University of Crete
Department of Physics

A thesis submitted for the degree of

Bachelor of Science

In Physics

Measuring Galaxy Distances Using Cross-Correlation Methods

Author:
Giannis Markoudakis

Supervisors:
Dr. Andreas Zezas
Dr. Carolina Casadio

October, 2023

Contents

List of Figures

Abstract

We are interested in developing a tool that can accurately measure the redshift of distant galaxies, while also quantifying of the measurement's confidence. To accomplish that, we utilized the RVSAO package, which is a set of tools operating in the IRAF environment. xcsao, which is the basic task of RVSAO, implements the cross-correlation method to estimate the redshift of an object, using a template spectrum as reference. We have enriched the template list provided by RVSAO, by adding Sloan Digital Sky Survey III (SDSS-III) spectra as well as composite ones, covering a wide range of galaxy types with redshifts reaching up to $z = 1.814$. These spectra were pre-processed appropriately, in order to match the RVSAO template format. We have modeled a relation between the confidence of a distance measurement and the signal-to-noise (SNR) of the spectra, and also executed several simulations to thoroughly assess the method's capabilities. We confirm the efficiency and stability of the method for a plethora of galaxy types, such as different types of Active Galactic Nuclei (AGN) and quasars.

1 Introduction

1.1 Active Galactic Nuclei

Comprehending the evolution of galaxies, as well as the manner in which their emitted light interacts gravitationally with non-luminous particles, is of great cosmological importance. This understanding has the potential to answer fundamental questions, including the post-Big Bang universe's evolution and the elusive nature of dark matter as revealed through gravitational lensing and other methods. One gateway for probing the universe is via active galaxies, also known as active galactic nuclei (AGN) [6]. AGN represent galaxies capable of generating vast quantities of energy across the electromagnetic spectrum by accumulating mass onto a supermassive black hole (BH) located at the center of the galaxy. Figure 1 displays a simulation of mass accreting in the central BH of a spiral galaxy [9].

AGN can be further divided into classes according to their spectral characteristics (emission line width, radio emission, etc), the most prevalent being:

- Seyferts: Mostly spiral galaxies, emitting in X-rays, showing both narrow and broad emission lines.
- Radio galaxies: Mostly elliptical galaxies, with strong radio emission, showing both narrow and broad emission lines.
- Quasi-stellar radio sources, commonly known as quasars, with immense amount of luminosity, spanning from gamma rays to the radio.

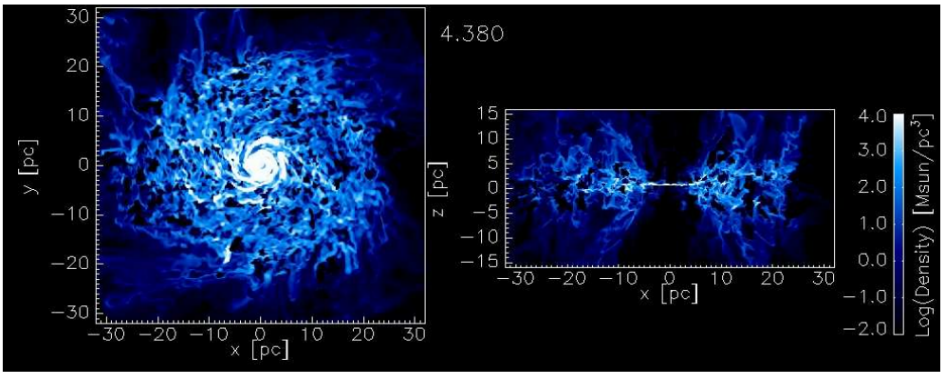

Figure 1: Face-on and edge-on density maps of a central toroidal structure resulting from gas infall towards a black hole (BH) in the center of a spiral galaxy [9].

1.1.1 Quasars

Among these astronomical objects, quasars stand out as particularly remarkable. Quasars emit prodigious amounts of energy that span a vast swath of the electromagnetic spectrum, all while being located at extreme distances from Earth. Their great distances are also designated alternatively their redshift, which can reach significant values, such as the recorded observation of up to z = 7.642 [15]. This redshift indicates that the radiation we receive from quasars has traveled across cosmic expanses and through vast stretches of time, allowing us to peer into the distant past of the universe. Given these characteristics, the study of quasars and AGN in general, is of primary importance in advancing our understanding of the universe's early epochs and its subsequent evolution.

1.2 Redshift & Radial Velocity

Accurately measuring the redshift of galaxies is of primary importance in the field of extragalactic astrophysics and cosmology. In physics, a redshift is the increase of the wavelength in electromagnetic radiation, which is a consequence of the Doppler effect, also known as Doppler shift. The redshift z of an object is defined as:

$$z = \frac{\lambda_{obs} - \lambda_{emit}}{\lambda_{emit}}, \tag{1}$$

where λ_{obs} is the observed photon wavelength, and λ_{emit} is the actual photon wavelength emitted from the object [12]. This shift is caused by the relative motion between the observer (Earth), and the object (galaxy) being observed, which originate from Doppler motions and/or the general expansion of the universe, also referred to as the Hubble flow. Exactly because of this flow the vast majority of the

galaxies appear to move away from us resulting in teh observation of longer wavelength radiation than emitted by the galaxy. Therefore, z can can be expressed using a similar quantity, the radial velocity u, using the equation:

$$1 + z = \sqrt{\frac{1 + \frac{u}{c}}{1 - \frac{u}{c}}}, \tag{2}$$

where c is the speed of light. For non-relativistic motions where $u << c$, Eq. 2 can be written as:

$$z = \frac{u}{c}. \tag{3}$$

Radial velocity is also related to the distance d of an object using the Hubble law:

$$u = H_0 d, \tag{4}$$

where H_0 is the Hubble constant. Consequently, Using Eq. 3 and the Hubble law, we can also estimate the distance d of the object emitting the light, which can be expressed as follows:

$$d = \frac{cz}{H_0}. \tag{5}$$

Redshifts and radial velocities can provide a plethora of information for the galaxies observed, such as their distance and relative motions, quantities that are essential for understanding the evolution of the universe. Consequently, we aimed to characterize a tool that can automatically estimate the radial velocity of a galaxy, given its spectrum.

1.3 Cross-Correlation Method

For this task, we implemented the cross-correlation technique, according to which, given a template spectrum in rest frame ($z = 0$) as reference, the radial velocity of an object spectrum can be estimated. Initially, the two spectra, after being discretely resampled to the same number of bins and wavelength range, are normalized in order to ensure that spectral lines for both spectra will be considered during the process. Subsequently the two spectra are Fourier transformed. The product of the transforms produces the cross-correlation function, the highest peak of which determines the radial velocity of the object spectrum. An example of the cross-correlation function is showcased in Figure 2. In more detail, The cross-correlation function is calculated by shifting the input spectrum relative to the template spectrum over a range of possible velocity shifts (redshifts). At each shift, the two spectra are multiplied point by point, and the resulting product is summed. This process generates a correlation curve as a function of velocity. The redshift (velocity) of the target object is determined by identifying the peak of the

Figure 2: Cross-correlation function profile. The horizontal axis represents the radial velocity or redshift. The location of the peak designates the result redshift of a spectrum.

cross-correlation curve. This peak corresponds to the radial velocity that maximizes the similarity between the input and template spectra. Tonry and Davis [14], who utilized this method to estimate radial velocities of galaxies, defined the R value, originating from their r-statistic, which can represent the confidence of a redshift measurement. Of course, the higher the similarity between the object and template spectra, the higher the R value will be. Therefore, it is vital to have an adequate number of template spectra, covering a wide range of galaxy types and their spectral properties, in order to obtain an accurate redshift measurement for all types of galaxies.

1.4 RVSAO

Our pipeline is based on IRAF's [13] contributed package called RVSAO, which is comprised of several tasks that can perform actions like removing the continuum, shifting spectra, and identifying spectral lines. xcsao is the heart of the package, responsible for performing the cross-correlation. Tasks included in RVSAO provide a wide variety of adjustable parameters in order to optimize the process for the needs of each project [8]. Some of these include:

- the specification of the wavelength interval where the cross-correlation process will be applied. If not specified, the shared region between the input spectrum and the current template is automatically selected.
- The option to zero-pad the spectral transforms in Fourier space. Fine-tuning

this parameter can occasionally yield more precise results, depending on the characteristics of the spectra under examination.

- Possible renormalization of the input spectrum.
- An interpolation mode for fitting and, potentially, removing the continuum from either spectra.

While these parameters are initially set, slight adjustments are made based on the specific properties of the spectra analyzed.

1.5 Sloan Digital Sky Survey

For the purpose of our work, we used galaxy spectra from the latest release of the Sloan Digital Sky Survey III (SDSS-III), which aims to probe the nature of dark matter by tracing the large-scale mass structure of the universe using massive galaxies and quasars [2] [5]. Their catalog contains moderate-resolution optical spectra, along with important information such as redshift, signal-to-noise ratio (SNR), and nuclear photometry magnitudes for each filter. In general, SDSS is an survey that has made major contributions to the field of astronomy and cosmology over the years. It was initiated with the purpose of constructing a comprehensive 3D map of the universe, mapping the distribution of stars, distant galaxy clusters and other celestial objects.

2 Data Overview

2.1 RVSAO Template Spectra

RVSAO itself offers a wide range of spectra, spanning from stars to galaxies, with varying spectral characteristics. It also provides artificial absorption and emission line spectra, such as those displayed in Figure 3. Some of the artificial spectra predominantly feature only one or two dominant lines were discarded from the template list to avoid potentially misleading results, since these individual lines could be associated with any emission line in the object spectrum. RVSAO can read only linearly binned spectra, the bin size of each one being stored in their header, along with other information for the spectra. The header can also host some special keywords that can regulate how the spectra will be manipulated during the cross-correlation process, like whether the continuum will be subtracted or divided (if removed at all), whether narrow emission lines will be removed, etc. For the purpose of measuring radial velocities, the SNR of the original spectrum must be preserved, so subtracting the continuum is preferred for all spectra. The other

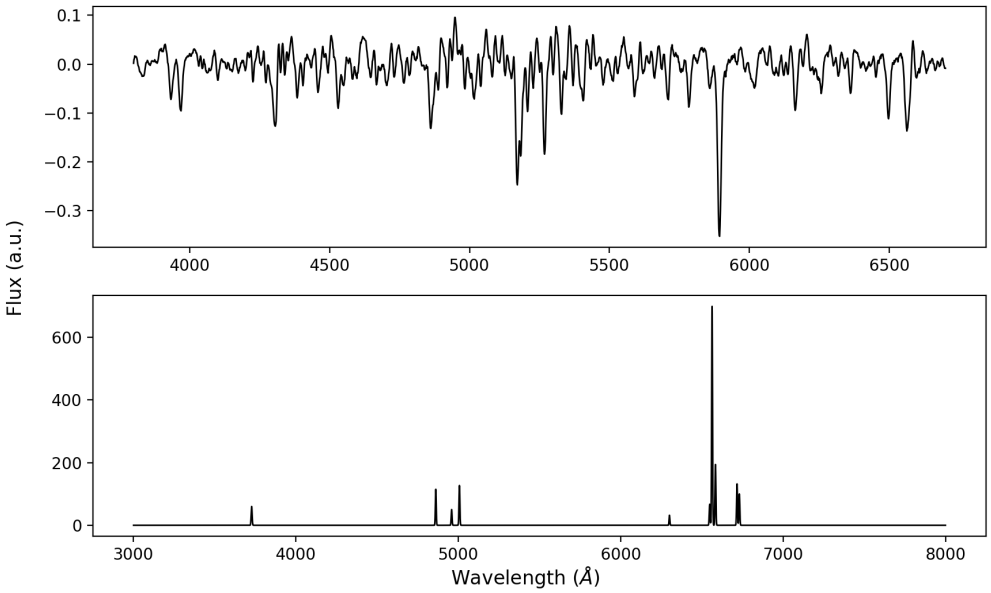

Figure 3: Artificial template spectra provided by RVSAO. The top panel displays a spectrum dominated by absorption lines. The bottom panel displays a spectrum dominated by known emission lines.

keywords are adjusted discretely for other spectra, depending on the properties of the galaxy at hand.

2.2 SDSS-III Spectra

We utilized 63 spectra sourced from the SDSS-III, encompassing a broad spectrum of galaxy types, such as Seyfert galaxies and quasars, spanning various redshifts, reaching values up to $z = 1.814$. SDSS-III employs logarithmic binning to store their digital spectra, so the spectral axis is represented in logarithmic wavelength units instead of linear wavelength units. Two of these spectra are displayed in Figure 4. Important quantities and properties of the spectra are also provided. Some of those are the following:

- Signal-to-noise ratios (SNRs), each computed within the corresponding interval of the different filters (ugriz) used by SDSS.
- Median signal-to-noise ratio (SNR) computed for the entire digital spectrum.
- Photometric magnitudes within a 3 arcsecond aperture (fiber magnitudes) for each one of the filters used by SDSS.

The most important features of the spectra we will be utilizing are the following:

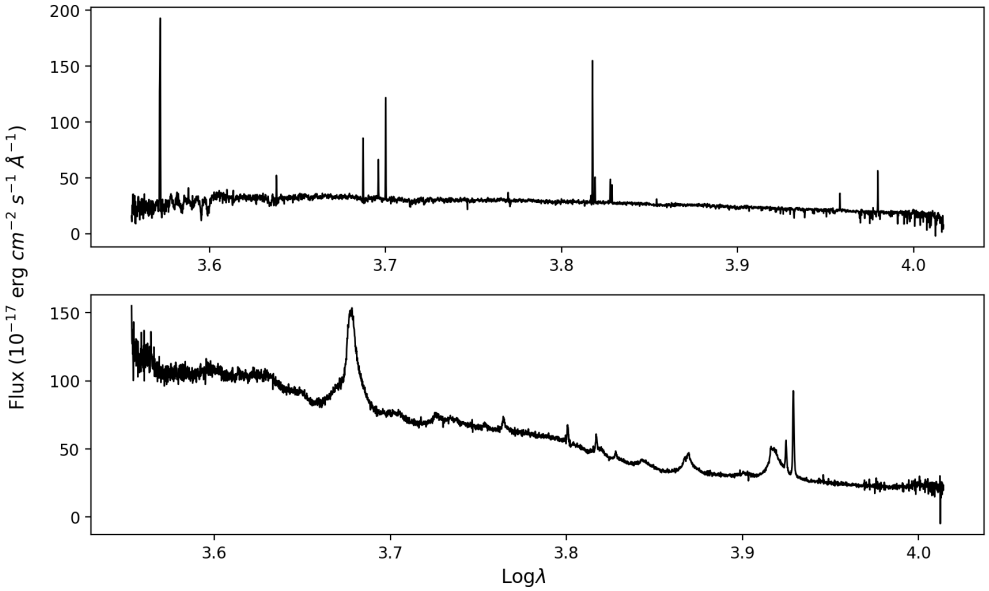

Figure 4: SDSS-III spectra displayed in logarithmic spectral axis, where the wavelength is expressed in Angstrom. The top panel depicts a spectrum of an AGN with $z = 0.0007$. The bottom panel depicts a spectrum of a quasar with $z = 0.7$.

- Flux density, expressed in $10^{-17} erg/cm^2/s/\mathring{A}$ units.
- Logarithmic wavelength $Log\lambda$, where λ is expressed in Angstrom ($\mathring{A}$) units.
- Inverse variance of flux, which is calculated through the standard deviation of the spectrum.

2.3 Composite Spectra

In addition to the RVSAO and SDSS-III template spectra, we utilized composite galaxy spectra generated from an extensive collection of SDSS spectra. Given an adequate number of sample spectra, a composite median spectrum can be constructed [4]. This process involves shifting the sample spectra to the rest frame, normalizing them, and finally stacking them, to produce a median spectrum. In particular, we used a published composite quasar spectrum and a composite Seyfert 1 spectrum, both derived from SDSS data [1] [11].

The quasar sample used to generate the median composite spectrum consists of 2200 quasar spectra, spanning a wide redshift range of $0.044 \leq z \leq 4.789$. The final spectrum covers a rest-wavelength range from 800 to 8555 $\mathring{A}$, and is mainly dominated by emission lines. The composite quasar spectrum is displayed in Figure 5.

The composite Seyfert 1 spectrum was produced from 10112 galaxies, spanning a redshift range of $0 \leq z \leq 0.793$. The generated rest-frame spectrum spans from 2122 to 8555 $\mathring{A}$, and is displayed in Figure 5.

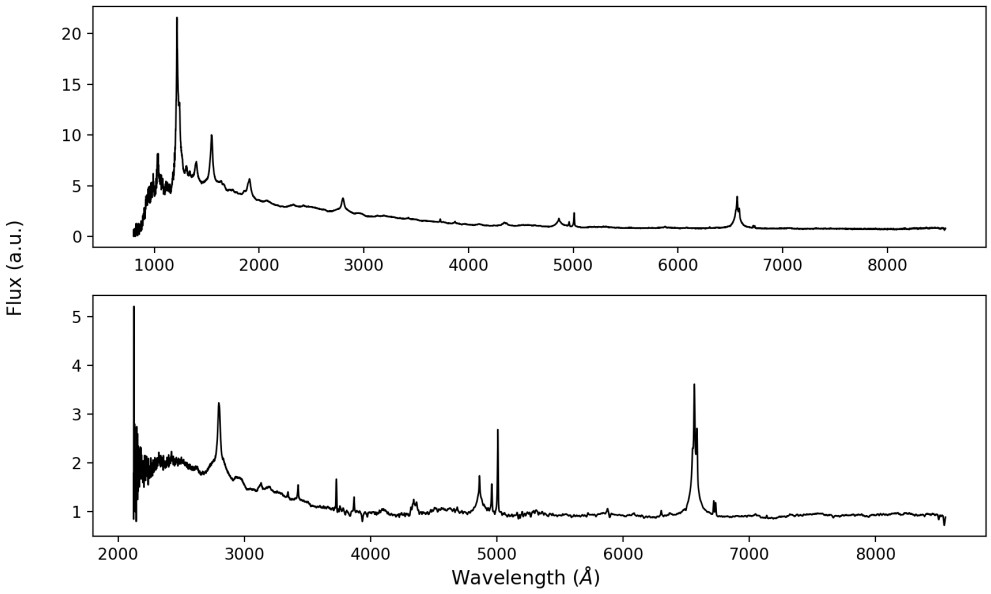

Figure 5: Composite spectra generated from SDSS sample spectra. The top panel depicts a composite spectrum of a quasar produced from 2200 spectra with redshift range $0.044 \leq z \leq 4.789$. The bottom panel depicts a spectrum of a Seyfert 1 produced from 10112 spectra with redshift range $0 \leq z \leq 0.793$.. The vertical axis in both spectra is expressed in arbitrary units.

3 Pipeline Overview

3.1 Pre-Processing Spectra

3.1.1 Resampling

Since RVSAO's xcsao, which implements the cross-correlation technique exclusively accepts spectra with linear wavelength binning, the spectra obtained by SDSS-III cannot be directly utilized, since the are logarithmically binned. Therefore, the spectra need to be resampled first, converting their wavelength axis into a linear format as mandated by the RVSAO format. In order to efficiently execute this resampling process, we utilized a tool [3], that preserves the integrated flux $F_\nu d\nu$. We arbitrarily chose the linear bin size equal to $1\mathring{A}$, and also updated

the header of the spectra. A spectrum before and after the resampling process is showcased in Figure 6. The flux variance was also resampled accordingly, and was

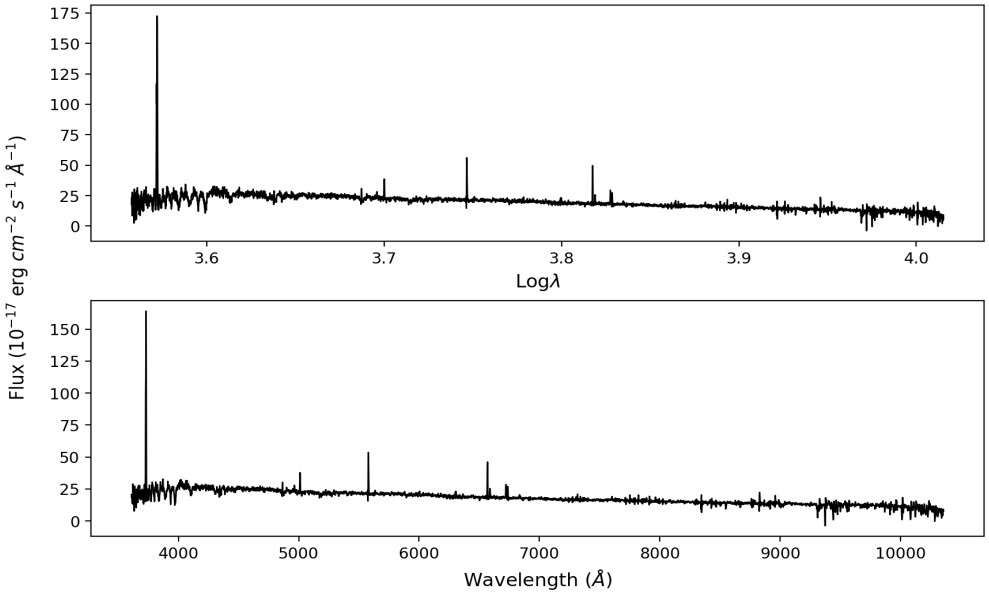

Figure 6: SDSS-III spectrum before and after the resampling process. The top panel showcases the original spectrum with logarithmic binning. The bottom panel showcases the same spectrum after it is resampled in linear binning.

subsequently converted to flux error using the equation:

$$error = \frac{1}{\sqrt{ivar}}, \tag{6}$$

where *ivar* is the resampled inverse variance. As a consequence, the SDSS-III spectra now have the desired format, and can be utilized by RVSAO, both as input and template spectra.

3.1.2 Shifting Spectra to the Rest Frame

RVSAO's xcsao is capable of estimating the redshift z of an input spectrum by cross-correlating it with a template spectrum. However, when the template spectrum itself has redshift $z_{temp} \neq 0$, then the resulting z will deviate from the actual value by z_{temp}. Therefore, it is imperative that all template spectra are shifted to rest frame (i.e. $z_{temp} = 0$). The RVSAO templates and the composite spectra are already in rest frame, but additional SDSS-III spectra we introduce as templates require redshift removal before they are utilized as templates. To address this, we

developed a pipeline to shift the observed SDSS-III spectra. In order to do that, the spectral axis needs to be shifted first as follows:

$$\lambda_{rest} = \frac{\lambda_{obs}}{1+z}, \tag{7}$$

where z is the redshift of the spectrum, and λ_{rest} and λ_{obs} are the rest and observed wavelengths, respectively. After the rest spectral axis is obtained, an appropriate modification to the flux, caused by the redshift, needs to be made as well. This modification can be made using the expression:

$$F_{rest} = F_{obs}(1+z)^3, \tag{8}$$

where F_{rest} and F_{obs} are the rest and observed fluxes, respectively [10]. The same procedure was consistently applied for all SDSS-III spectra, and their header was updated to indicate that the spectrum is shifted to $z = 0$. Figure 7 depicts a quasar spectrum before and after it was shifted to rest frame. All of the pre-processed spectra were subsequently cross-correlated with RVSAO templates to confirm the effectiveness of the procedure.

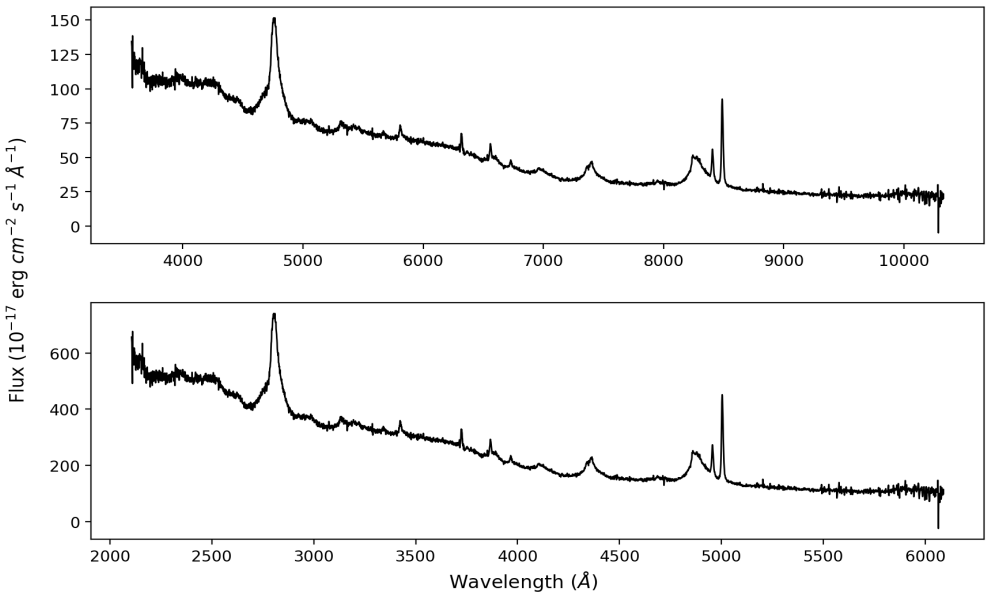

Figure 7: SDSS-III quasar spectrum before and after redshift removal. The top panel showcases the observed spectrum. The bottom panel showcases the same spectrum after it is shifted to rest frame.

3.2 Spectra Generation

In order to perform statistical tests to probe the capabilities of xcsao, including the efficiency of the method under increasing spectral noise, or the parameters the R value depends on, we developed a pipeline to generate new spectra of different SNR. These new spectra are derived from the existing SDSS-III spectra at our disposal. The process involves taking a spectrum, denoted as A, and creating a new spectrum, B, by sampling flux values from a Gaussian distribution. The mean value μ of this distribution is set to the flux value of A. The standard deviation σ of the distribution is chosen to be equal to the flux error multiplied by a positive factor α, which depends on the desired amount of noise to be added to the new spectrum. For instance, setting $\alpha = 1$, replicates the input spectrum as faithfully as possible, while setting $\alpha = 2$ doubles the noise. By iterating through all the flux points (bins) of spectrum A, we generate a unique yet similar spectrum. Figure 8 illustrates generated spectra derived from an AGN spectrum, illustrating different amounts of noise (i.e. different α values). By creating these new spectra, which

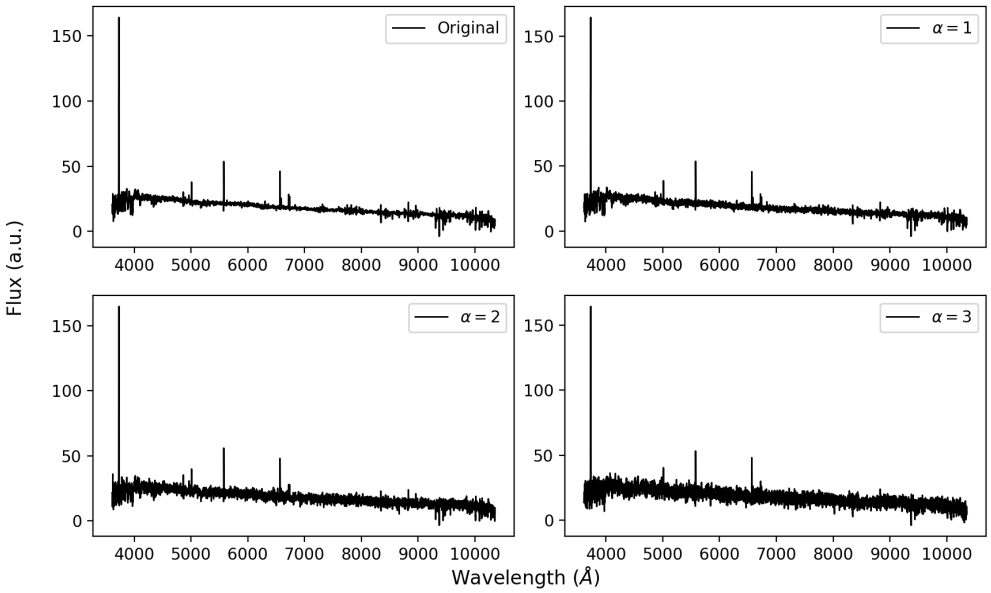

Figure 8: Generated spectra with different amounts of noise, i.e.different α values (see section 3.2). The top left panel depicts the original AGN spectrum from SDSS-III used to produce the others. The top right panel showcases a generated spectrum with regular noise ($\alpha = 1$), whereas the bottom left and bottom right showcase spectra with double ($\alpha = 2$) and triple noise ($\alpha = 3$), respectively.

are representative of a specific spectrum but incorporate increasing noise levels, we can conduct a range of statistical simulations. These simulations enable us to

assess the effectiveness of the cross-correlation method under conditions of both high and low Signal-to-Noise Ratio (SNR) spectra. Our specific goal is to model the behavior of essential parameters utilized by RVSAO, including the 'R' value, resulting radial velocity, and radial velocity error, across a range of spectra of varying spectral types and SNR.

4 Results

4.1 Accuracy Among Templates

Our primary objective is to clarify that the cross-correlation performed by RVSAO's xcsao provides accurate results. In order to to accomplish that, we will cross-correlate a spectrum, with known redshift, with the RVSAO templates, including the composite spectra discussed in the previous section. Ultimately, the radial velocity and associated error given by the best template (i.e. the template that gives the highest R value), should sufficiently reproduce the actual radial velocity of the target spectrum. In addition to that, the results of the other templates should not deviate considerably from the ones given by the best template, as long as their results are confident. For that purpose, and for all the following test we will execute, we used a spectrum of the 3C449 radio galaxy [7], depicted in Figure 9, which has $z = 0.017$. The results produced by xcsao for all the current

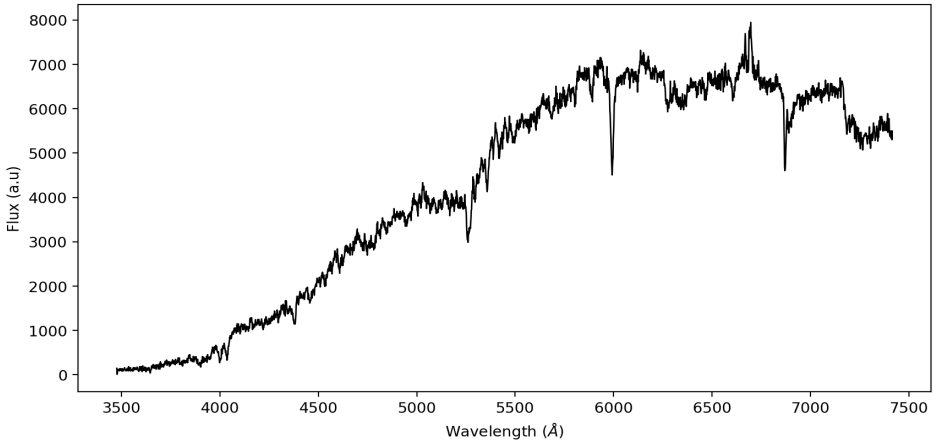

Figure 9: Spectrum of 3C449 radio galaxy with $z = 0.017$.

templates are showcased in Figure 10. The best template, with $R = 33.27$, yields a radial velocity measurement of $u = 5107.41\text{kms}^{-1}$ which, using Eq. 3, translates to a redshift $z = 0.017025$. The corresponding error is $\delta u = 10.825\text{kms}^{-1}$, expressed in velocity units, or $\delta z = 3.6 \times 10^{-5}$ for the redshift. Examining the top panel of

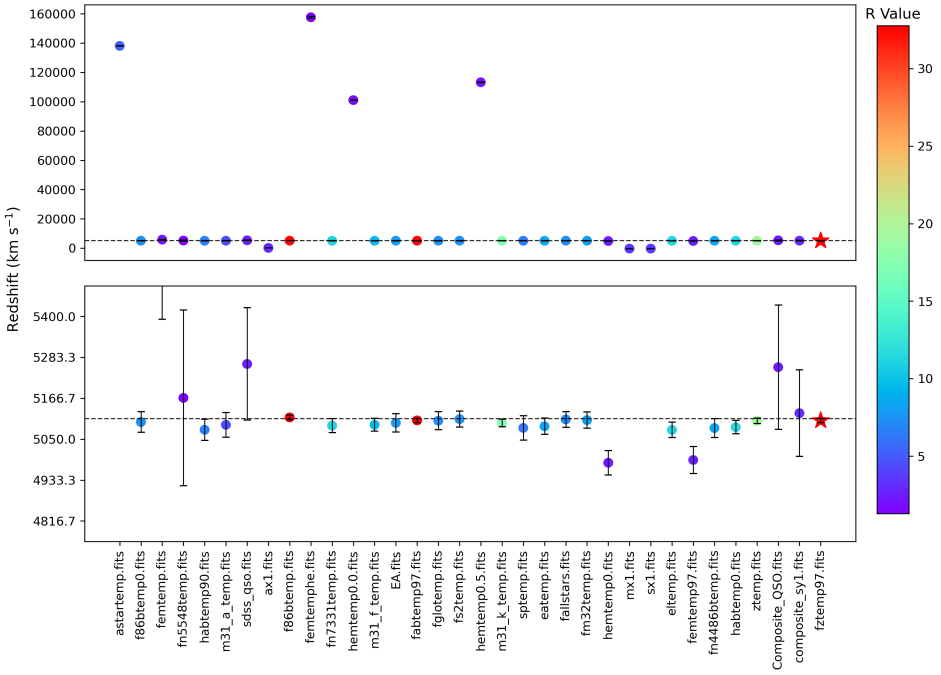

Figure 10: Results of an AGN spectrum cross-correlated with RVSAO and composite templates. The vertical axis represents the redshift, expressed in kms^{-1}, along with the corresponding error, and the horizontal axis features the templates. The colors indicate the confidence of the measurement (R value), while the best template is designated with a star symbol. The bottom panel showcases the same graph, limited in the region where the majority of the measurements are.

Figure 10, we observe that, besides a few outliers, the majority of templates give results within a narrow radial velocity region. From the bottom panel, it becomes evident that the template spectra which give satisfactory results have a higher R value, as indicated from the colorbar. On the other hand, the few outlier spectra present a very low R value and substantial error, which suggests that they are a poor match to the specific target spectrum.

4.2 Noise sensitivity

Our next task is to investigate how RVSAO's parameters behave with increasing levels of noise. Despite the fact that the cross-correlation method provides satisfactory results for the high SNR spectra from SDSS-III, we ultimately aim to test its consistency even with noisier spectra. We are chiefly interested in the consistency of the redshift measurement, although we anticipate that, with the introduction of more noise, the R value will decrease, and the error will increase. To conduct

this analysis, we employed the spectra generation pipeline discussed in the previous section to create 500 new spectra. These spectra were categorized into five groups, each containing 100 spectra. Each group of spectra was associated with a specific level of noise (as designated by the factor α, which takes the values of 1, 2, 3, 5, 10, respectively for each group) relative to the original spectrum. All 500 spectra were independently cross-correlated with the available RVSAO templates and composite spectra. The results for each group were aggregated and are presented in the histograms of Figure 11. Upon inspection of the histograms in the first column in the figure, we observe that, for spectra with regular amount of noise, the max R value is centered around the value of 32, whereas it decreases to approximately 11, for 10 times noisier spectra. Accordingly, the redshift error is 11kms^{-1} on average for high SNR spectra, and increases to about 30kms^{-1} for the noisiest spectra of the sample. Conversely, while the standard deviation of the redshift gradually increases as more noise is added to the spectra, the majority of the measurements remain closely clustered around 5100kms^{-1} for all 500 samples.

4.3 Best Template Consistency

While we found that the increased noise does not significantly affect the precision of the measured redshift, it is also useful to check whether the best template, for a specific input spectrum, remains consistent as the SNR of the spectrum diminishes. Employing the same 500 spectra generated previously, we followed the same procedure as before, but this time we focused solely on the best template, as suggested from the xcsao pipeline. The statistics for the best template from the analysis of the whole sample are displayed in Figures 12, 13. Upon initial inspection of these figures, it becomes evident that the best template does not remain the same for every spectrum within each group. In more detail, as the noise remains low ($\alpha = 1$), there are two candidate best templates, with "fztemp97" getting the vast majority of votes. However, as more noise is added to the spectra, the second candidate, "f86btemp", has increasingly higher preference among the samples, ultimately surpassing "fztemp97" for spectra with 5 times the amount of the initial noise ($\alpha = 5$). Moreover, as the SNR decreases, a minority of spectra exhibit higher R values for different templates compared to the initial two. Consequently there are a total of four candidate templates in the last panel of Figure 13 ($\alpha = 10$). Nevertheless, we consider their influence on the results to be insignificant, since they each of them receives less than 10% of the votes.

4.4 RVSAO Quantities Dependency on SNR

Finally, we aim to model how the R value, the redshift, and its associated error depend on the SNR of the input spectra. While Figure 11 provides an initial

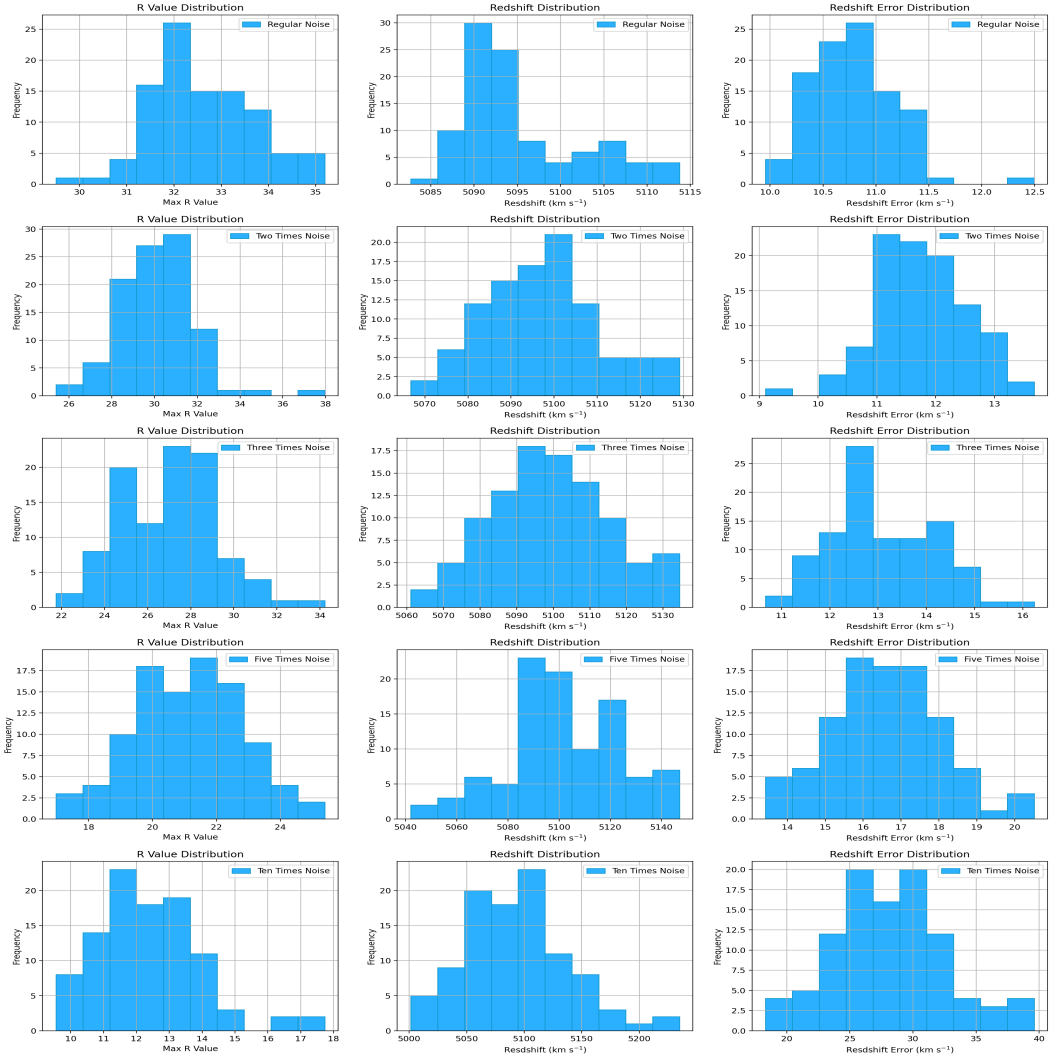

Figure 11: Histograms displaying the frequency of RVSAO results for 5 groups of spectra. Each row shows a group of 100 generated spectra with a certain amount of noise, compared to the original spectrum: Regular noise ($\alpha = 1$), 2 times noise ($\alpha = 2$), 3 times noise ($\alpha = 3$), 5 times noise ($\alpha = 5$), and 10 times noise ($\alpha = 10$). The columns depict the distribution of the R value, the redshift (expressed in velocity units), and the redshift error, given by the best template, for each spectrum.

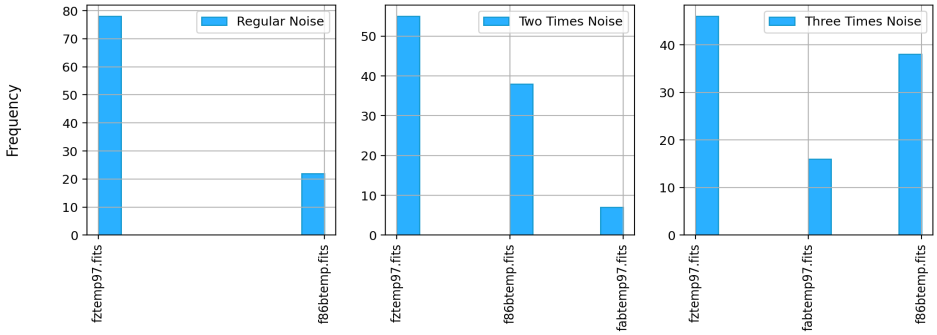

Figure 12: Histograms displaying the frequency of RVSAO best templates for 3 groups of spectra. Each histogram showcases a group of 100 generated spectra with a certain amount of noise, compared to the original spectrum: Regular noise ($\alpha = 1$), 2 times noise ($\alpha = 2$), and 3 times noise ($\alpha = 3$).

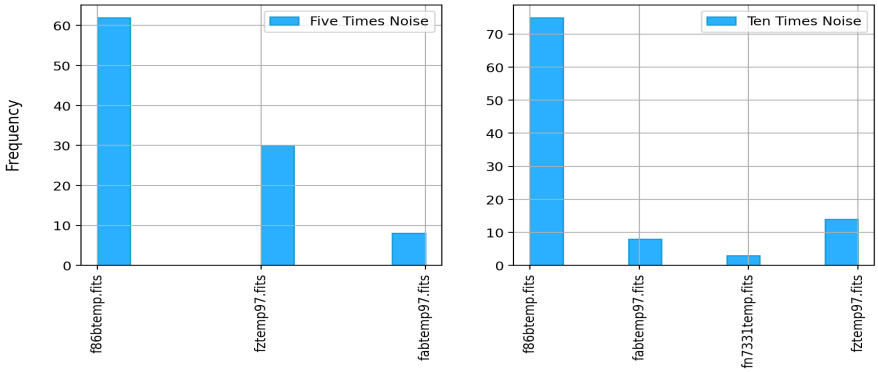

Figure 13: Histograms displaying the frequency of RVSAO best templates for 2 groups of spectra. Each histogram showcases a group of 100 generated spectra with a certain amount of noise, compared to the original spectrum: 5 times noise ($\alpha = 5$), and 10 times noise ($\alpha = 10$).

understanding of how RVSAO parameters are expected to behave, noise itself is an abstract concept. Therefore, we will convert it into SNR, a widely used and easily calculable metric for spectra. For each group of generated spectra, we computed the SNR (in the same spectral region, as they essentially represent the same spectrum with varying levels of noise) for all 100 spectra. Subsequently, we calculated the mean SNR value of all the spectra in a group, and assigned the mean value in each group. Afterwards, all generated spectra were consecutively used as input in the xcsao pipeline. More specifically, we calculated the median value of the R values, as estimated from the best template, for each group. In addition, the mean redshift, and the standard deviation were also computed for each group.

Finally, we determined the mean redshift error for each group. Using these results, we generated the scatter plots shown in Figure 14. From the first plot in this Figure, we observe that the median R value consistently decreases as noise levels increase. In particular, it starts at approximately 9 for spectra with SNR above 40 and drops below 8 for spectra with SNR lower than 20. On the other hand, a similar, reversed, trend is observed in the second scatter plot, depicting how much redshift measurements deviate from each other as SNR is decreased. Specifically, the standard deviation is approximately 10kms^{-1} for spectra with SNR above 50, and greater than 70kms^{-1} for low SNR spectra. The panel depicting the redshift dependency on SNR, suggests that the estimated redshift is not very sensitive on the noise of the spectrum, whereas the error increases from 10kms^{-1} to about 85kms^{-1}, as more noise is introduced.

5 Discussion & Future Work

In general, every statistical test performed, explicitly confirms the efficiency of the xcsao pipeline, which, along with the redshift estimation, provides errors and a confidence metric.

Firstly, we confirmed the accuracy of redshift estimation through cross-correlation, as the standard deviation between the computed and actual redshifts is approximately $\delta z = 0.001$. We also observed that the vast majority of template spectra yield satisfactory results, besides a few outlier instances. As previously mentioned, if the template has major differences in its spectral properties with the object spectrum, essentially meaning that there are no emission or absorption features that can be used in the cross-correlation process, then xcsao will provide unsatisfactory results. However, as shown in Figure 10, xcsao's inaccurate redshift measurements are accompanied by a relatively low R value, which indicates the poor confidence of the specific result. On the other hand, the best fitting templates generally give reliable results, paired with a relatively high R value. This

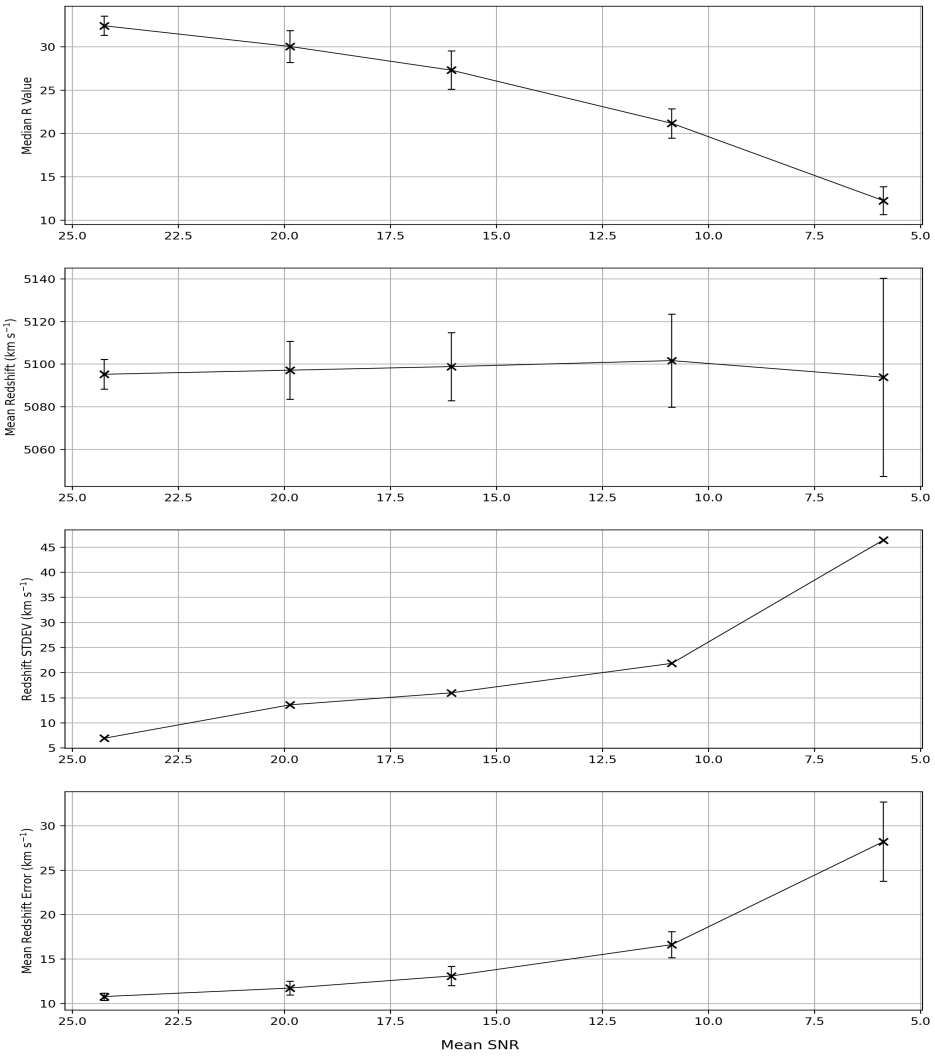

Figure 14: Scatter plots showcasing the dependency of RVSAO results on signal-to-noise ration (SNR). The vertical axis of the top panel represents the median maximum R value, and the error bars represent the standard deviation of the maximum R value. The second panel depicts the relation between the mean redshift and the SNR. The error bars represent the redshift standard deviation, which is also explicitly plotted in the third panel. The bottom panel showcases how the mean redshift error behaves as SNR changes. The error bars represent the standard deviation of the error.

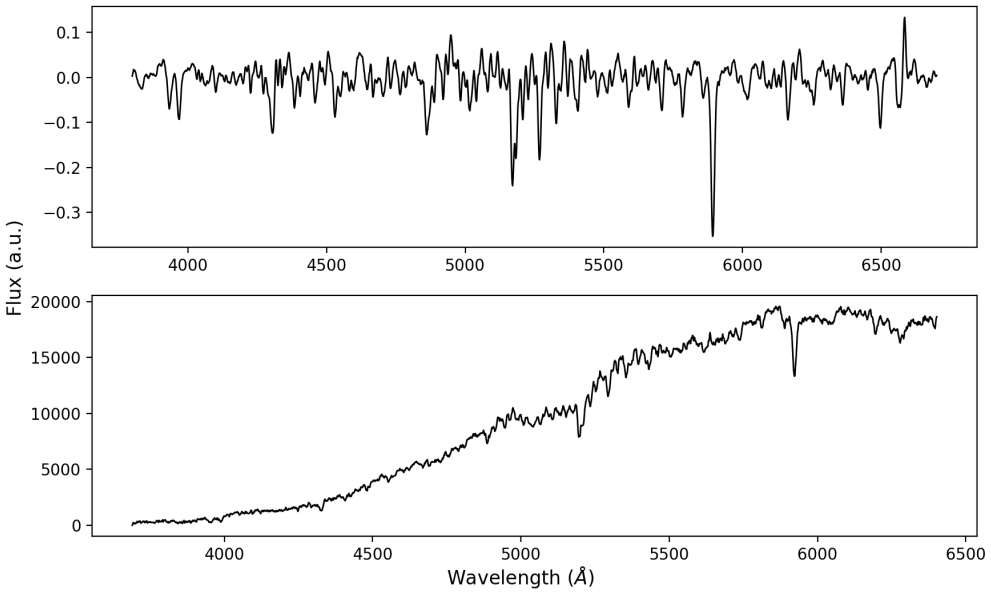

Figure 15: Two RVSAO templates dominated by absorption lines. The top panel displays the "fztemp97" template, which is an artificial, continuum subtracted, spectrum with a few distinct absorption lines. The bottom panel depicts the "f86btemp" template, which represents a galaxy spectrum dominated by absorption lines.

test validates the general reliability of the method for relatively high SNR spectra, thus indicating that the outlier templates are not a good spectral matches for the specific target spectrum.

Next, we investigated the performance of xcsao for noisier spectra. By generating spectra and cross-correlating them with the available templates, we produced the histograms of Figure 11. The behaviour of the R value and the redshift error were expected, since noisier spectra make the cross-correlation process more challenging and less accurate. Essentially, the distinct spectral lines of the input spectrum are not easily distinguishable, thus rendering the matching between input and template spectrum more challenging. Despite that, we also observed that the redshift estimation remained nearly equally reliable throughout all noise levels. This result reassures that the cross-correlation technique performs equally well on both high and low noise spectra when estimating redshifts or radial velocities, which is the fundamental goal.

Furthermore, Figures 12, 13 demonstrated that the best template does not remain consistent throughout different amounts of noise. However, all of these four templates are dominated by absorption lines as showcased in Figure 15, where the two most voted templates are depicted. Despite the difference in their continuum,

the two specific templates of the figure, display the same characteristic lines that also agree with the object spectrum, which verifies that the best template choice is not stochastic. It is important to note that for most other SDSS-III spectra tested (AGN and quasars included), the best template remained consistent even with increasing noise. Therefore, we deduce that the best template choice depends on the spectral properties of each specific spectrum.

Lastly, Figure 14 gave us a perspective of how RVSAO parameters behave with lower signal-to-noise spectra. More specifically, the R value, which provides a metric of the measurement's confidence, predictably decreases as SNR decreases. Despite that, the decline is not abrupt, so R value still remains adequately high for spectra with SNR equal to approximately 10. Consequently, we verify that the cross-correlation process can give a confident result on both high and low noise spectra. On the other hand, the redshift, which is the primary result, remains roughly consistent for all levels of noise. This confirms the reliability of the method even for noisy spectra, as long as there are representative templates that adequately resemble the object spectrum. Additionally, the redshift's standard deviation and error keep increasing as the R value decreases. Therefore, for low signal-to-noise spectra, an average of many cross-correlation runs can potentially provide a more accurate result, due to the high stochasticity accompanying noisier spectra.

All in all, the statistical simulations performed confirm the efficiency of the xcsao's cross-correlation technique in estimating the redshift of galaxies for both high and low signal-to-noise spectra. The pipeline was verified on other spectra provided by SDSS-III, and similar results were obtained. With the final addition of the 63 SDSS-III spectra to the template list, we enrich the sample and highlight the reliability of the method for a variety of galaxy types, ranging from nearby ellipticals to distant AGN and quasars.

However, the future addition of more spectra, can possibly be beneficial, as the ultimate goal is to obtain the most accurate result possible. Alternatively, the currently available template spectra can be aggregated to produce composite spectra, in order to cover a wider range of spectral types.

6 Conclusion

We have performed a thorough study of the cross-correlation method utilized by xcsao, and can therefore confidently verify the efficiency of our pipeline to estimate radial velocities of galaxy spectra in a wide redshift range. More specifically:

- We have confirmed that xcsao provides sufficiently accurate results for the vast majority of templates.

- We have observed that the redshift results remain consistent even if the amount of noise added in the spectrum is drastically increased.
- We have developed relations between the essential RVSAO parameters and the signal-to-noise ratio (SNR) of the input spectra.

With the addition of the SDSS-III and composite spectra, the template list finally consists of more than 100 spectra, with a wide variety of spectral properties. All in all, this tool can be utilized for a plethora of surveys in the foreseeable future, providing a stable way of estimating galaxy redshifts.

References

[1] Daniel E Vanden Berk et al. "Composite quasar spectra from the sloan digital sky survey". In: *The Astronomical Journal* 122.2 (2001), p. 549.

[2] Adam S Bolton et al. "Spectral classification and redshift measurement for the SDSS-III baryon oscillation spectroscopic survey". In: *The Astronomical Journal* 144.5 (2012), p. 144.

[3] AC Carnall. "SpectRes: a fast spectral resampling tool in Python". In: *arXiv preprint arXiv:1705.05165* (2017).

[4] László Dobos et al. "A high-resolution atlas of composite Sloan Digital Sky Survey galaxy spectra". In: *Monthly Notices of the Royal Astronomical Society* 420.2 (2012), pp. 1217–1238.

[5] Daniel J Eisenstein et al. "SDSS-III: Massive spectroscopic surveys of the distant universe, the Milky Way, and extra-solar planetary systems". In: *The Astronomical Journal* 142.3 (2011), p. 72.

[6] Lei Hao et al. "Active galactic nuclei in the sloan digital sky survey. I. Sample selection". In: *The Astronomical Journal* 129.4 (2005), p. 1783.

[7] KI Kellermann, IIK Pauliny-Toth, and PJS Williams. "The spectra of radio sources in the revised 3C catalogue". In: *Astrophysical Journal, vol. 157, p. 1* 157 (1969), p. 1.

[8] Michael J Kurtz and Douglas J Mink. "Rvsao 2.0: Digital redshifts and radial velocities". In: *Publications of the Astronomical Society of the Pacific* 110.750 (1998), p. 934.

[9] Hagai Netzer. "Revisiting the unified model of active galactic nuclei". In: *Annual Review of Astronomy and Astrophysics* 53 (2015), pp. 365–408.

[10] Bradley M Peterson. *An introduction to active galactic nuclei*. Cambridge University Press, 1997.

[11] Nihan Pol and Yogesh Wadadekar. "Seyfert 1 composite spectrum using SDSS Legacy survey data". In: *Monthly Notices of the Royal Astronomical Society* (2016), stw2763.

[12] Peter Schneider. *Extragalactic astronomy and cosmology: an introduction*. Springer, 2014.

[13] Doug Tody. "The IRAF data reduction and analysis system". In: *Instrumentation in astronomy VI*. Vol. 627. SPIE. 1986, pp. 733–748.

[14] John Tonry and Marc Davis. "A survey of galaxy redshifts. I-Data reduction techniques". In: *Astronomical Journal, vol. 84, Oct. 1979, p. 1511-1525.* 84 (1979), pp. 1511–1525.

[15] Feige Wang et al. “A luminous quasar at redshift 7.642”. In: *The Astrophysical Journal Letters* 907.1 (2021), p. L1.