GaussPy` decomposes a spectrum into Gaussian components: A represents the peak amplitude of the curve, μ is the wavelength at which the peak amplitude is located, and w is the full width at half maximum (FWHM) of the curve.

Though `GaussPy` made Gaussian decomposition more straightforward, there were still several considerations that had to be made in regards to how the spectra were decomposed. The primary arguments used to tweak the decomposition are the α and signal to noise threshold parameters. First, the star-free emission line spectra were isolated to a range between 6400 and 6700 Angstroms. This was done so that `GaussPy` did not attempt to decompose emission lines besides those that were of interest to this project[8]. The α parameters determine how much to smooth the derivative of the curve when decomposing a spectra and the signal to noise threshold ignores values below the supplied ratio when modeling the Gaussian components. `GaussPy` has multiple options for decomposition, including both one and two phase decomposition. Two phase decomposition uses two separate values for α. This is especially useful for analysis such as this, as one value of α can be used to find broad velocity components and another can be used for narrow velocity components. Altering the values of these parameters can have a noticeable effect on how many components are fitted to a spectrum.

[4]The documentation on the `least_squares()` function is available here.

[5]It seems that the wavelengths of some of the data sets were offset ever so slightly, likely while making redshift corrections, so that the peaks of the observed lines aren't exactly at the expected wavelengths. However, these differ by, at most, a few angstroms, and the code was written in such a way that this should not have an effect on the results or analysis.

[6]c must be nonzero

[7]The GitHub repository for `GaussPy` can be found here; the documentation can be found at this link.

[8]Namely: the $[NII]$ line at 6548 Å, the Hα line at 6563 Å, and another $[NII]$ line at 6583 Å

For example, Figures 47 and 58 are plots of the same spectrum decomposed using different smoothing parameters. While Figure 47 may have a slightly worse fit to the data, it represents a more plausible physical situation, as the decomposition in Figure 58 identifies two components for only one of the $[NII]$ lines; it's unlikely for the $[NII]$ lines of a spectrum to demonstrate a different number of components (Kriss [1994]).

One area `GaussPy` struggled with was decomposition of particularly noisy spectra. Figure 55 is an example of a spectrum with a very low signal to noise ratio – it's difficult to differentiate between actual signal peaks and the noise. Unfortunately, in cases such as these, even lowering the signal to noise threshold in `GaussPy`, as shown in Figure 57, didn't seem to result in a more meaningful decomposition.

Emission lines can decompose into multiple Gaussian components, with each component likely indicating a contribution to the emission line from a different part of the galaxy (Stirpe [1990]). These components can reveal information about the makeup of the galaxy. For example, the presence of a broad Hα emission line component strongly suggests that the galaxy may have an active galactic nucleus. A broad velocity Hα is characterized by a width that corresponds to a $\Delta v \geq 1000 \frac{km}{s}$ (Stirpe [1990]). While Δv is not directly provided when decomposing the star-free emission line spectra into their components, there is a useful relation that allows for this to be calculated.

$$\frac{\Delta v_0}{c} = \frac{\Delta v}{v} = \frac{\Delta \lambda}{\lambda} = \frac{\Delta z}{z+1} \tag{8}$$

Equation 8 demonstrates the relationship between the rest frame velocity interval Δv_0 and quantities like the frequency (v), the wavelength (λ), and the redshift (z). Because the data is provided already corrected to the rest wavelength, rearranging part of Equation 8 results in a useful relation.

$$\Delta v_0 = \frac{\Delta \lambda}{\lambda} \cdot c \tag{9}$$

Equation 9 can be used to determine whether a component satisfies the requirement for broad Hα that $\Delta v \geq 1000 \frac{km}{s}$, with $\Delta\lambda$ being the FWHM of the component as determined by `GaussPy`, λ being the wavelength of the emission line (For Hα, $\lambda = 6563$Å), and c being the speed of light in a vacuum.

For the spectra that decomposed in such a way that there was a possible broad velocity component for the Hα line, the broadest Gaussian component centered at or very close to 6563Å was identified and its Δv was calculated to determine if it was a broad velocity component.

$$\mathrm{F}(\lambda) = \int_{\lambda_0}^{\lambda} \Phi(\lambda^{'}) \mathrm{d}\lambda^{'} \tag{10}$$

$$\mathrm{L} = 4\pi d^2 \cdot \mathrm{F}(\lambda) \tag{11}$$

In order to compare the luminosity of broad velocity Hα components to their widths, the total flux of the components had to be calculated. Equation 10 is the mathematical representation of the function written to find the total flux of a Gaussian component, where $\Phi(\lambda^{'})$ is the flux density as a function of wavelength and $(\lambda - \lambda_0)$ is the FWHM of the given Gaussian component. The integral was calculated using the `scipy.integrate` function `quad()`. Equation 11 is the mathematical expression that represents the function used to calculate the luminosity of a given Gaussian component, where d is the distance to the galaxy (listed in Table 1) and $F(\lambda)$ is the Flux calculated in Equation 10. The `units` module of `astropy` was used for these calculations to ensure that the correct unit conversions were applied.

3 Results

3.1 Emission Line Ratios

In the investigations of emission line ratios, it was found that using the HII region boundary line worked to an extent for classifying the source of ionization. The BPT diagram followed the trend of the HII region boundary line. In Figure 1, there is a clearly visible distinction between galaxies that were ionized primarily by young stars compared to galaxies ionized by active galactic nuclei. However, there were some galaxies that stuck out. In Figure 1, there are six galaxies to the left and above the boundary line. The existence of those galaxies in that location may have been due to the boundary line equation used not fitting precisely to the data points. Due to the fact

that there were only six galaxies that fell in this category, it was decided to continue to label these galaxies as being AGN-ionized when investigating correlation with other quantities. There is also one star that did not follow any trend. That galaxy is colored black in Figure 1 to show that it is an outlier from the rest. It was not taken out of the data due to the fact that it did not have significant uncertainty, although the reason for its location is unclear. Another notable feature of the BPT Diagram is the area circled in black. This was where galaxies could have been either ionized by active galactic nuclei or young bright stars.

3.1.1 Ionization State

As shown in Figure 2, there was no correlation between the ionization state and the ionization source.

3.1.2 Dust Content

As shown in Figure 3, there was no correlation between the dust content and the ionization source.

3.1.3 Hα Luminosity

As shown in Figure 4, there was no correlation between the Hα luminosity and the ionization source.

3.1.4 Line-to-Continuum Ratio

As shown in Figure 5, there appeared to be a correlation between the line-to-continuum ratio and the ionization source. According to the figure, galaxies ionized by bright young stars have lower line-to-continuum ratios than AGN-ionized galaxies.

3.1.5 Velocity Width

The strongest correlation found was between the velocity width and the ionization source, as shown in Figure 6. Galaxies ionized by active galactic nuclei generally had much higher velocity widths than star-ionized galaxies.

3.2 Emission Line Widths

For the purposes of this project, each star-free emission line spectrum was decomposed with the smoothing parameters $\alpha_1 = 1E-5$ and $\alpha_2 = 0.8$ and a signal to noise threshold of $snr = 5.0$ to maintain consistency[9]. These decompositions are shown in Figures 47 through 56.

3.2.1 Gaussian Components

Using the above parameters for decomposition, most of the star-free emission line spectra decomposed into plausible Gaussian components. Of the ten spectra analyzed, three decomposed into three narrow velocity components corresponding to each emission line (shown in Figures 47, 48, and 50) and five decomposed into more than three components (shown in Figures 29, 31, 32, 33, and 34). Two of the spectra (Figures 55 and 56) had unexpected behavior and were not included in either category.

3.2.2 Presence of Broad Velocity Hα Gaussian Components

Of the five star-free emission line spectra that had potential broad velocity Hα components, three had velocity intervals that exceeded $1000\frac{km}{s}$, those being the spectra shown in Figures 51, 52, and 53[10]. The precise values for the flux, luminosity, and velocity interval for these candidates can be found in Table 5. A plot of the FWHM versus the luminosity of the broad velocity Hα components is shown in Figure 59.

[9] Experimenting with different values for α_1 and α_2 yielded better possible decompositions in some cases, but many of were unlikely physically or didn't identify any additional Gaussian components.

[10] While Figure 53 has an unlikely difference in number of components for the two $[NII]$ lines for the parameters used here, it was still included in the candidates for broad velocity Hα components, as this spectrum decomposed in a more plausible manner for slightly altered parameters while retaining the same broad velocity Hα component.

One thing to note when considering Figure 59 is that there is some uncertainty inherent to the calculations of the FWHM of any of the Gaussian components. Since the star-free emission line is found by optimising a scale factor, some uncertainty is introduced in the data itself, and using different parameters during the Gaussian decomposition can alter the FWHM of the components. In addition, the data itself is subject to the effects of noise, which can also affect the processing. Considering these factors, the FWHM has a mild degree of uncertainty in its value and shouldn't be taken as exact. That being said, the Gaussian components that were found to be broad velocity Hα components had velocity intervals a fair amount above the threshold, so they likely would still qualify as broad velocity Hα components under a more exacting analytical regime.

4 Conclusions

4.1 Emission Line Ratios

In conclusion, the use of emission line ratios to plot a BPT diagram is useful in order to classify galaxies and learn more about how they are being ionized. Unfortunately, the ratios are not always successful in determining information about particular ionization states, dust content within the galaxy, or Hα luminosity. This was surprising as it was expected that there would have been more correlation between the ionization state and the mechanisms behind how the gas within a galaxy was being ionized. This could be due to the high amounts of uncertainty that existed in the data, or because of how close the wavelengths are on the spectrum. However, emission line ratios are helpful for looking at velocity widths and line-to-continuum patterns.

4.1.1 Areas for Future Study

This project has shown a lot of opportunity for examining what causes the lack of correlation between the emission line ratios and ionization states, dust content within the galaxy, or Hα luminosity. This project also shows opportunity for further examining galaxies that fall into a region where they could be excited by either AGN or young stars.

4.2 Emission Line Widths

Analysis of the Gaussian decomposition star-free emission line spectra provides some insights into the galaxies the emission lines are from. Before exploring this, however, it's worth discussing why Figures 55 and 56 were considered not suitable for further analysis, as they provide some instructive lessons for future work in this area. Figure 55 corresponds to a noticeably noisy spectrum; it has the lowest signal to noise ratio of any of the star-free emission line spectra. For cases such as these, `GaussPy` struggled with accurate identification of emission lines. Since `GaussPy` smooths the derivative less for smaller values of α, logic would dictate that lowering the smoothing parameters would allow for a more accurate fit. However, even when the smoothing parameters were altered, there was a point at which `GaussPy` was unable to provide a more complex fit, and lowering the values had no effect. This represents a challenge with `GaussPy` and would be worth exploring further. In the case of Figure 56, there was an underlying Gaussian component detected for almost all tested values of α that didn't correspond to any emission line studied. It's possible that this is due to a less-than-optimal scale factor for that spectrum, which would mean not all of the contributions from the stars in the galaxy were removed.

As for the star-free emission line spectra that decomposed in such a way that broad velocity Hα components were found, they provide interesting, if inconclusive, results. From Figure 59, it's apparent that there is a possible positive correlation between the width of these components and their luminosity. This is fairly intuitive if one understands these broad components to be caused by active galactic nuclei. If the source is, say, gas orbiting a black hole, stronger emission lines could suggest the gas is more photoionized, though there are a variety of plausible models (Baldwin [1997]). The luminosity values of these components also roughly match those of many active galactic nuclei (LaMassa et al. [2010]), strengthening the possibility that these are correctly identified broad velocity Hα components. While three data points are not enough to claim this correlation is definitive, it serves as evidence that this method could be a useful means of identifying active galactic nuclei of other galaxies and could be expanded to analyse other galaxies.

4.2.1 Areas for Future Study

On the topic of expansion, there are several ways in which this project could be explored further to derive even more meaningful results. Within the current bounds of the project, it would be beneficial to calculate some numerical uncertainty that propagates throughout the various manipulations of the data, if only to quantitatively identify any areas in which the uncertainty could be minimized. As mentioned previously, it's possible that the scale factor optimisation may be a source of error when isolating the star-free emission line spectra. There are other, more advanced, means of minimizing functions that could be explored[11]. These could potentially allow for spectra with inconclusive decompositions to be analysed, and spectra that already demonstrate possible broad velocity Hα components could be made more exact.

To a similar end, the means in which the star-free emission line spectra are decomposed could be expanded upon. `GaussPy` is a fairly robust library that offers several options for Gaussian decomposition. One feature that wasn't utilized for this project is algorithmically determining the values of the α parameters by using a synthetic or known training set. This wasn't used because of the relatively low number of spectra that had to be decomposed; the training program was computationally expensive and caused some hardware issues[12]. If more galaxies were to be analysed, it would be beneficial to implement this so that the decomposition of each spectra is not limited to a single set of α parameters. `GaussPy` isn't the only library that offers the ability to decompose spectra, though. `GaussPy+`[13] (Riener et al. [2019]) is a library that builds upon `GaussPy` and reduces the potential for experimental bias by using machine-learning to determine the best fitting parameters for a spectrum without relying on initial guesses[14]. This library is likely more advanced than the scope of this project requires, but it would be a useful tool for a broader survey of galaxies. There are doubtless other libraries for Gaussian decomposition of spectra that may offer other utilities for determining precise decompositions[15]. This project served as an informative exploration of the information that can be derived from decomposing emission lines into Gaussian components, and further exploration with a larger data set would potentially reinforce the conclusions suggested by the results outlined in this paper.

[11] `scipy` has several more advanced options for least squares optimisation that could be compared.

[12] The source of the issue seems to be the use of the `AVX` instruction set at some point in the algorithm; `AVX` is a powerful, albeit demanding, instruction set. The hardware issues likely stemmed from `GaussPy` attempting to use 80% of each available core during the training algorithm. This resource usage could be tweaked and, for a larger project, remote computing on a more powerful system is an option that would be worth exploring.

[13] Referred to interchangeably as `GaussPyPlus`

[14] `GaussPyPlus` is maintained in a GitHub repository that can be found here.

[15] Attempts to automate the Gaussian decomposition of spectra aren't even particularly new; Kriss [1994] outlines a method that was developed in the mid 1990s.

5 Figures

5.1 Emission Line Ratios

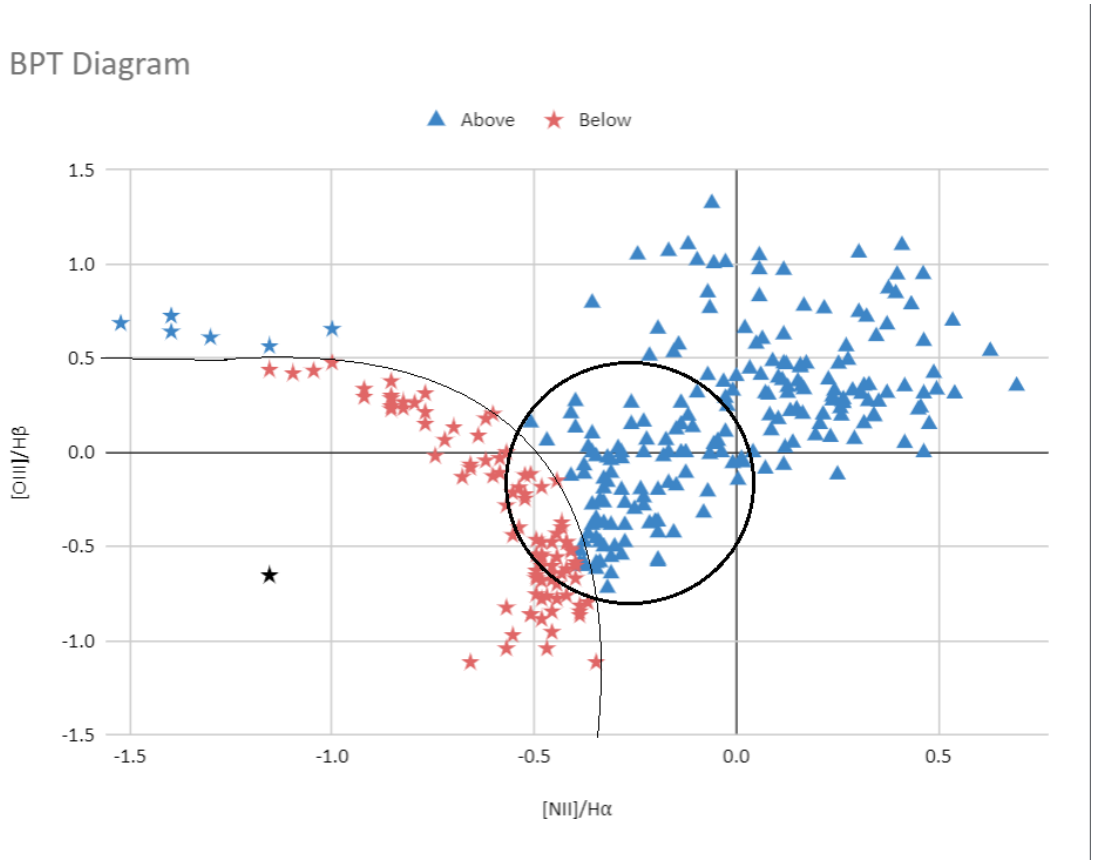

Figure 1: BPT Diagram categorized by ionization source, using HIII region boundary line. All galaxies that fall above the boundary line are categorized as being ionized by an active galactic nuclei (AGN) and are represented by a blue triangle. Galaxies below the boundary line are categorized as being ionized by a young bright star and are represented by a pink star.

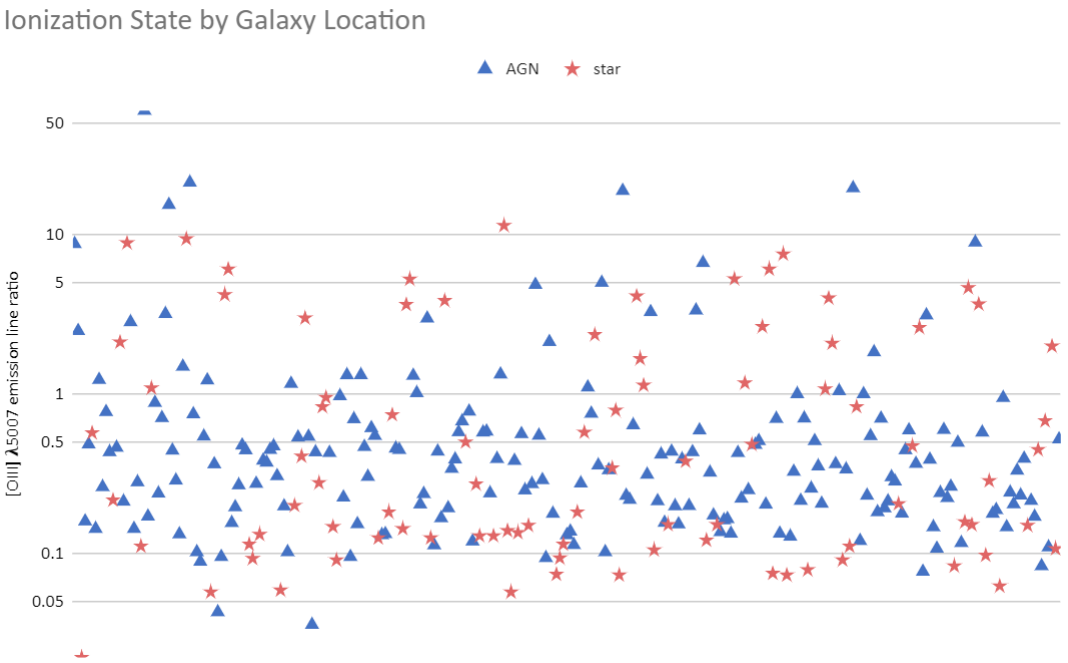

Figure 2: Scatter plot of ionization state, categorized by galaxy location. Ionization state found using O[III] emission line ratio of each galaxy. Classifications used are the same as in Figure 1.

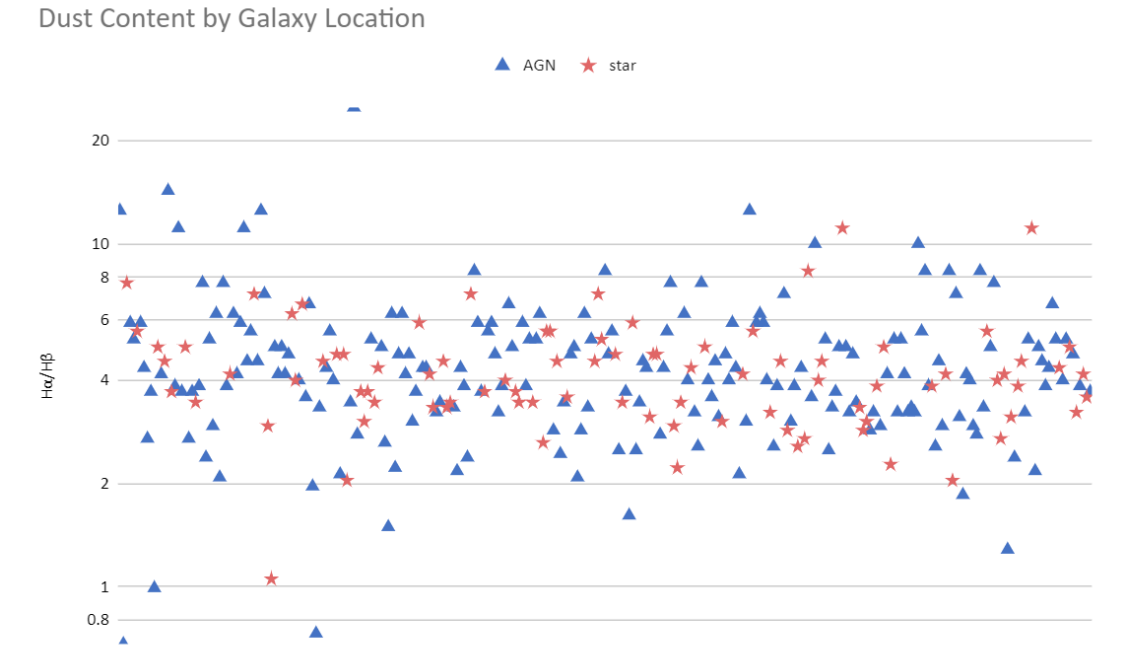

Figure 3: Scatter plot of dust content, categorized by galaxy location. Dust content found using Hα/Hβ of each galaxy. Classifications used are the same as in Figure 1.

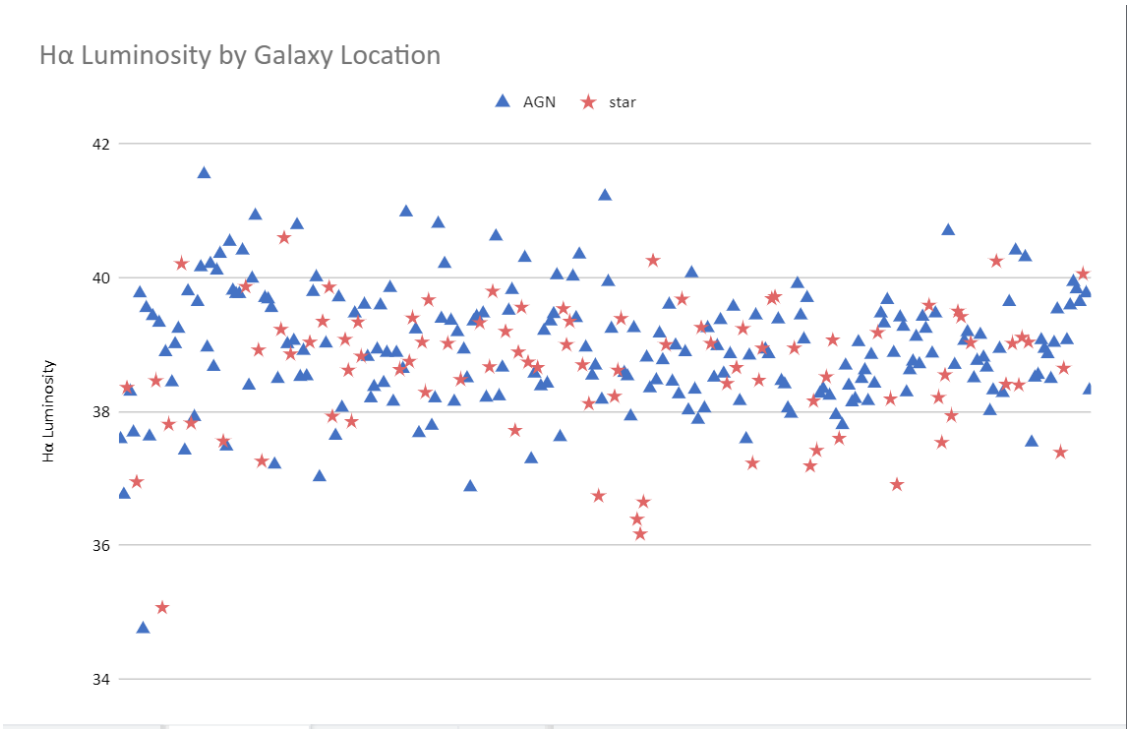

Figure 4: Scatter plot of Hα luminosity, categorized by galaxy location. Classifications used are the same as in Figure 1.

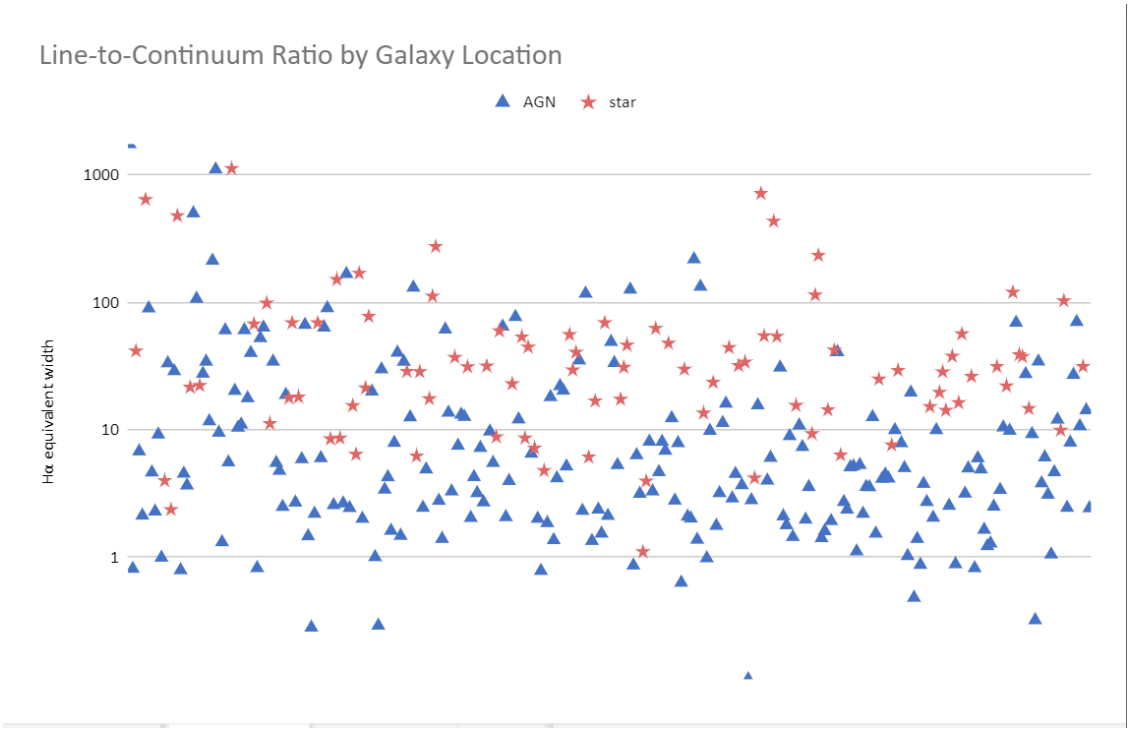

Figure 5: Scatter plot of line-to-continuum, categorized by galaxy location. Line-to-continuum found using Hα equivalent width of each galaxy. Classifications used are the same as in Figure 1.

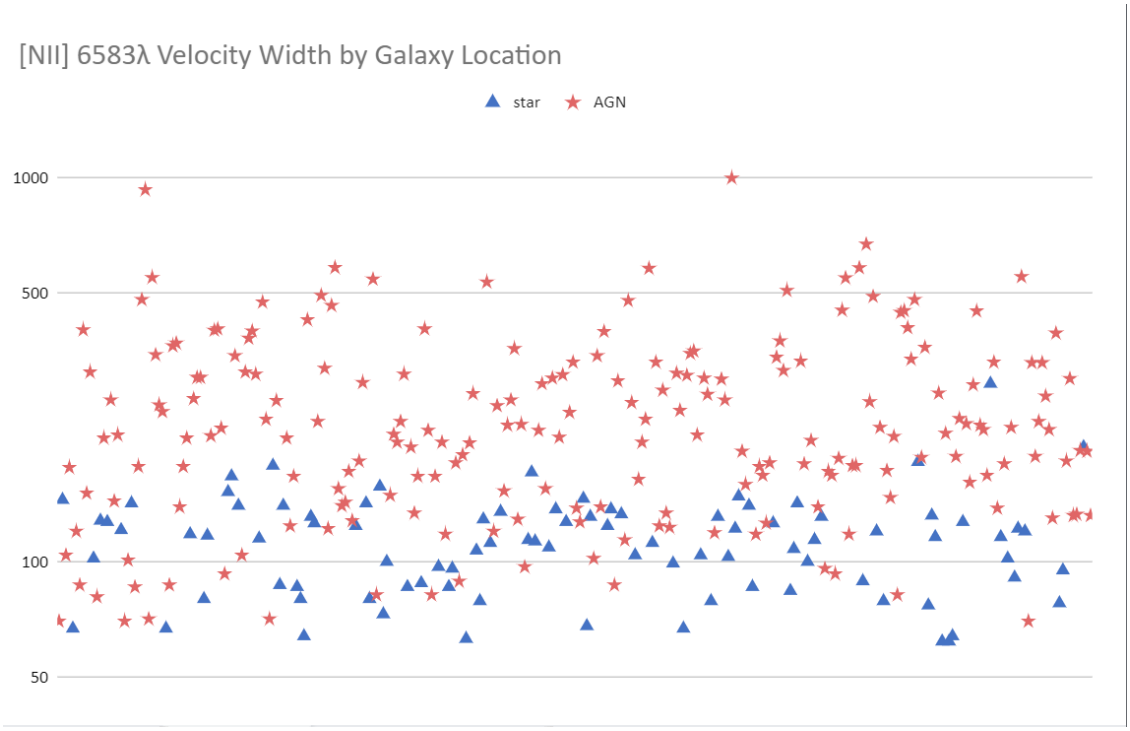

Figure 6: Scatter plot of velocity width, categorized by galaxy location. Classifications used are the same as in Figure 1.

5.2 Emission Line Widths

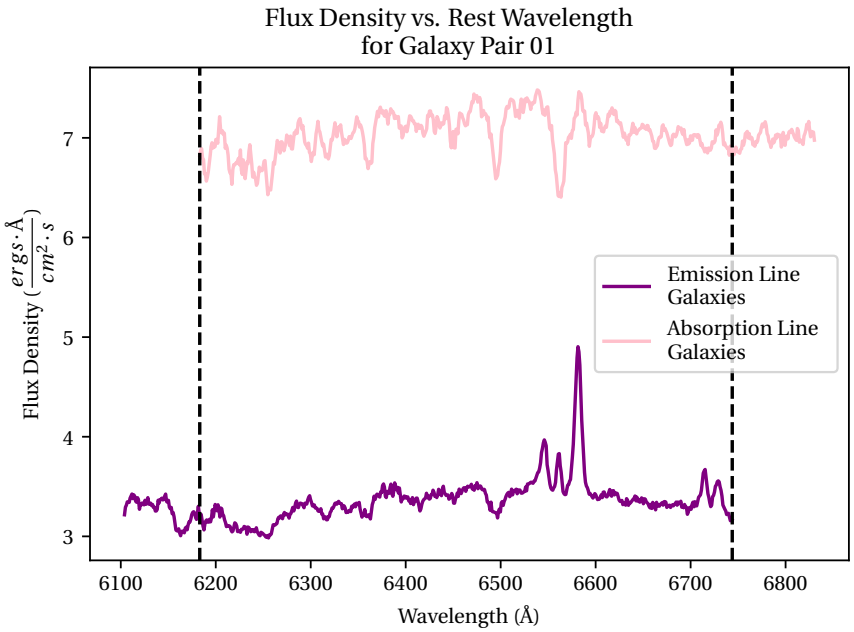

Figure 7: Cutoff wavelengths for original absorption and emission lines. Absorption line sets the minimum wavelength, emission line sets the maximum.

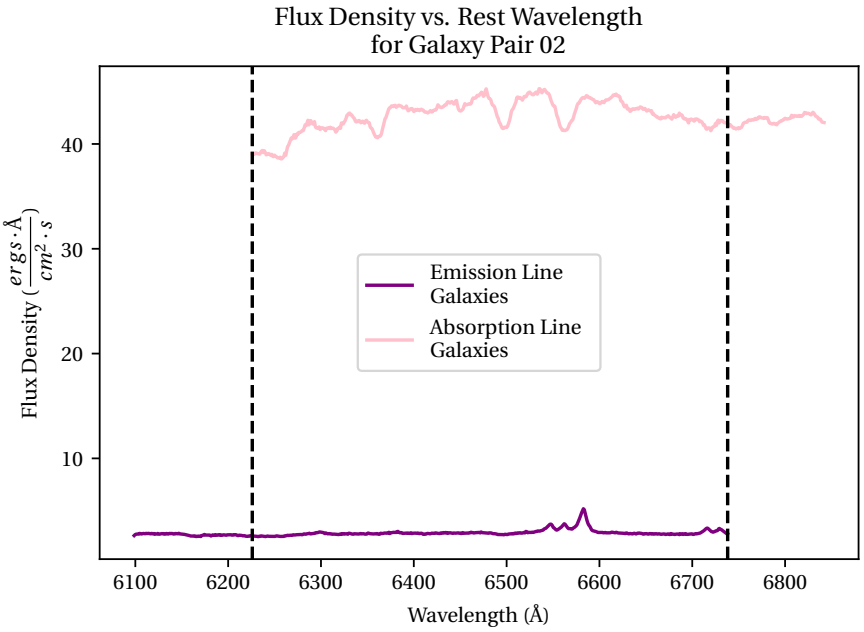

Figure 8: Cutoff wavelengths for original absorption and emission lines. Absorption line sets the minimum wavelength while the emission line sets the maximum wavelength.

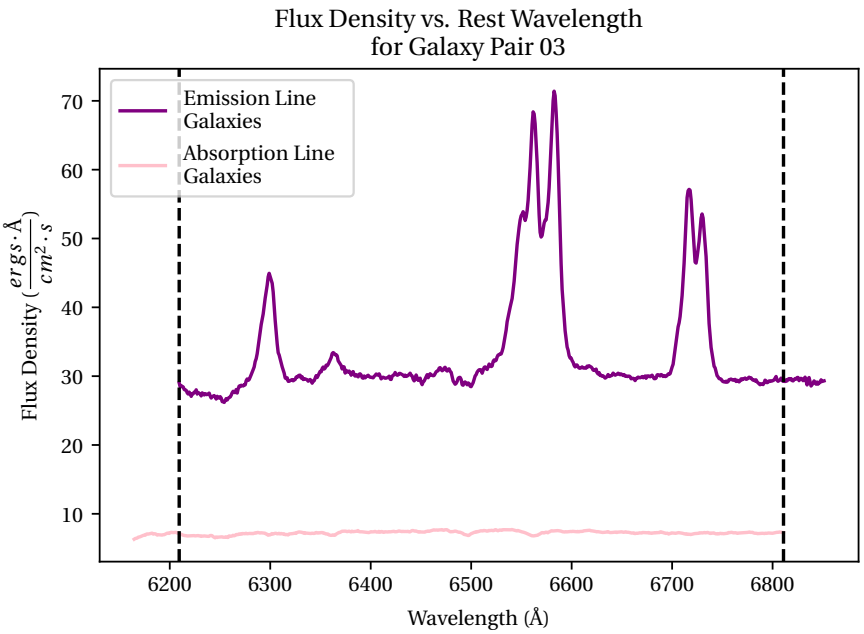

Figure 9: Cutoff wavelengths for original absorption and emission lines. Emission line sets the minimum wavelength and the absorption line dictates the maximum. One of the few cases where the magnitude of the emission line is consistently larger than the magnitude of the absorption line.

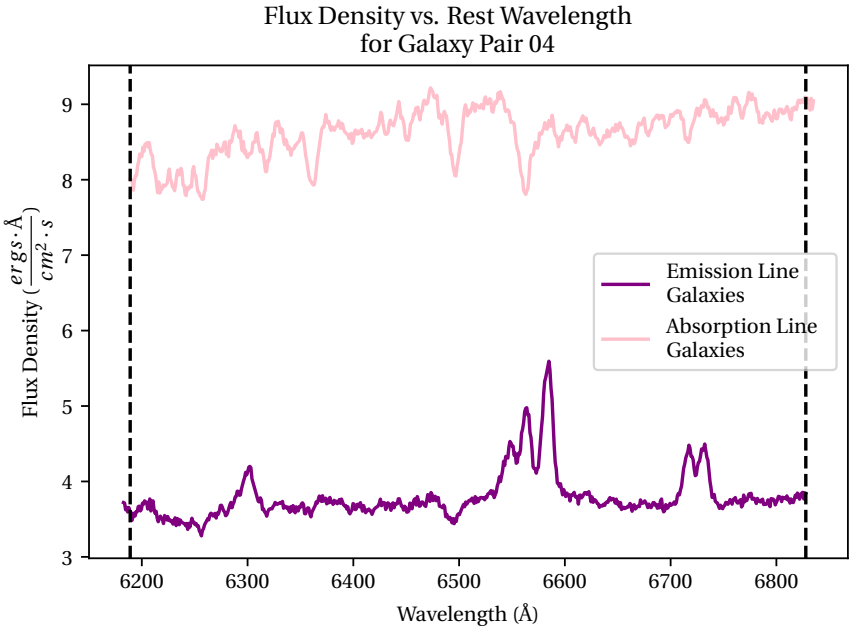

Figure 10: Cutoff wavelengths for original absorption and emission lines. Wavelength ranges of spectra are roughly the same.

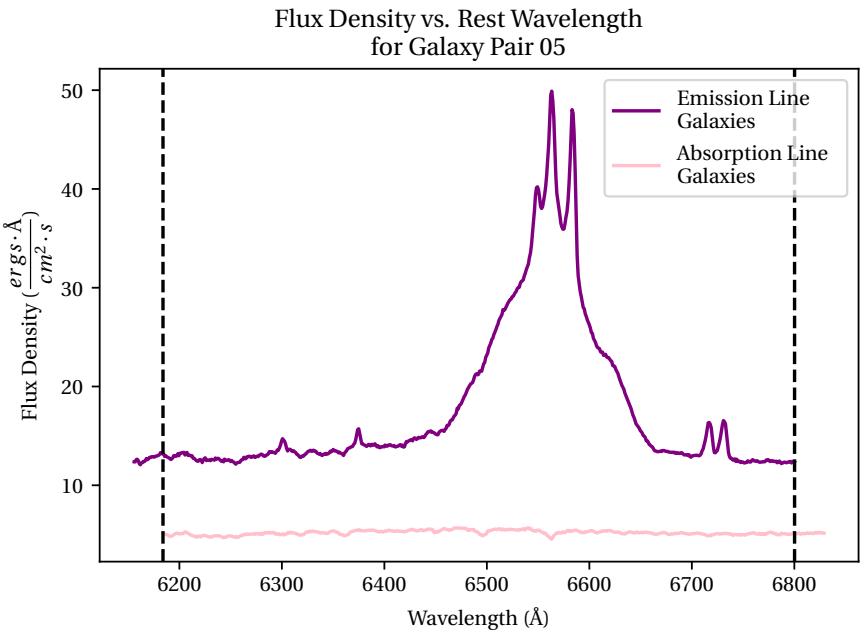

Figure 11: Cutoff wavelengths for original absorption and emission lines. Absorption line set by the minimum wavelength, maximum wavelength set by emission line. One of the handful of cases where the magnitude of the emission line is markedly greater than the magnitude of the absorption line.

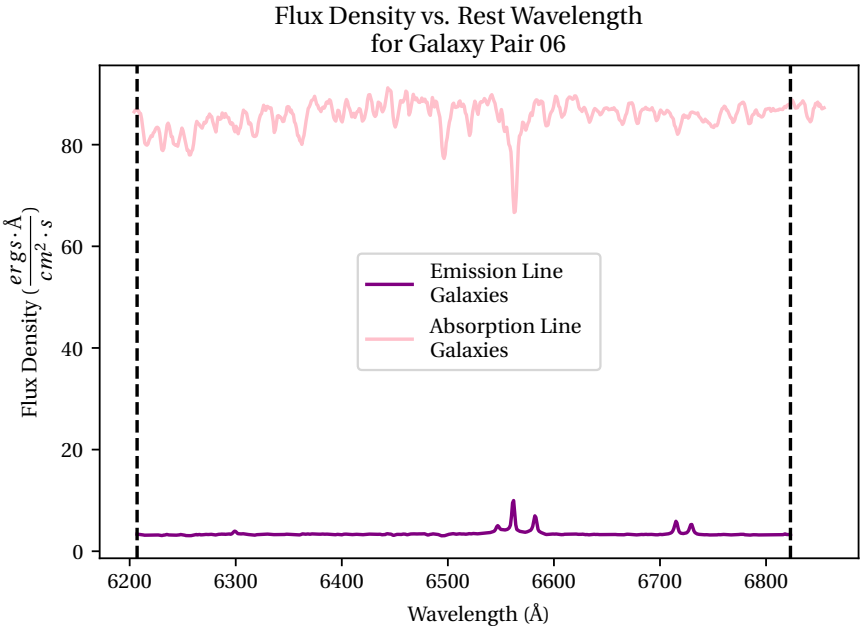

Figure 12: Cutoff wavelengths for original absorption and emission lines. Wavelength minimum is roughly the same, though the maximum is set by the emission line. Magnitude of absorption line is markedly greater than the magnitude of the emission line

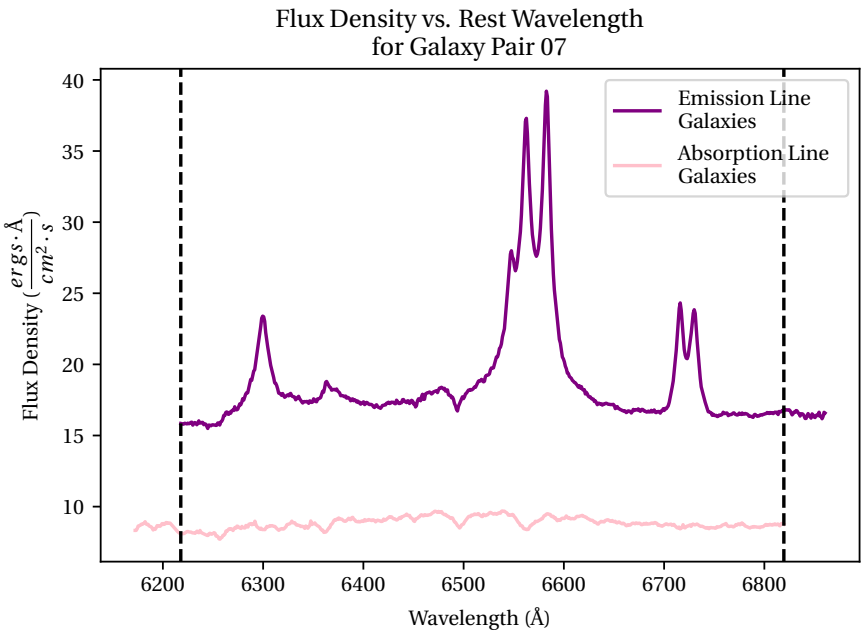

Figure 13: Cutoff wavelengths for original absorption and emission lines. Emission line determines the minimum wavelength, absorption line sets the maximum. One of the handful of cases where the magnitude of the emission line is a good deal larger than the absorption line.

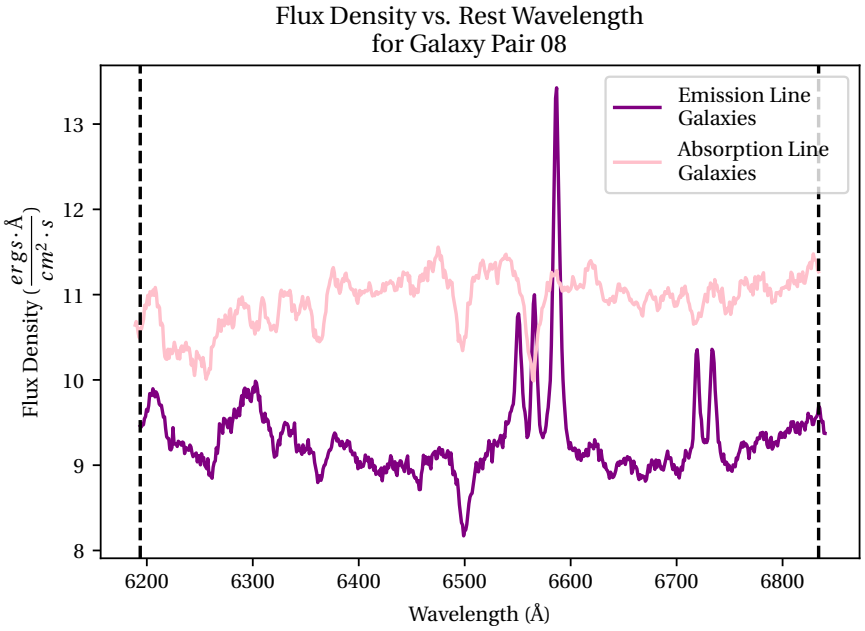

Figure 14: Cutoff wavelengths for original absorption and emission lines. Wavelength ranges are roughly the same. Rare case of the magnitude of the spectra intersecting at points.

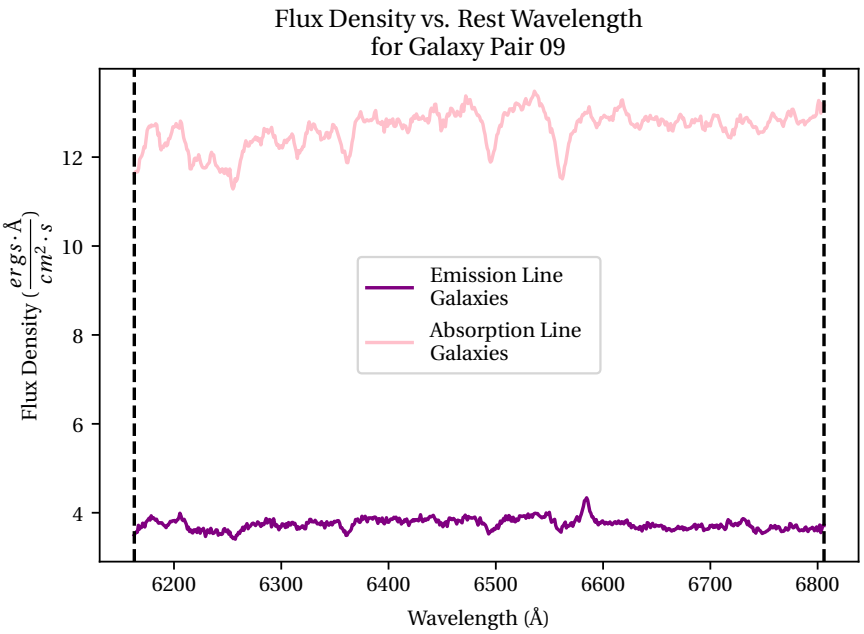

Figure 15: Cutoff wavelengths for original absorption and emission lines. Wavelength ranges are roughly similar.

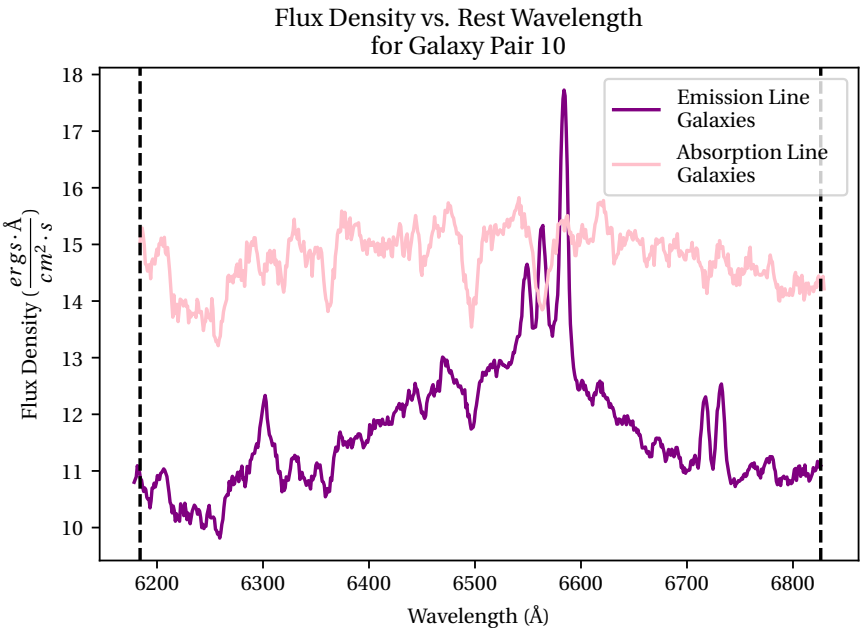

Figure 16: Cutoff wavelengths for original absorption and emission lines. Wavelength ranges are approximately the same. Rare case where the magnitude of the absorption and emission line spectra intersect

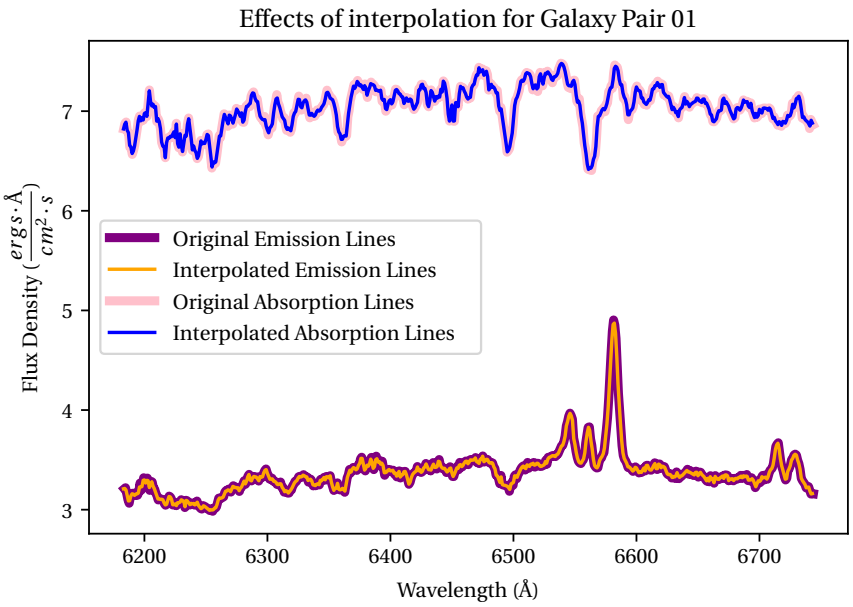

Figure 17: Interpolation of spectra.

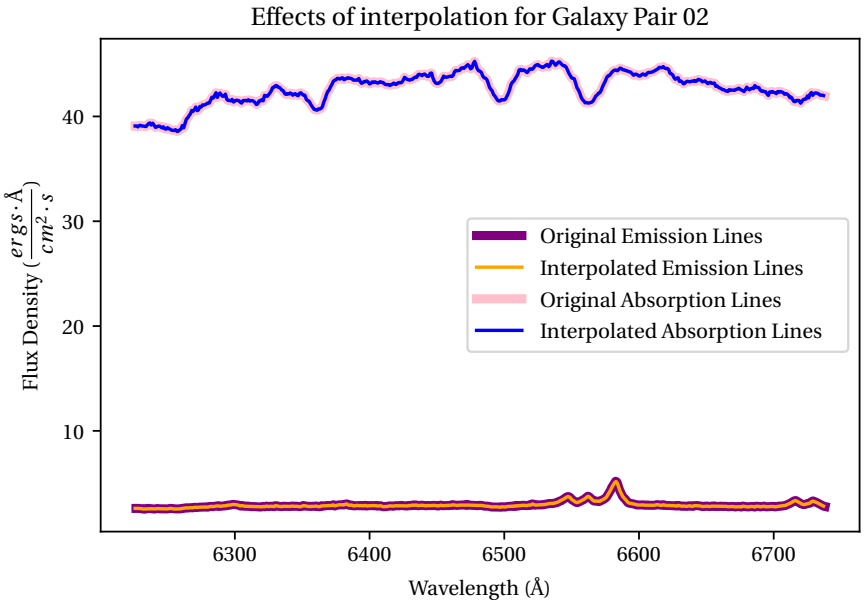

Figure 18: Interpolation of spectra.

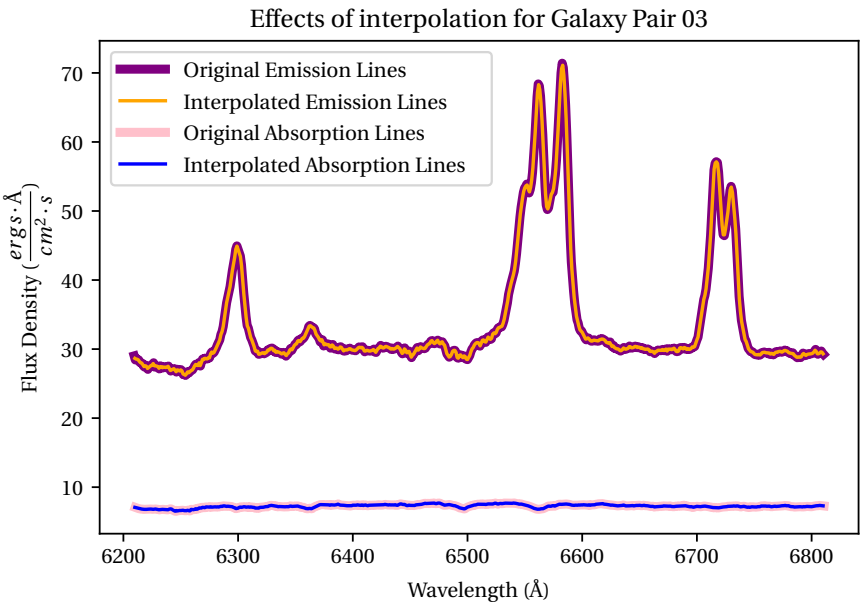

Figure 19: Interpolation of spectra. Emission line spectrum has noticeable emission lines both in and outside the range of interest.

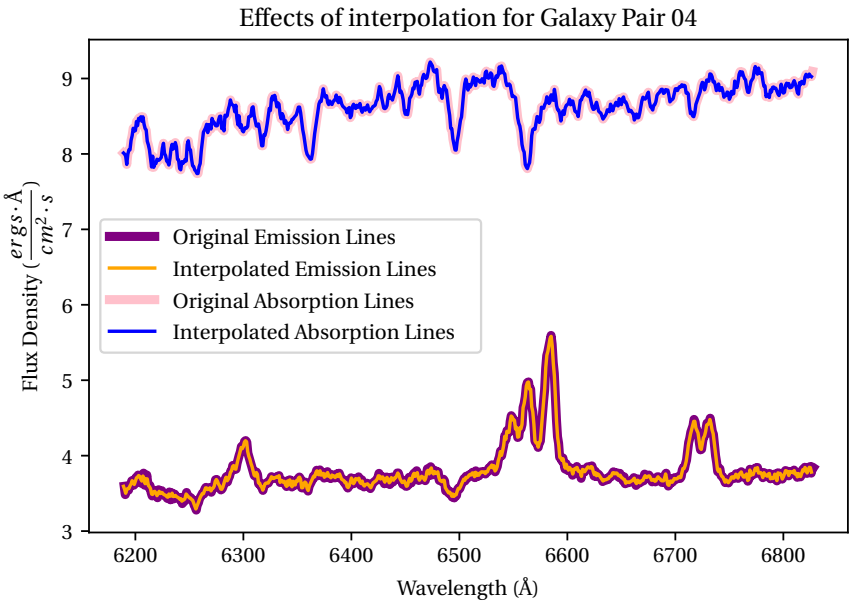

Figure 20: Interpolation of spectra. Emission line spectrum has noticeable emission lines both in and outside the range of interest.

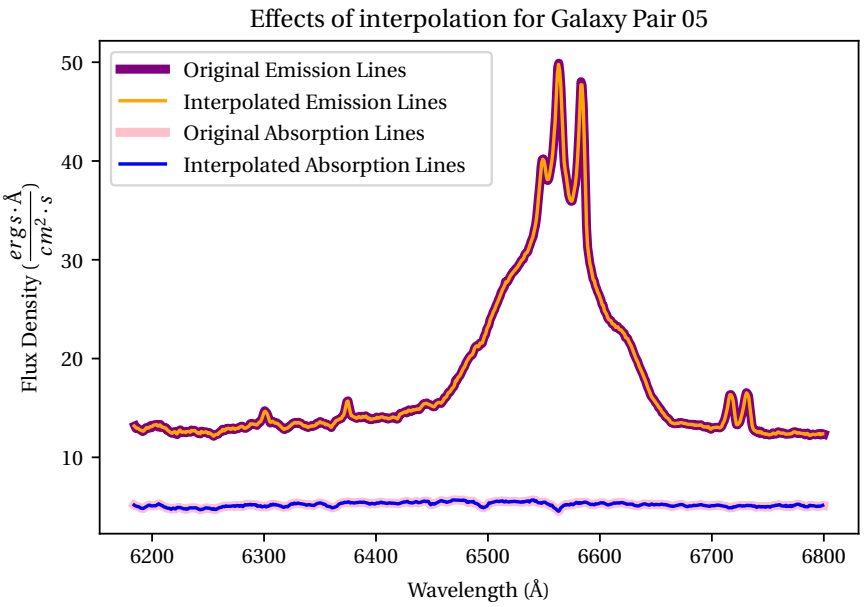

Figure 21: Interpolation of spectra.

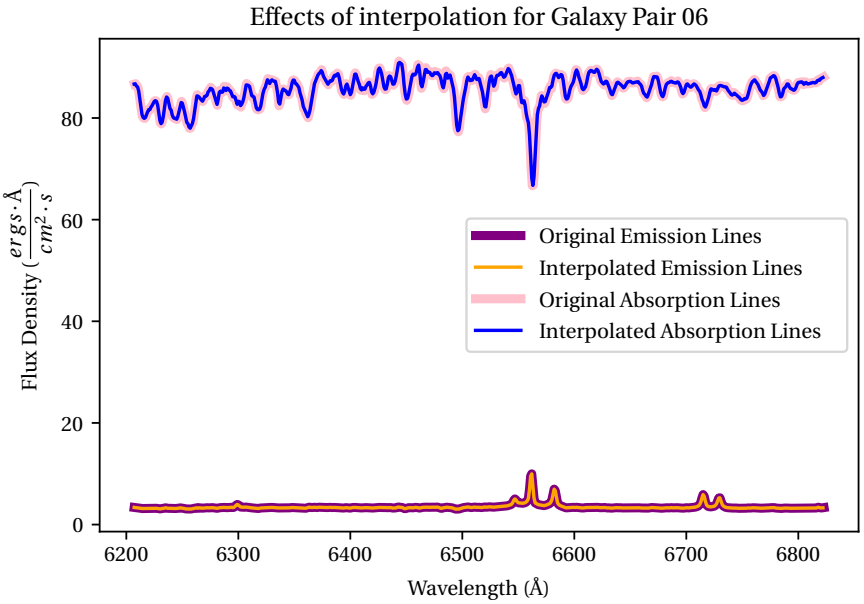

Figure 22: Interpolation of spectra.

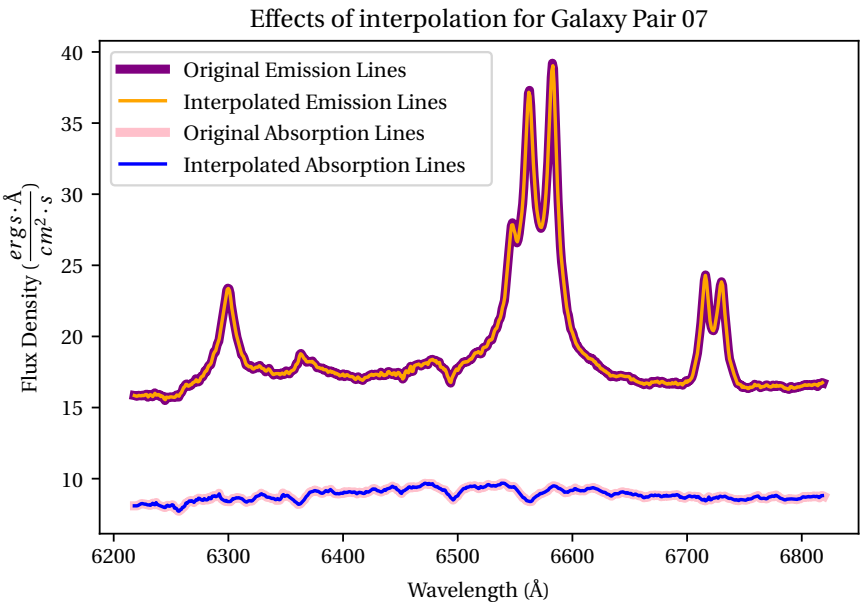

Figure 23: Interpolation of spectra. Emission line spectrum has noticeable emission lines both in and outside the range of interest.

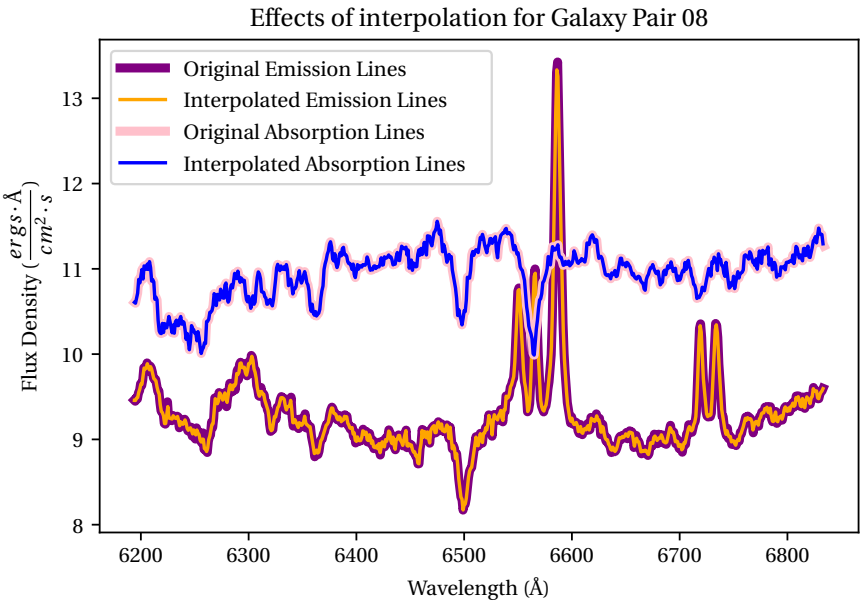

Figure 24: Interpolation of spectra. Very prominent emission line is present within range of interest.

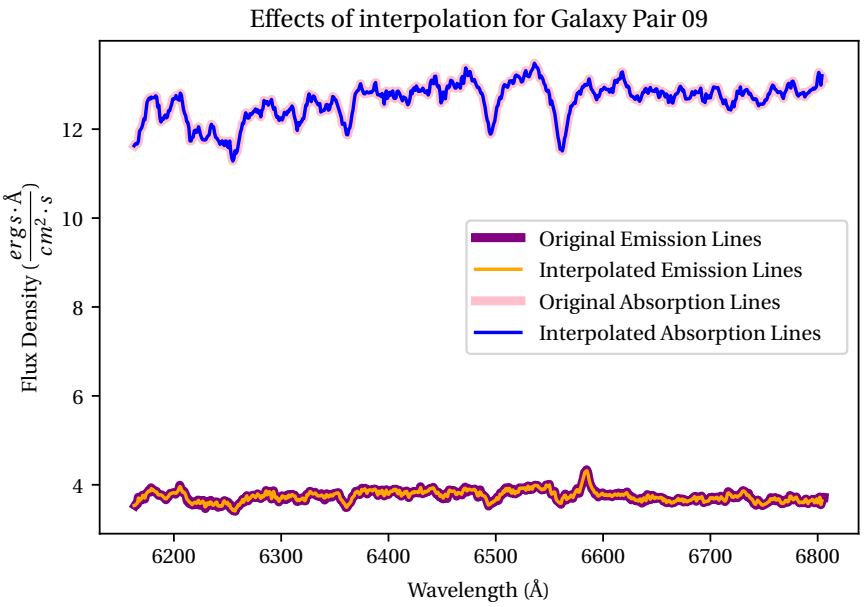

Figure 25: Interpolation of spectra.

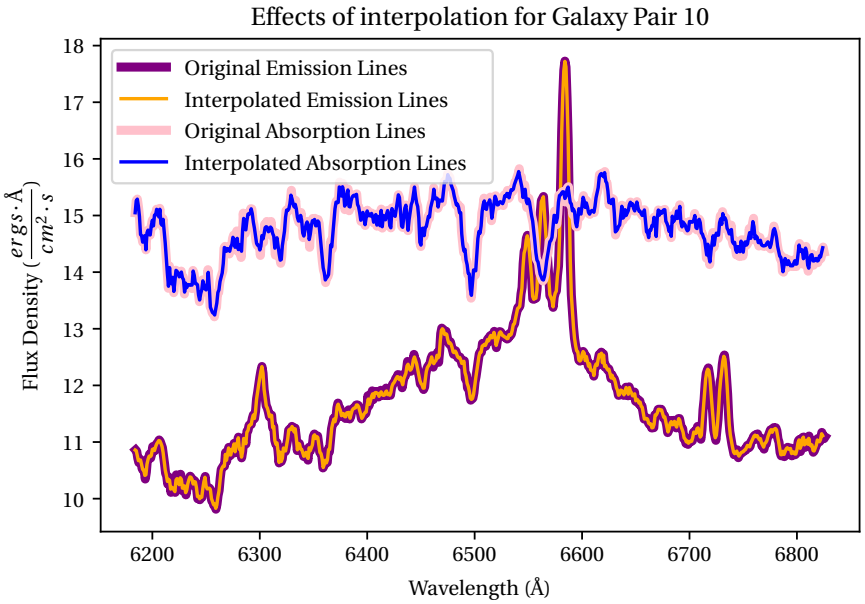

Figure 26: Interpolation of spectra. Very prominent emission line is present within range of interest.

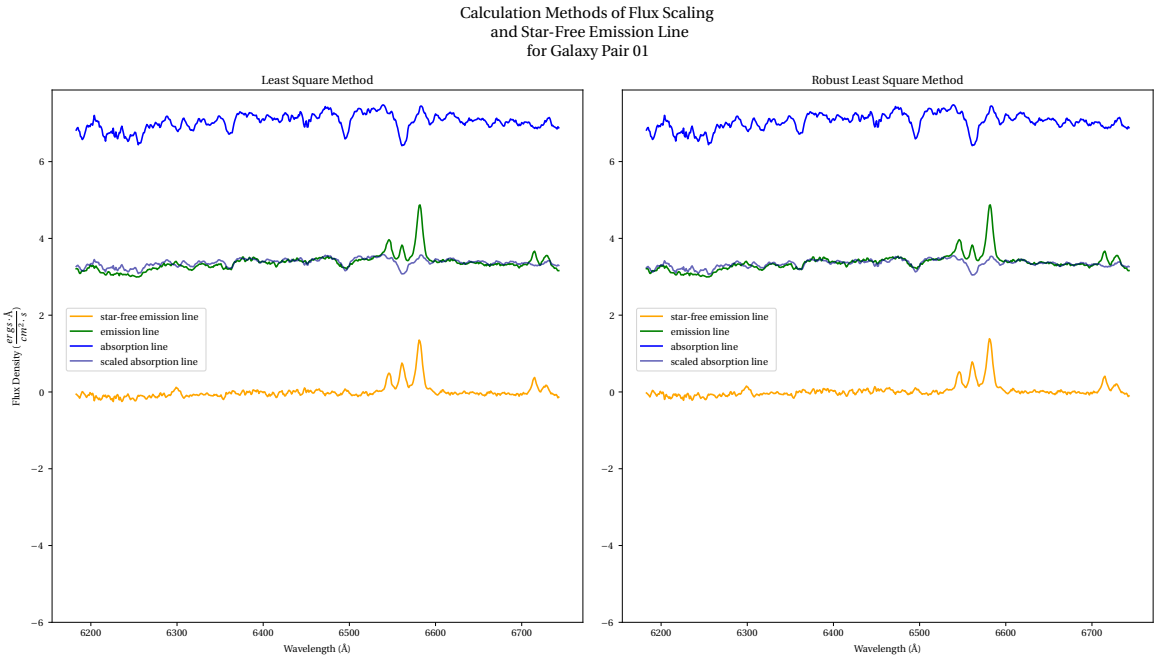

Figure 27: Comparison of optimisation methods for absorption line scaling. Methods produce roughly the same scaled absorption line.

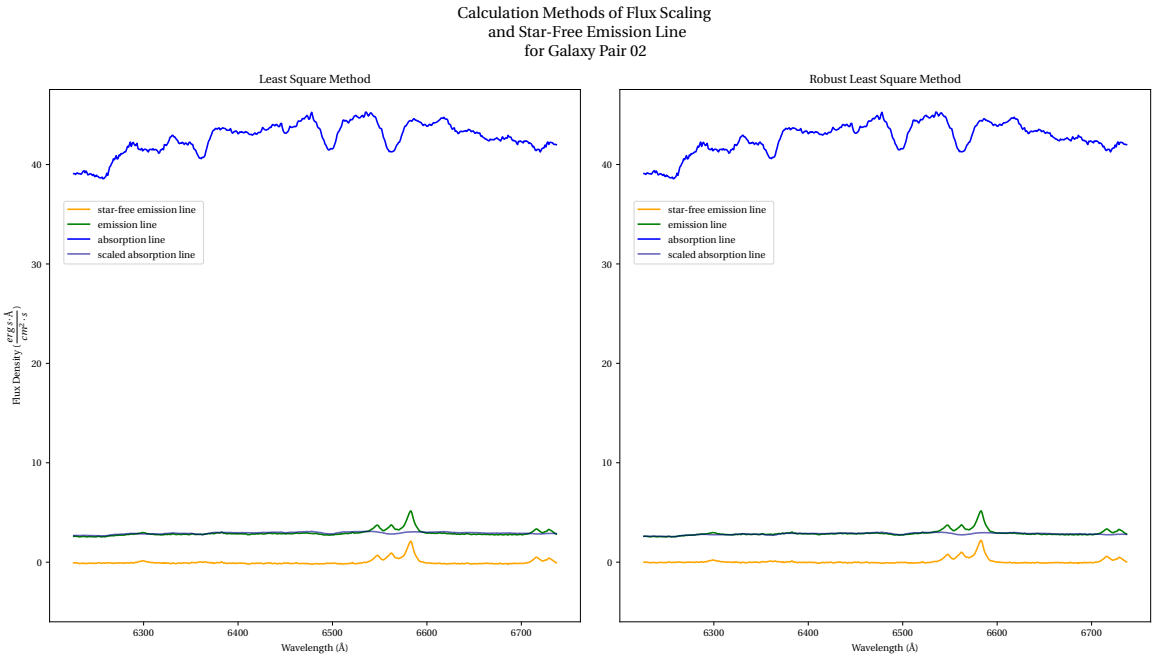

Figure 28: Comparison of optimisation methods for absorption line scaling. Difference in methods aren't noticeably different, probably because the emission line spectrum has few pronounced emission line peaks.

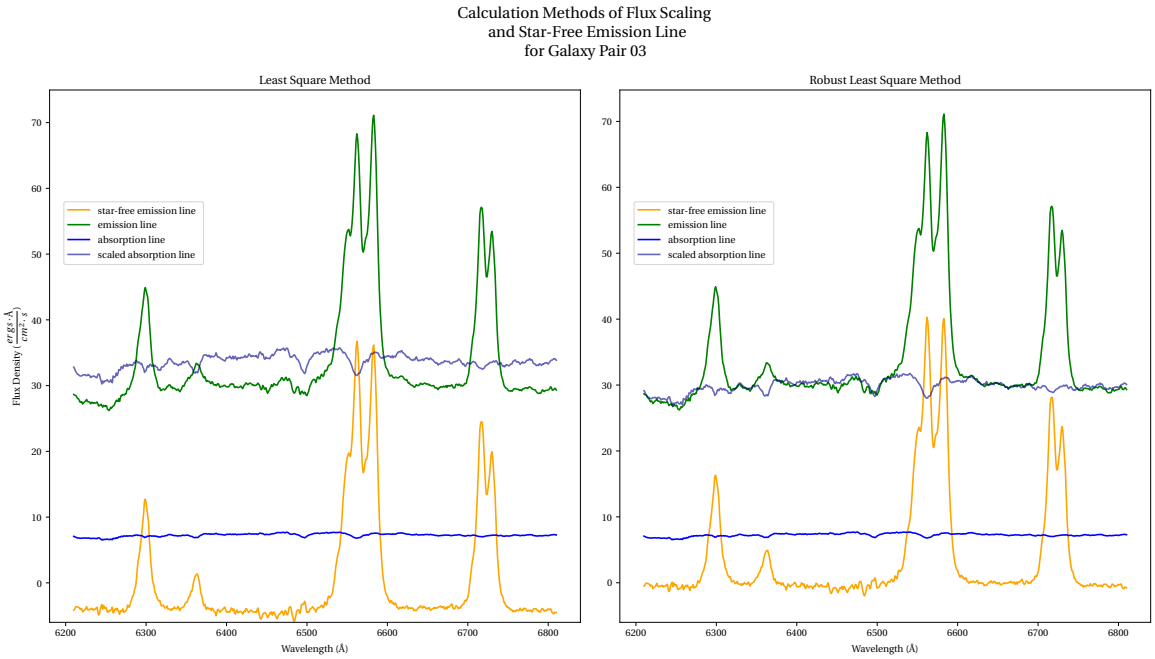

Figure 29: Comparison of optimisation methods for absorption line scaling. Robust least squares method has a notably closer fit.

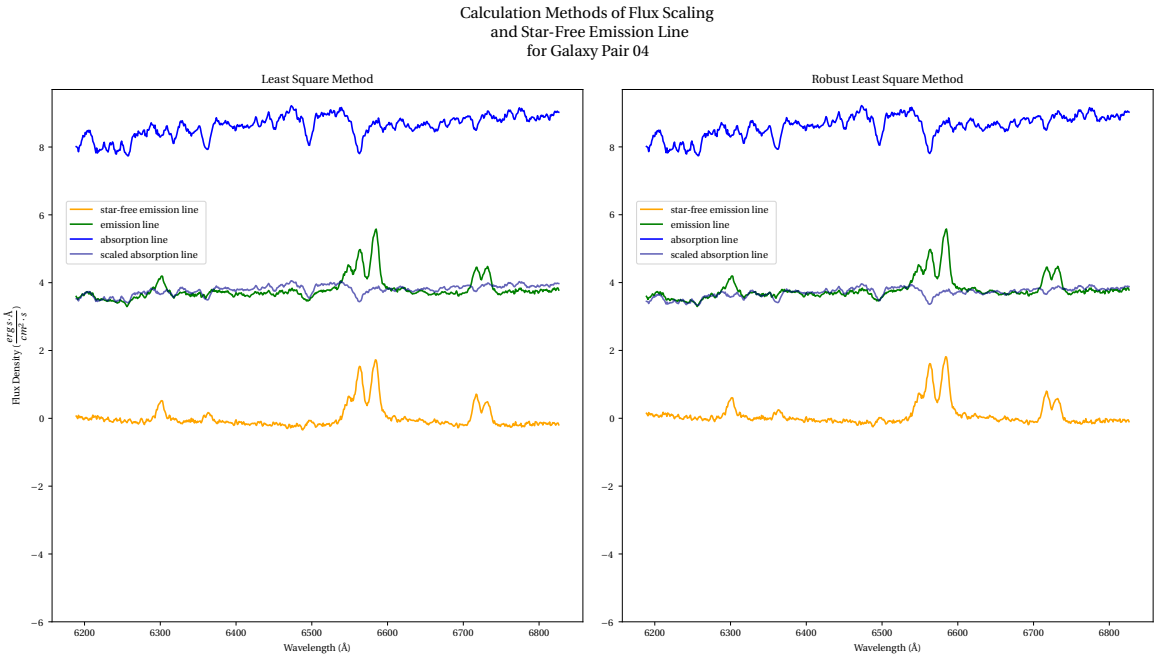

Figure 30: Comparison of optimisation methods for absorption line scaling. Robust least squares has a very slight edge in fitting accuracy even though the emission lines aren't particularly pronounced in this spectrum.

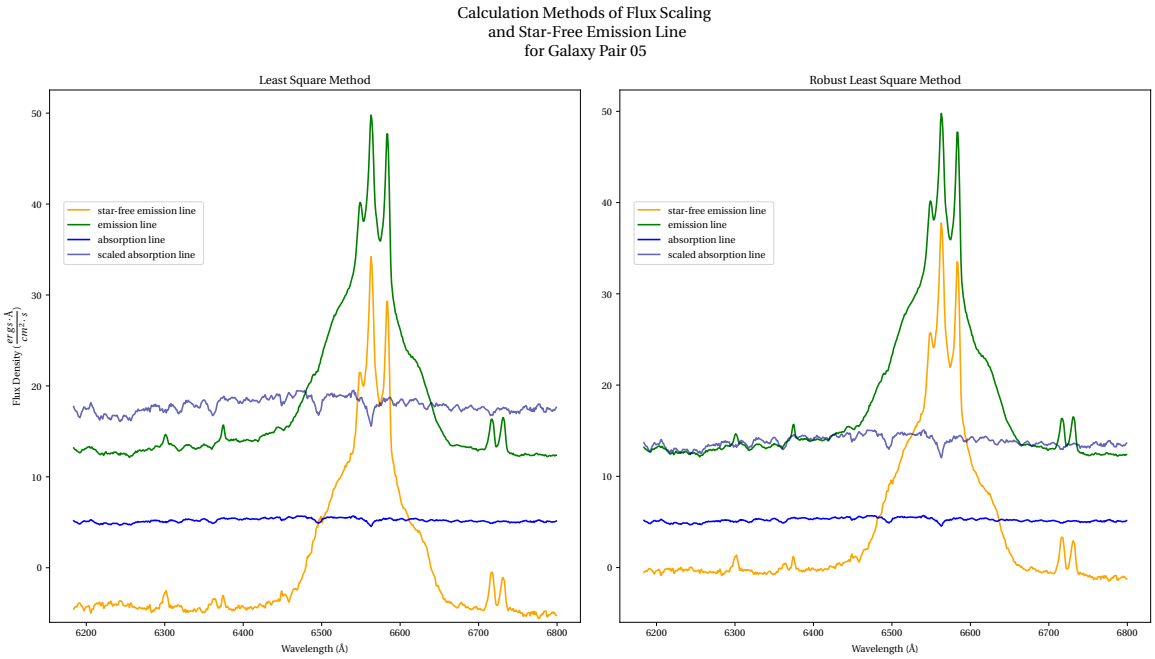

Figure 31: Comparison of optimisation methods for absorption line scaling. The robust least squares method produces a scaled absorption line spectrum that is significantly better fitted to the emission line spectrum

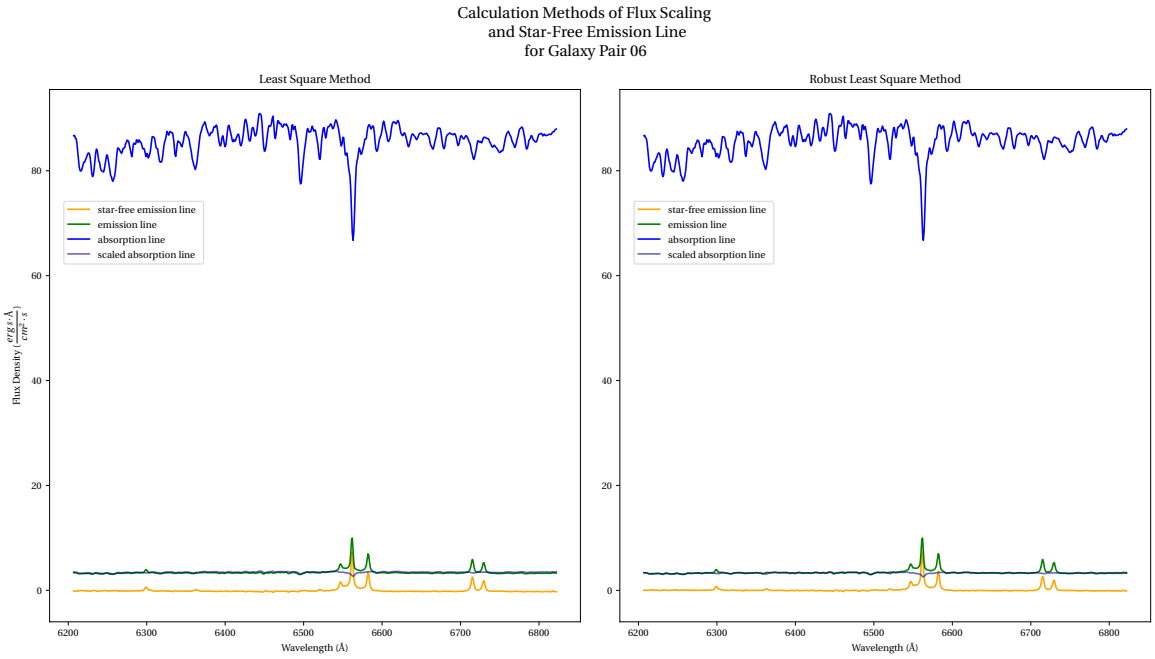

Figure 32: Comparison of optimisation methods for absorption line scaling. Methods demonstrate no clear difference, though the emission line spectrum is diminutive in comparison to the original absorption line spectrum and has fairly small emission line peaks.

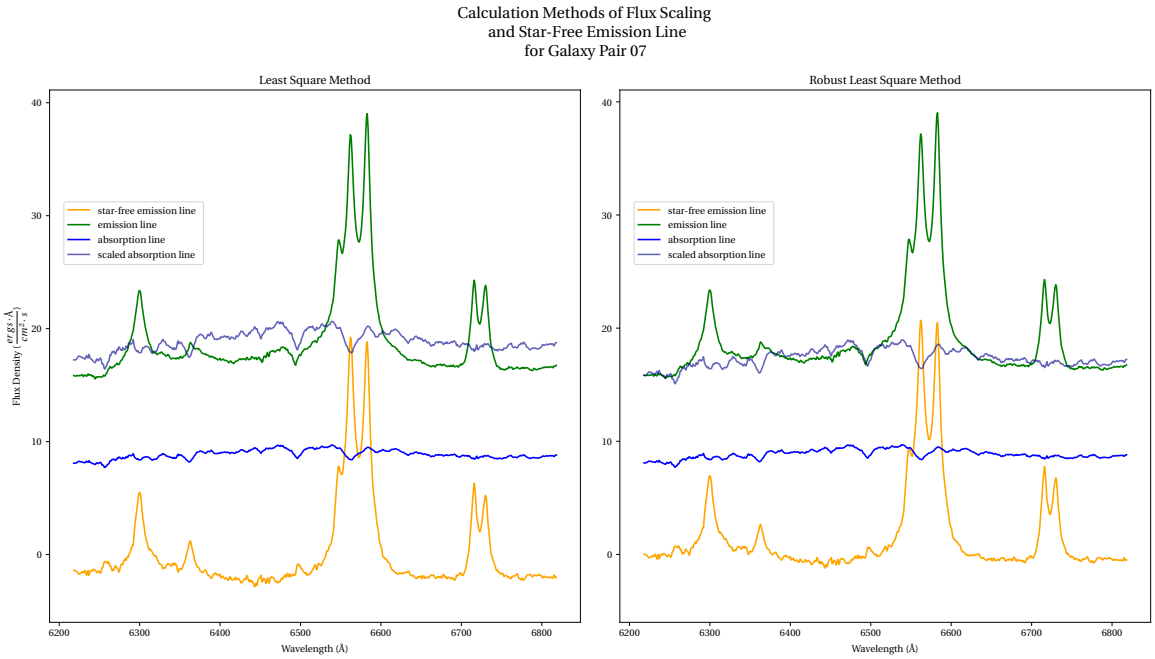

Figure 33: Comparison of optimisation methods for absorption line scaling. Robust least squares method produces a noticeably more accurate scale factor

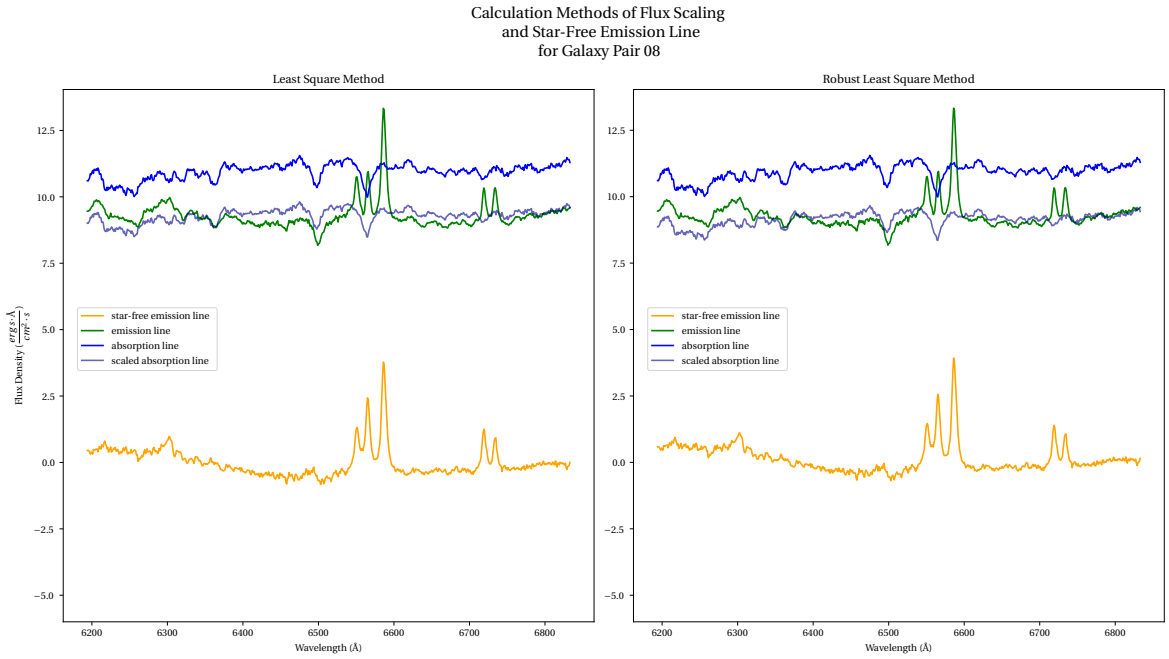

Figure 34: Comparison of optimisation methods for absorption line scaling. Robust method is slightly better at scaling the absorption line spectrum to the emission line spectrum.

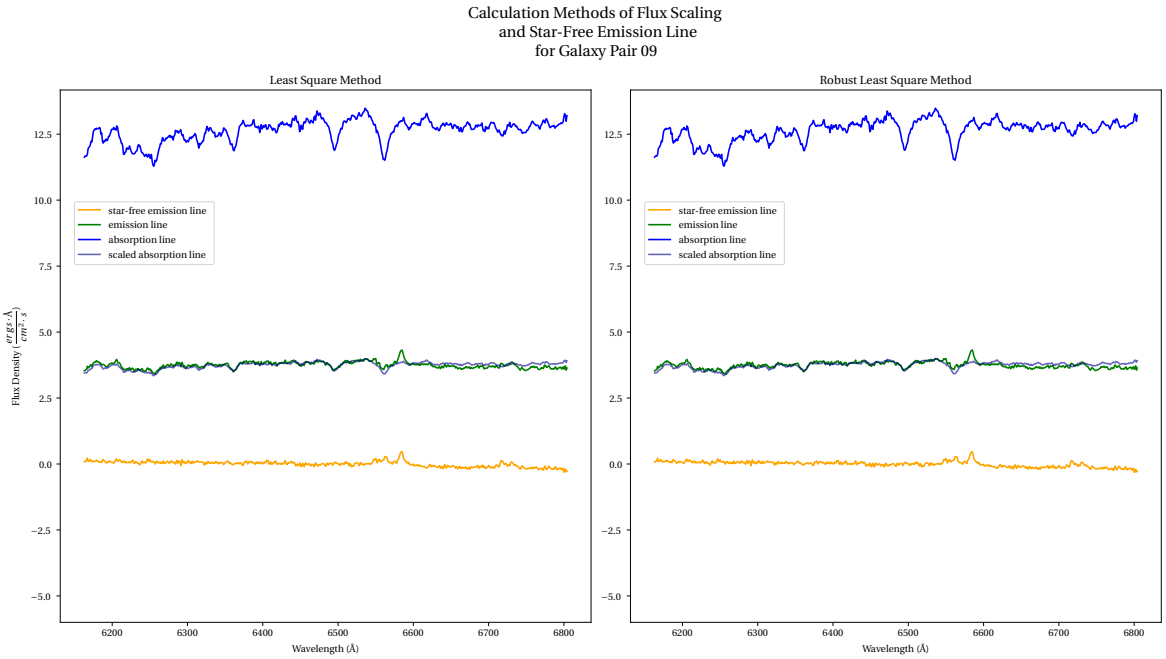

Figure 35: Comparison of optimisation methods for absorption line scaling. There is no noticeable difference between the methods, likely because the emission line spectrum shows few noticeable peaks

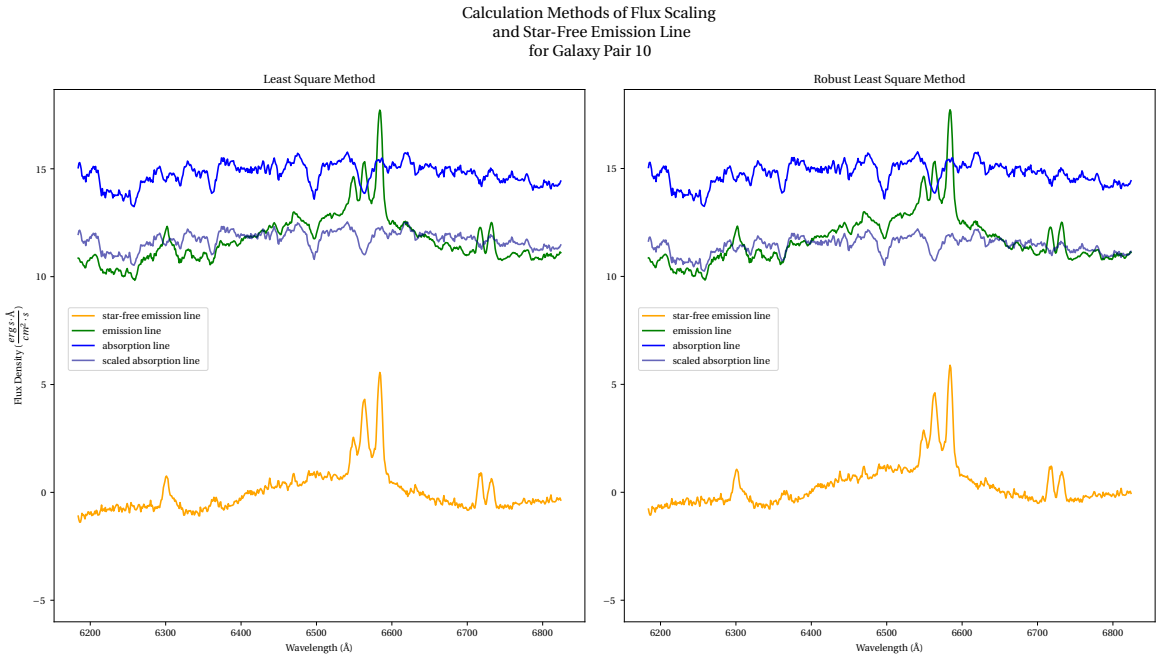

Figure 36: Comparison of optimisation methods for absorption line scaling. Robust least squares method provides better fit but still unable to properly scale the absorption line for some wavelengths.

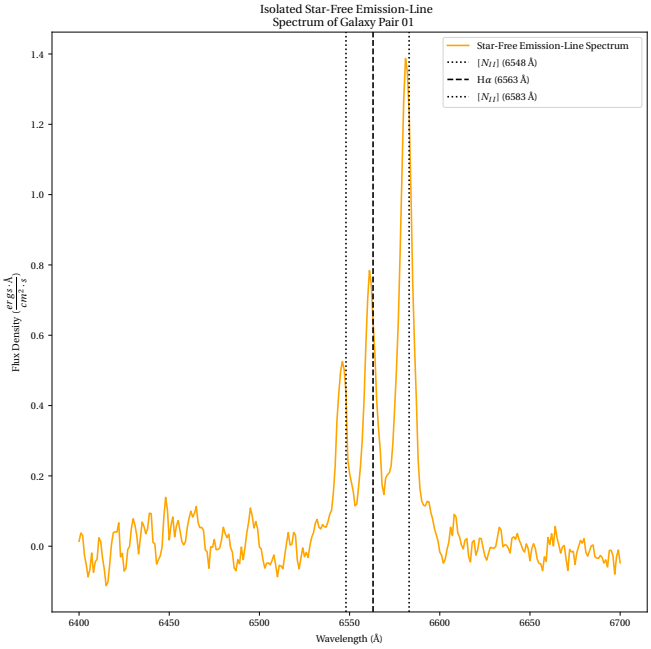

Figure 37: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms. Emission Lines appear to be slightly offset to the left.

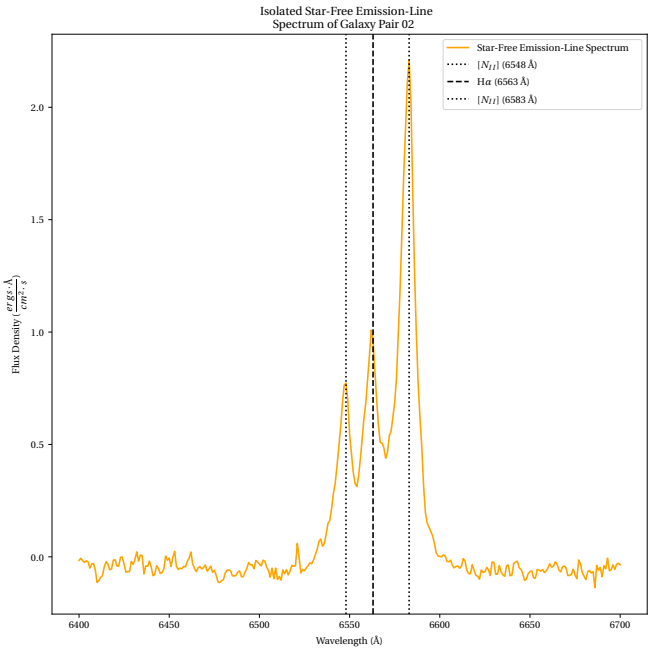

Figure 38: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

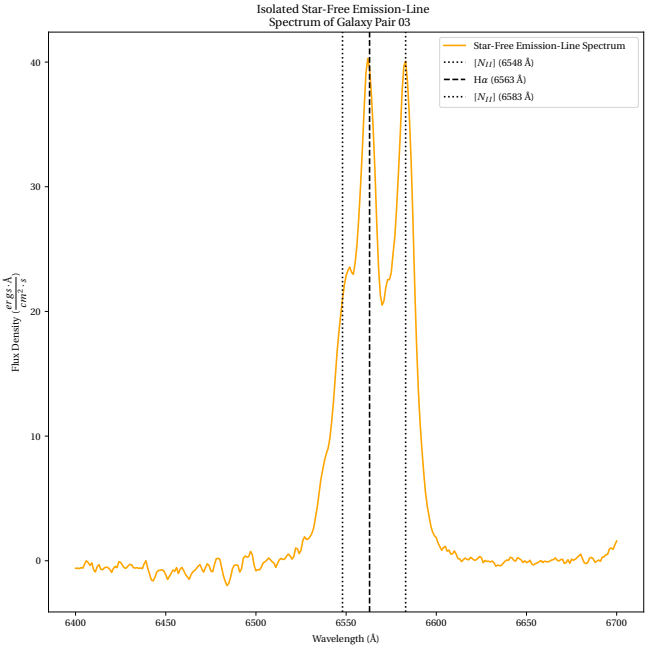

Figure 39: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

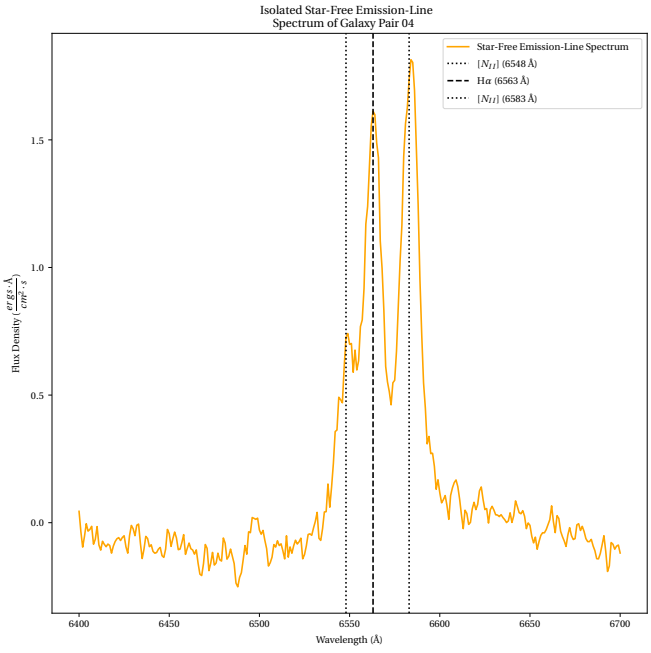

Figure 40: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

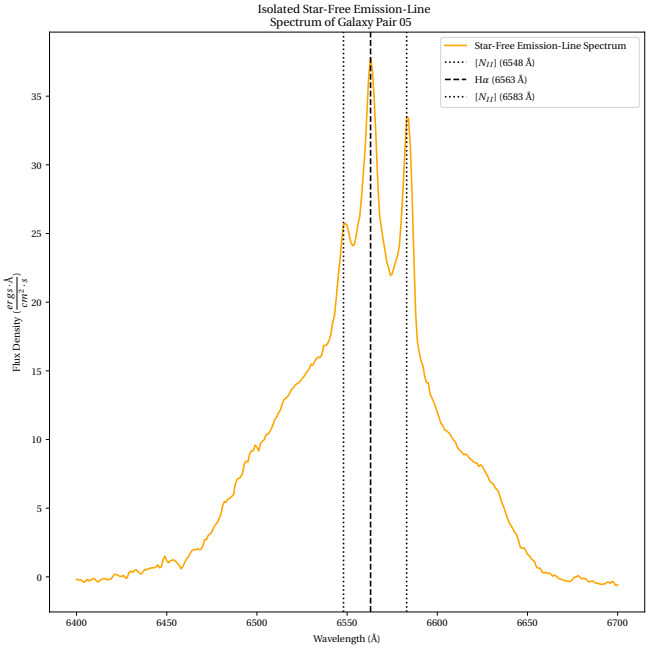

Figure 41: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

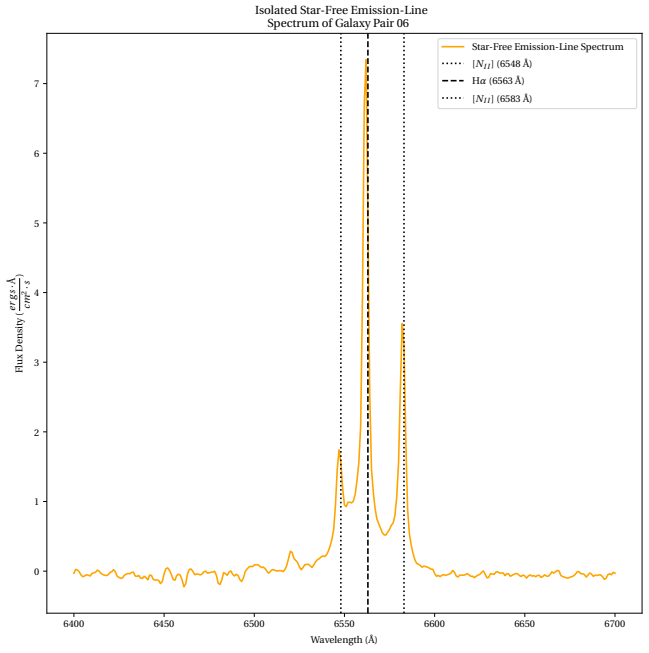

Figure 42: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms. Emission Lines appear to be slightly offset to the left, made especially apparent by the pronounced Hα line.

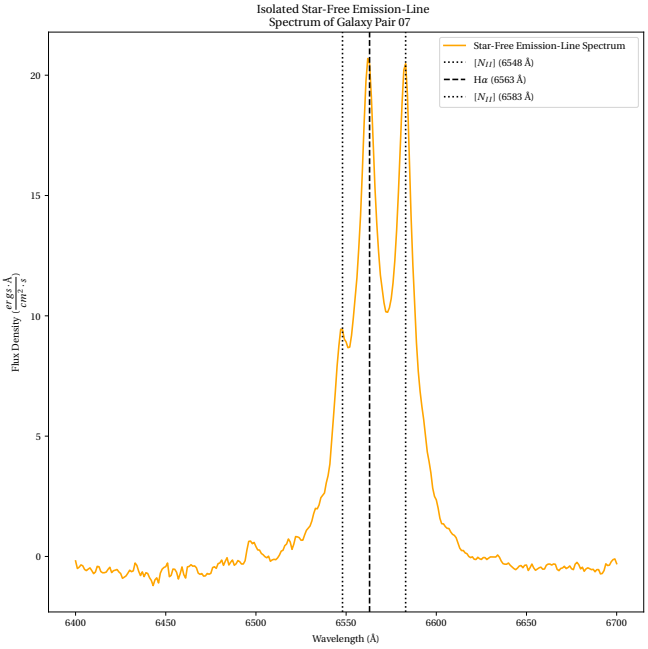

Figure 43: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

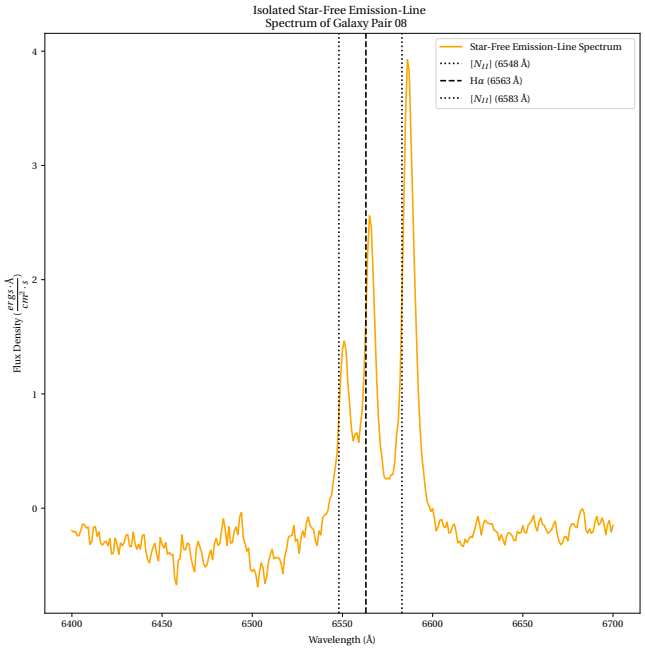

Figure 44: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms. Emission Lines appear to be slightly offset to the right.

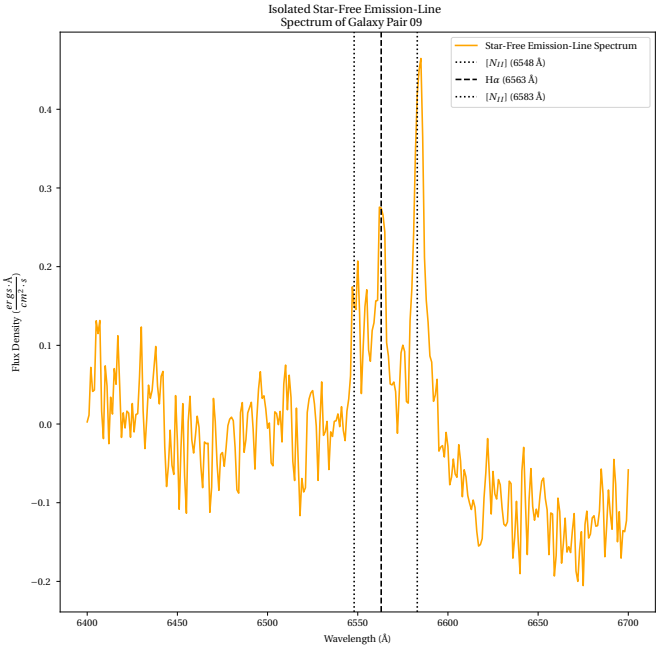

Figure 45: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms. While the first $[NII]$ seems to be present, its peak is only slightly larger in magnitude than the noise of the spectrum.

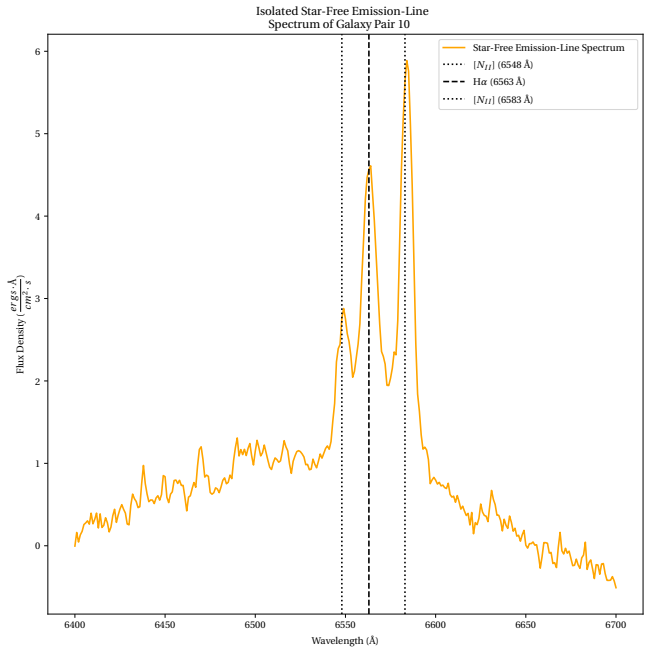

Figure 46: Star-Free Emission Line Spectrum isolated to between 6400 and 6700 Angstroms.

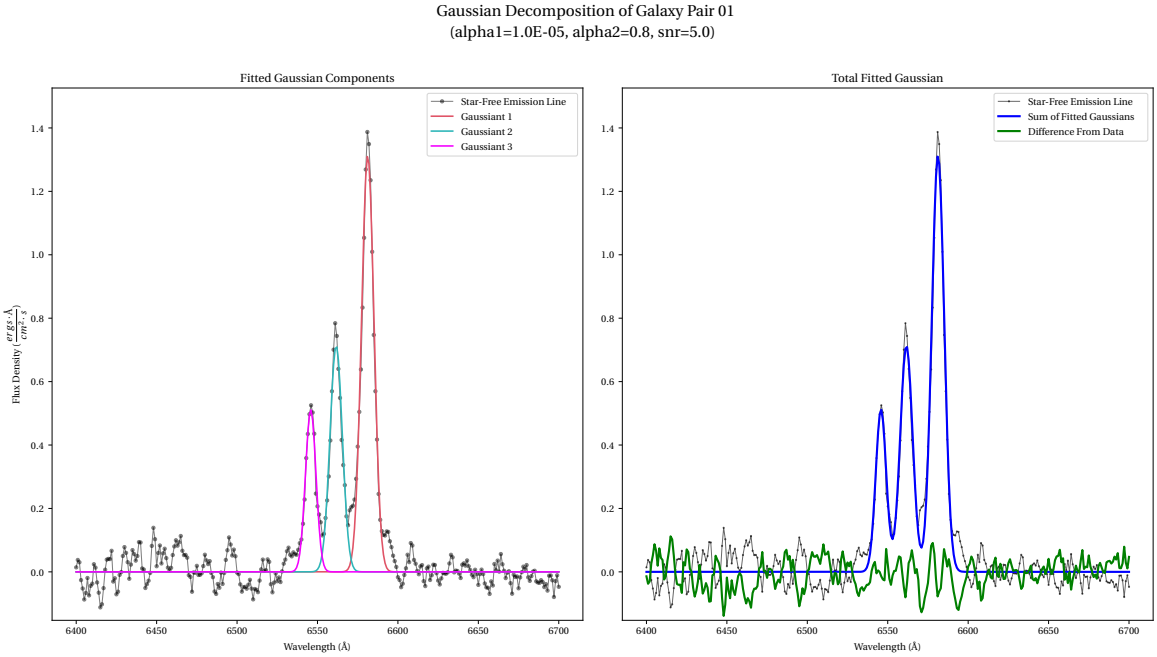

Figure 47: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components.

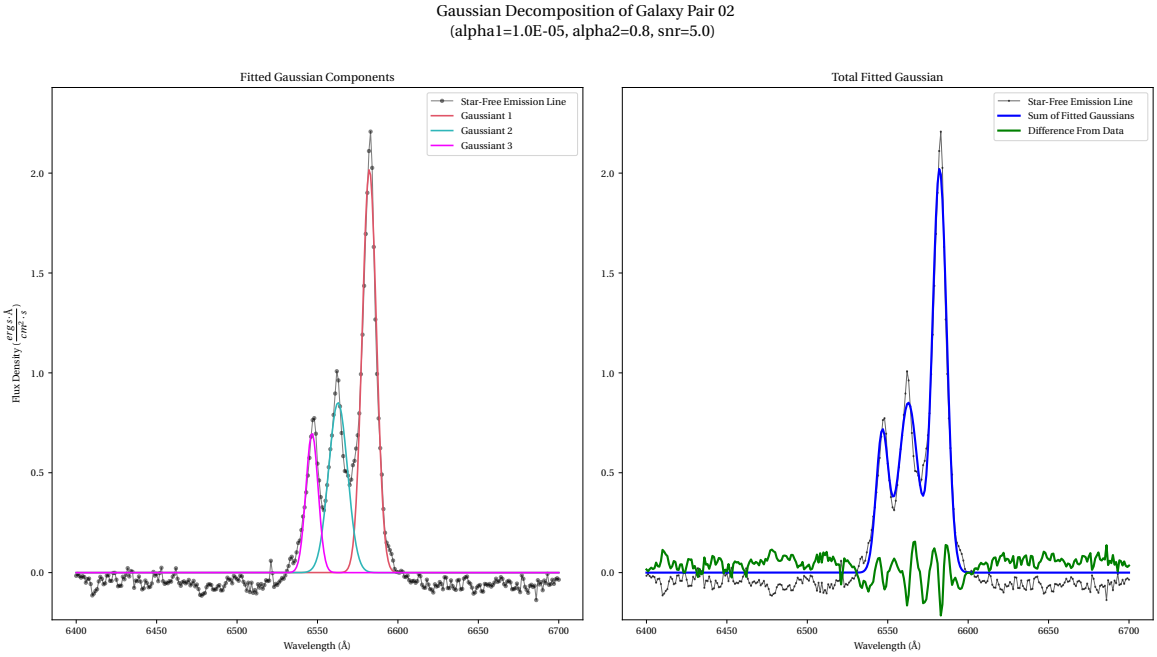

Figure 48: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components.

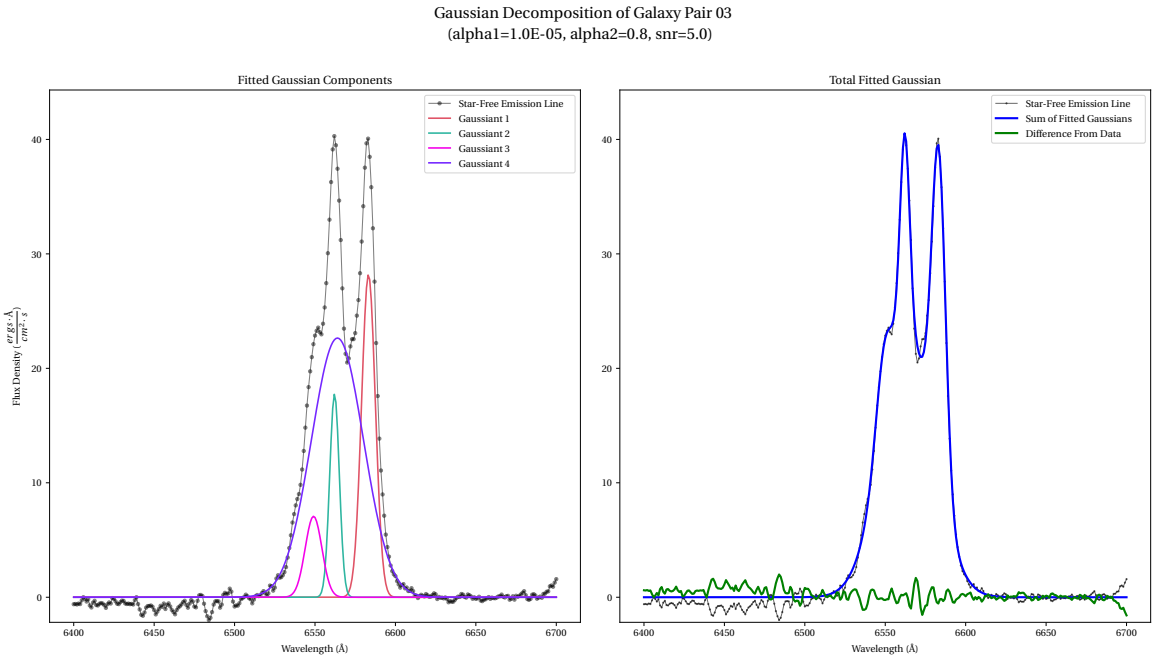

Figure 49: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components and a potential broad velocity Hα component.

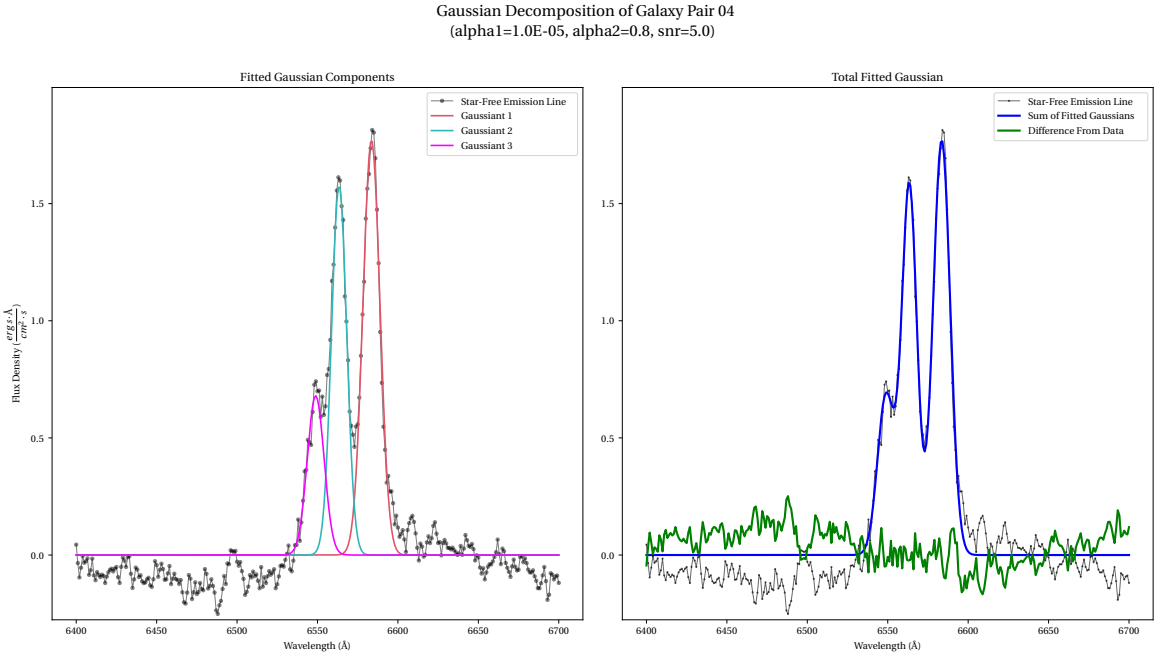

Figure 50: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components.

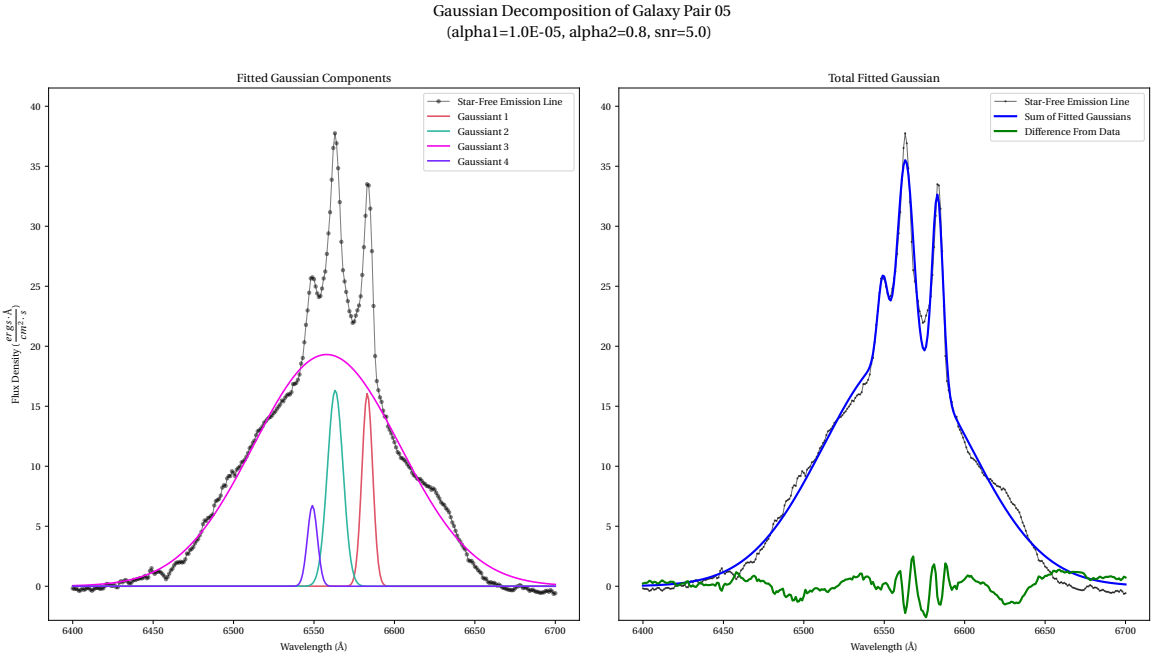

Figure 51: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components and a potential broad velocity Hα component. This is the broadest component of any of the spectra.

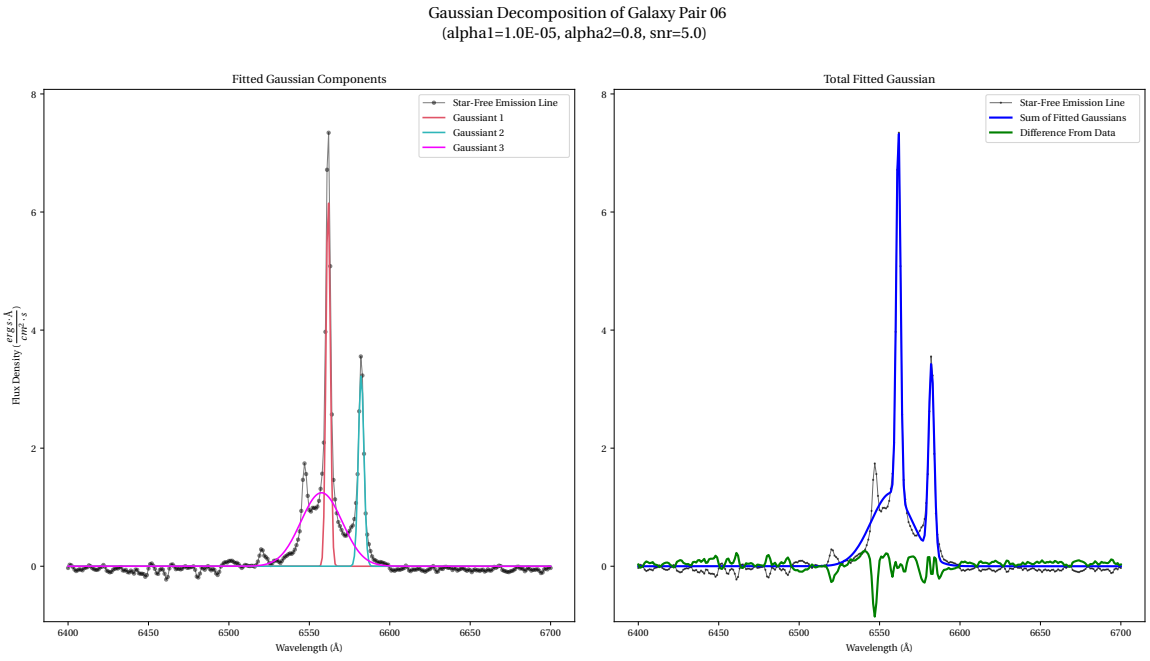

Figure 52: A Gaussian Decomposition using the constant parameters chosen for the analysis. With these parameters, the spectrum did not decompose with all three narrow components being identified, but a potential broad velocity Hα component is identified.

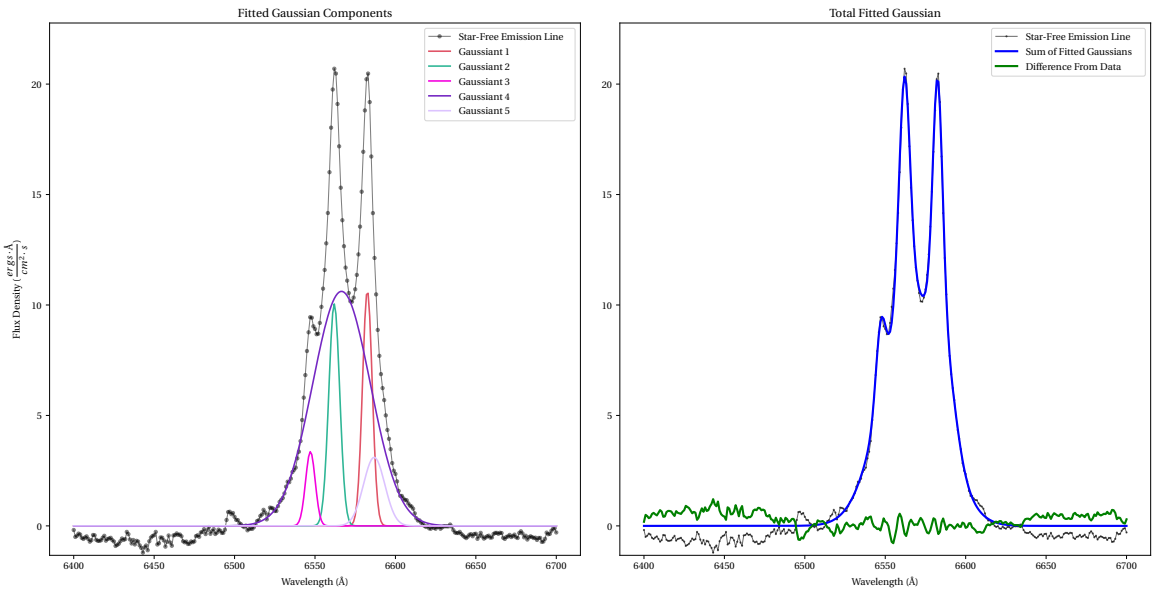

Figure 53: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum demonstrates three narrow velocity components and a potential broad velocity Hα component, but there is also an additional [NII] component that may not be physically plausible.

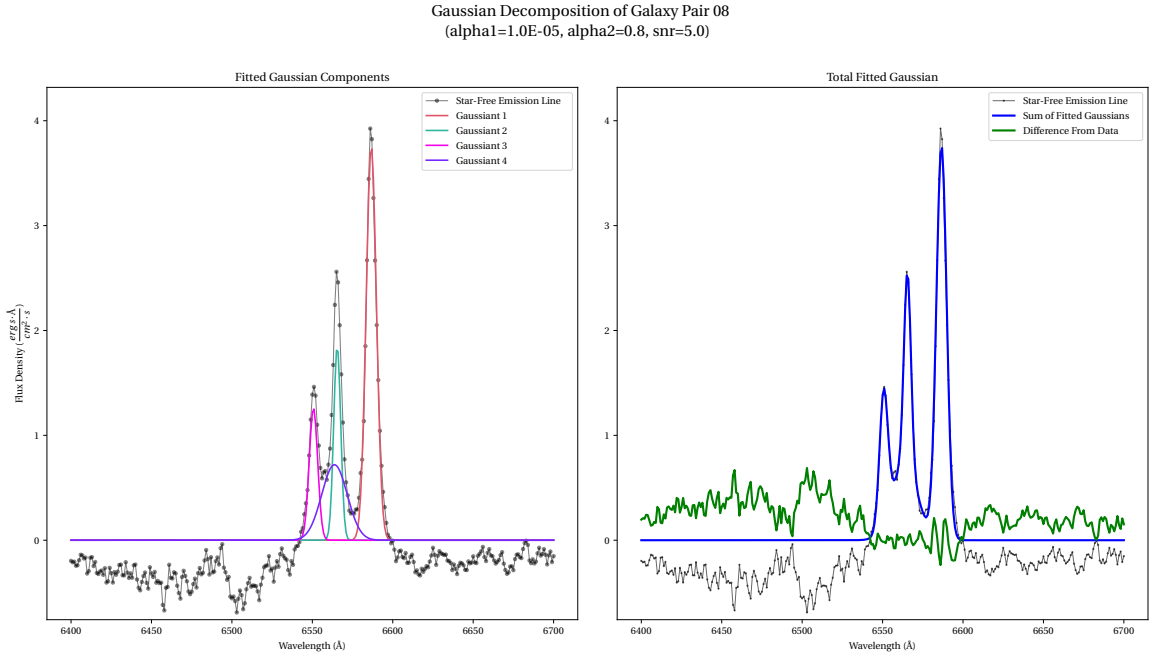

Figure 54: A Gaussian Decomposition using the constant parameters chosen for the analysis. This spectrum decomposed to three narrow components and a potential broad velocity Hα component.

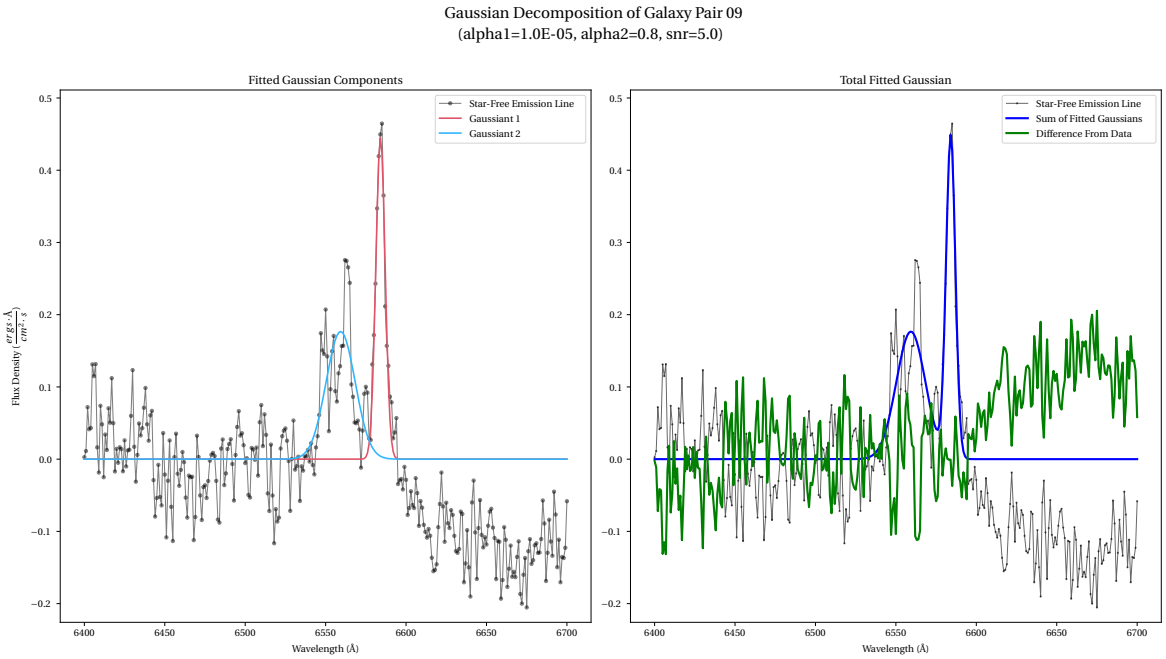

Figure 55: A Gaussian Decomposition using the constant parameters chosen for the analysis. The signal to noise ratio of this spectrum was too low for `GaussPy` to properly decompose.

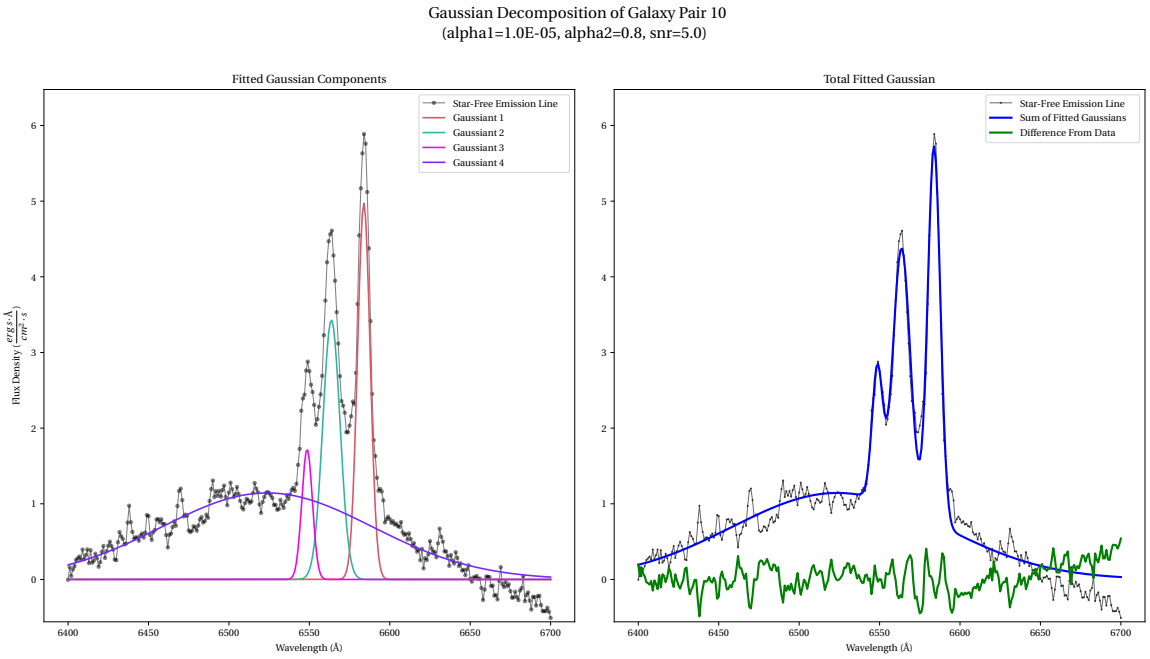

Figure 56: A Gaussian Decomposition using the constant parameters chosen for the analysis. While three narrow velocity components are identified, there is also an unidentified broad component at an unexpected wavelength.

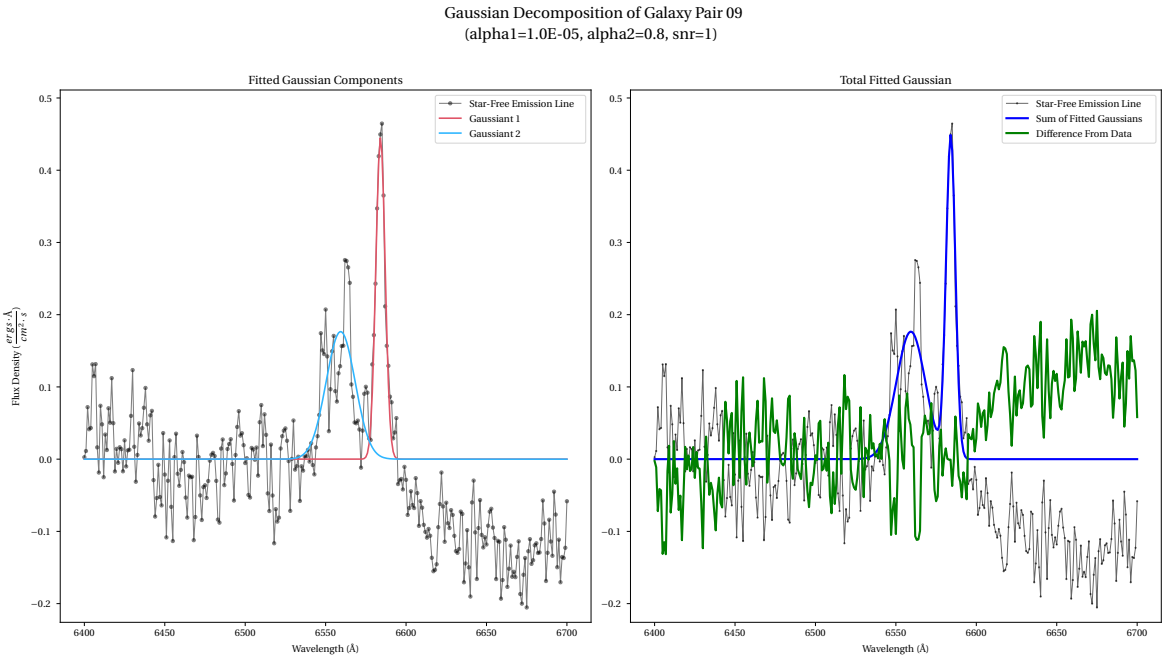

Figure 57: An example of a lowered signal to noise threshold parameter applied to a noisy spectrum. Changing this parameter did not alter the decomposition behavior of `GaussPy` in this case.

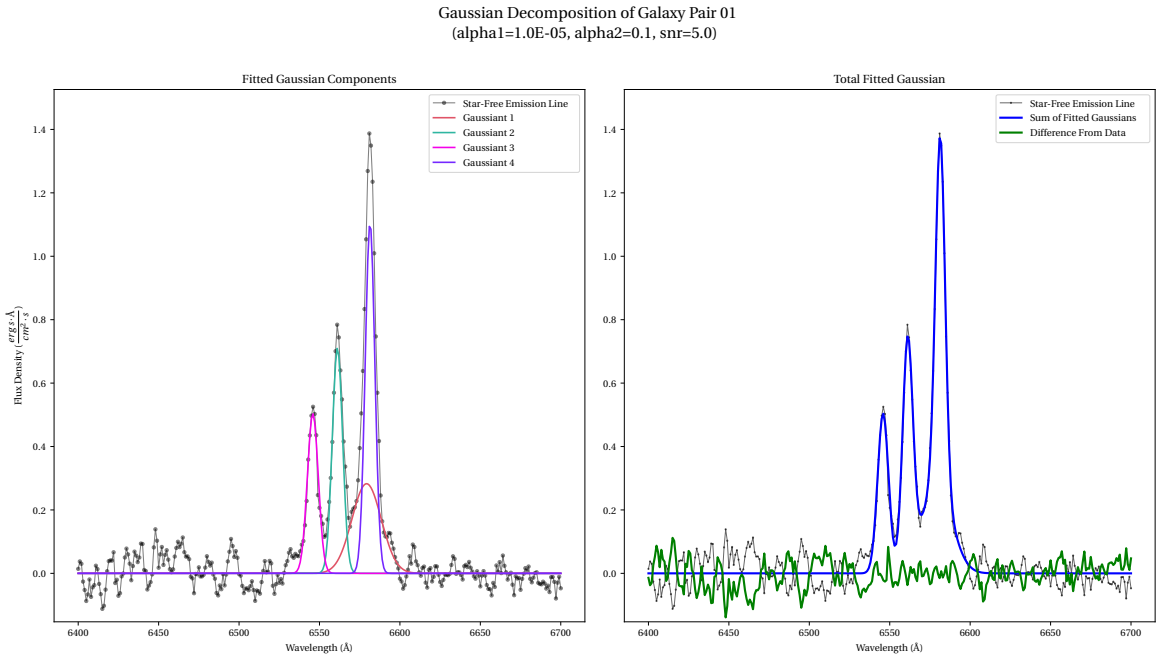

Figure 58: An example of a Gaussian decomposition that is more accurate to the data while not necessarily being likely. Notice the secondary component fitted to only one of the $[NII]$ lines.

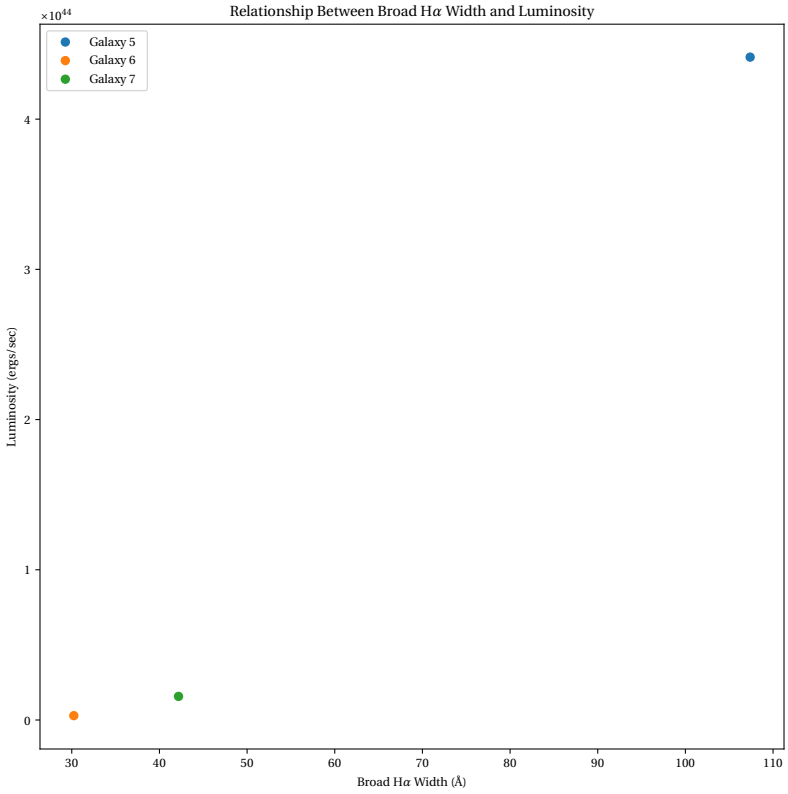

Figure 59: Relationship between width of broad Hα component and luminosity for spectra that demonstrate an Hα component with $\Delta v \geq 1000 \frac{km}{s}$

6 Appendix

6.1 Emission Line Ratios

6.2 Emission Line Widths

Table 1: Distances of Galaxies

Galaxy	Distance (Mpc)
1	71.1
2	75.1
3	22.0
4	18.2
5	40.9
6	24.6
7	16.6
8	14.6
9	32.8
10	16.2

Table 2: Flux Density Scaling: Least Square Method

Galaxy	Wavelength Min (Å)	Wavelength Max (Å)	Mean Emission Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Mean Absorption Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Least Square Scale Factor	Least Square Mean Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Least Square Star-Free Emission Line Flux STD ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Least Square Median Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)
1	6183.0	6743.0	3.368606	7.033626	0.47901	-0.00058	-0.00058	-0.04181
2	6226.0	6737.0	2.923814	42.659414	0.06856	-0.00076	-0.00076	-0.10087
3	6210.0	6810.0	33.637388	7.246678	4.63913	0.01908	0.01908	-3.78526
4	6189.0	6826.0	3.795688	8.613204	0.44033	0.00306	0.00306	-0.09882
5	6184.0	6799.0	17.777310	5.203517	3.42991	-0.07027	-0.07027	-3.98654
6	6207.0	6822.0	3.453967	85.541495	0.04026	0.00971	0.00971	-0.14957
7	6218.0	6818.0	18.850824	8.849536	2.13050	-0.00314	-0.00314	-1.52983
8	6194.0	6833.0	9.297444	10.941038	0.84918	0.00655	0.00655	-0.15218
9	6163.0	6804.0	3.744116	12.632402	0.29627	0.00148	0.00148	0.01973
10	6184.0	6824.0	11.724450	14.760037	0.79480	-0.00684	-0.00684	-0.35076

Table 3: Flux Density Scaling: Robust Least Square Method

Galaxy	Wavelength Min (Å)	Wavelength Max (Å)	Mean Emission Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Mean Absorption Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Scale Factor	Robust Least Square Mean Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Star-Free Emission Line Flux STD ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Median Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)
1	6183.0	6743.0	3.368606	7.033626	0.47437	0.03208	0.03208	-0.00858
2	6226.0	6737.0	2.923814	42.659414	0.06662	0.08202	0.08202	-0.01880

Continued on next page

Table 3: Flux Density Scaling: Robust Least Square Method

Galaxy	Wavelength Min (Å)	Wavelength Max (Å)	Mean Emission Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Mean Absorption Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Scale Factor	Robust Least Square Mean Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Star-Free Emission Line Flux STD ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Robust Least Square Median Star-Free Emission Line Flux ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)
3	6210.0	6810.0	33.637388	7.246678	4.12085	3.77493	3.77493	-0.01671
4	6189.0	6826.0	3.795688	8.613204	0.43015	0.09068	0.09068	-0.01164
5	6184.0	6799.0	17.777310	5.203517	2.65063	3.98472	3.98472	-0.06670
6	6207.0	6822.0	3.453967	85.541495	0.03867	0.14642	0.14642	-0.01181
7	6218.0	6818.0	18.850824	8.849536	1.95674	1.53456	1.53456	-0.01875
8	6194.0	6833.0	9.297444	10.941038	0.83617	0.14888	0.14888	-0.00683
9	6163.0	6804.0	3.744116	12.632402	0.29667	-0.00350	-0.00350	0.01469
10	6184.0	6824.0	11.724450	14.760037	0.77345	0.30835	0.30835	-0.03436

Table 4: Flux Density Scaling: Method Comparison

Galaxy	Wavelength Min (Å)	Wavelength Max (Å)	Method Mean Difference ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Method STD Difference ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)	Method Median Difference ($\frac{ergs \cdot Å}{cm^2 \cdot s}$)
1	6183.0	6743.0	0.03266	0.00010	0.03323
2	6226.0	6737.0	0.08278	0.00022	0.08207
3	6210.0	6810.0	3.75585	0.00736	3.76855
4	6189.0	6826.0	0.08762	0.00081	0.08718
5	6184.0	6799.0	4.05499	0.03784	3.91984
6	6207.0	6822.0	0.13671	0.00208	0.13776
7	6218.0	6818.0	1.53770	0.00186	1.51108
8	6194.0	6833.0	0.14233	0.00166	0.14535
9	6163.0	6804.0	0.00498	0.00007	0.00504
10	6184.0	6824.0	0.31519	0.00218	0.31640

Table 5: Properties of Potential Broad Hα components

Galaxy	Galaxy Distance (m)	Component Flux ($\frac{ergs}{cm^2 \cdot s}$)	Component Luminosity ($\frac{ergs}{s}$)	Component Velocity (km/s)	Is this a Broad Hα Component?
2	6.788491e+17	92.102277	5.333677e+42	560.044171	no
4	1.262042e+18	2204.716460	4.412760e+44	4909.197682	yes
5	7.590767e+17	40.061163	2.900713e+42	1382.198544	yes
6	5.122225e+17	476.719282	1.571773e+43	1925.642978	yes
7	4.505089e+17	14.283568	3.642951e+41	850.734414	no

References

J. A. Baldwin, M. M. Phillips, and R. Terlevich. Classification parameters for the emission-line spectra of extragalactic objects. , 93:5–19, February 1981. doi: 10.1086/130766.

G. M. Stirpe. Broad emission lines in active galactic nuclei. I. an atlas of H-alpha and H-beta profiles. , 85:1049, November 1990.

Robert R. Lindner, Carlos Vera-Ciro, Claire E. Murray, Snežana Stanimirović, Brian Babler, Carl Heiles, Patrick Hennebelle, W. M. Goss, and John Dickey. Autonomous Gaussian Decomposition. , 149(4):138, April 2015. doi: 10.1088/0004-6256/149/4/138.

G. Kriss. *Fitting Models to UV and Optical Spectral Data*, volume 61 of *Astronomical Society of the Pacific Conference Series*, page 437. 1994.

J. A. Baldwin. Broad emission lines in active galactic nuclei. *International Astronomical Union Colloquium*, 159: 80–95, 1997. doi: 10.1017/S0252921100039713.

Stephanie M. LaMassa, Tim M. Heckman, Andrew Ptak, Lucimara Martins, Vivienne Wild, and Paule Sonnentrucker. Indicators of Intrinsic Active Galactic Nucleus Luminosity: A Multi-wavelength Approach. , 720(1): 786–810, September 2010. doi: 10.1088/0004-637X/720/1/786.

M. Riener, J. Kainulainen, J. D. Henshaw, J. H. Orkisz, C. E. Murray, and H. Beuther. GAUSSPY+: A fully automated Gaussian decomposition package for emission line spectra. , 628:A78, August 2019. doi: 10.1051/0004-6361/201935519.