UNIVERSITY OF COPENHAGEN
FACULTY OR DEPARTMENT

Master thesis

Maria Hammerstrøm

Establishing the host galaxy contribution to spectroscopic data of the nuclear emission of active galactic nuclei

Supervisor: Marianne Vestergaard

Submitted on: 31 May 2017

Abstract

In order to estimate the mass of supermassive black holes found in the centres of active galactic nuclei (AGNs) from spectroscopic data, the contamination from the light emitted by the stars in the surrounding host galaxy must be determined in order to be subtracted.

I characterise the ability of a spectral decomposition software code I have written based on the Levenberg-Marquardt algorithm to accurately determine the host galaxy contribution to optical spectra of AGNs. I do this by creating a catalogue of 43,200 synthetic AGN spectra with various amounts of continuum emission, iron emission and host galaxy emission, effectively simulating different types of AGNs. I add different amounts of noise to the spectra before they are modelled by the decomposition code, in order to study the effect of noise on the results.

I find that the host galaxy contribution is accurately determined in the spectra with a galaxy ratio of 0.75 (measured at 6000 Å) to the observed spectrum. It becomes increasingly difficult to accurately determine the host galaxy emission as this contribution becomes relatively weaker, resulting in a significant overestimation of the host galaxy emission in these cases. I conclude that the spectral decomposition method is most applicable to low-luminosity Seyfert galaxies, though its applicability will depend on the desired accuracy and precision of the user.

Increasing the signal-to-noise ratio (S/N) from 5 to 50 does not affect the accuracy of the determination of the host galaxy contribution. Due to the increasing uncertainties and degeneracies between the different spectral components for low S/N ratios, the spectral decomposition method used in this thesis should only be applied to spectra with S/N $\gtrsim$ 10 pixel^{-1}.

Acknowledgements

This thesis was written as a part of a Master's Degree in Astrophysics at the University of Oslo, Norway. The research was done at the University of Copenhagen, Denmark and was supported by the Nordlys Exchange Programme.

I would like to thank my supervisor Marianne Vestergaard who helped make it possible for me to write my thesis with her at Dark Cosmology Centre in Copenhagen, and for proposing such an interesting project for me to work on. She has always believed in me, shown great interest in the project, and been of great help during the work on this thesis.

Thank you to Beverly Wills for providing one of the Fe II templates used in this thesis.

Last, but not least, a big thank you to my fiancé Vidar who has been incredibly supportive of me living in Copenhagen for almost ten months working on this project, and for always cheering me up and cheering me on.

Contents

Appendices

1 Introduction

One of the big unanswered questions in astrophysics is how galaxies and the supermassive black hole (SMBH) found in their centre have formed and evolved into the structures that we see throughout the Universe. To learn about the formation and evolution of galaxies and their SMBH, they must be studied early in their formation history and at different cosmological times. Because observations point to some form a coevolution of galaxies and their SMBHs (Kormendy & Ho 2013), studying how SMBHs feed and grow can help shed light on this topic.

1.1 Measuring black hole masses

Measuring the mass of SMBHs at various cosmological distances (i.e., at various cosmological times) can be used to trace the evolution of SMBHs. The mass of SMBHs may be measured by studying the motions of the stars orbiting the black hole in the central region of a galaxy, as the stars' motions will be influenced by the gravitational field of the black hole. But this method can only be used in galaxies that are relatively close to us, as it becomes difficult to resolve individual stars and their motions in galaxies that are farther away (Ferrarese & Ford 2005).

Active galactic nuclei (AGNs) can be used to measure SMBH masses in more distant galaxies. An AGN is the compact, energetic central region of a galaxy in which the SMBH is actively accreting matter. Matter is heated up as it is pulled towards the SMBH, releasing huge amounts of energy. This process makes AGNs some of the most luminous objects in the Universe. Due to their high luminosities, AGNs can be seen at great cosmological distances. The most distant AGN found to date is the quasar ULAS J1120+0641 (Mortlock et al. 2011). This quasar is observed at redshift $z = 7.085$, which means it had already formed when the Universe was only 770 million years old. It is estimated to have a SMBH in its centre with a mass of about $2 \cdot 10^9\,M_\odot$. This is a surprisingly massive object to exist at such an early time in the history of the Universe. It is not yet understood how SMBHs form and are able to grow so quickly (Latif et al. 2013).

The mass of SMBHs in AGNs at high redshifts can be determined using spectroscopic measurements of the AGN emission. The luminosity of the AGN varies over time in response to the material falling in towards the black hole. This change in luminosity will cause changes in the broad emission lines commonly seen in AGN spectra, which are emitted from a region further out in the AGN. The technique of reverberation mapping (Blandford & McKee 1982; Peterson 1993) can be used to study how long it takes for the change in luminosity to cause a change in the emission lines, effectively measuring the distance between the central region and the line-emitting region. Using this method on a number of AGNs has lead to the discovery of the R–L relation (Kaspi et al. 2000; Bentz et al. 2006, 2009). This relation connects the radius R of innermost line-emitting region surrounding the SMBH to the optical luminosity L of the AGN. Now that this relation has been established, it has become a powerful tool because it allows for the estimation of SMBH masses from single-epoch spectra directly by taking two measurements: the optical luminosity of the AGN at 5100 Å and the linewidth of the broad emission line Hβ. I describe the method in more detail in Chapter 2.

But measuring the luminosity of the AGN is not unproblematic. AGN spectra will inevitably contain a contribution from the light emitted by the stars in the host galaxy surrounding the AGN. This is because the long, wide slit that is used to take the spectroscopic measurement will not just cover the central AGN, but will also include a part of the galaxy. The light will contaminate the luminosity measurements of the AGN made from the spectra and result in an overestimation of SMBH masses (Denney et al. 2009). The stellar light contribution must therefore be subtracted from the spectra before the AGN luminosity is measured so that the SMBH mass can be accurately determined.

The stellar light contribution from the host galaxy has previously been determined using high-resolution images from the *Hubble Space Telescope (HST)* (e.g., Bentz et al. 2009). Because *HST* will only be operational for few more years, a different method is needed.

An alternative method for determining the host galaxy contribution to AGN spectra is performing a spectral decomposition of the AGN spectra. In this method the spectra are modelled by adjusting several spectral templates to observed AGN spectra using a fitting algorithm. The spectral templates represents the different types of emission that are known to contribute to the AGN spectrum: continuum emission, iron emission, Balmer emission, emission lines and the host galaxy emission (Wills et al. 1985). But with what certainty is this method able to determine the host galaxy contribution?

1.2 This work

The accuracy and precision to which the host galaxy contribution can be estimated using a spectral decomposition method, has so far only been studied using observational data (e.g., Dietrich et al. 2002a; Matsuoka et al. 2015; Barth et al. 2015). The problem with this approach is that the true contribution from the host galaxy emission is not known. In my thesis I test how successfully the method of spectral decomposition can estimate the host galaxy contribution by testing the method on synthetic AGN spectra. Because I make the spectra myself, the various spectral contributions are known.

I create synthetic AGN spectra using a variety of templates that represent known spectral features of AGNs. I describe how I do this in Chapter 3. I then created a catalogue consisting of a total of 43,200 synthetic spectra where I allow the different spectral features to vary in strength, thus effectively simulating different types of AGNs. The spectra are also made with different amounts of noise so that I can study how the noise affects the ability to recover the host galaxy emission.

I write my own spectral decomposition software based on the Levenberg-Marquardt algorithm (Moré 1978), which I describe in Chapter 4. Once this software has decomposed the spectra in my catalogue, I can measure and characterise how successfully the host galaxy contribution can be determined using this method. I do is by comparing the output values found in the spectral fitting process with the input values of the synthetic spectra. By decomposing the entire catalogue of synthetic AGN spectra, I can do a statistical study of how the strength of the various spectral components will affect the result. I present the results of the decomposition in Chapter 5. This study allows me to determine if using spectral decomposition software is a suitable method for determining SMBH masses in the future when space-based imaging is no longer a possibility. I discuss the implications of the results and contemplate possible improvements and future studies in Chapter 6, before I conclude my study in Chapter 7.

2 Active galactic nuclei (AGNs)

An active galactic nucleus (AGN) is an energetic, luminous and compact region found in the centre of some galaxies, both elliptical and spiral. An object is said to be an AGN if it has one or more of the following properties (Mo, van den Bosch & White 2010):

1. A compact, point-like nuclear region that outshines a similarly-sized region in a normal galaxy.
2. Non-stellar continuum emission.
3. Strong emission lines.
4. Variable continuum emission or variable emission lines on short time scales of a few days or months.

Observed AGNs have different subsets of these properties, all of which are rarely seen in normal galaxies. One commonly shared characteristic is that AGNs emit radiation across most of the electromagnetic spectrum, from radio to gamma rays.

2.1 The standard AGN model

The standard AGN model, illustrated in Figure 2.1, assumes the existence of a supermassive black hole (SMBH) as the 'central engine' (Blandford & Reese 1992). The AGN is thought to be fuelled by accretion of gas falling onto the SMBH. But this material does not fall directly onto the SMBH. The angular momentum of this material will instead cause the material to form an accretion disk surrounding the SMBH. When this material is drawn inwards towards the SMBH in a spiralling motion, it is heated up by viscous processes and radiates away its gravitational potential energy, which is observed as continuum emission (Sparke & Gallagher 2007). The continuum emission photoionises the gas regions surrounding the accretion disk, which produces the emission lines seen in AGN spectra (Peterson 1997). The radiation from the AGN can be obscured by a torus of very dense gas and dust surrounding the accretion disk, depending on our line of sight. Outside the torus lies the host galaxy. Some AGNs have been seen to have relativistic jets ejecting plasma from the central region into the intergalactic medium, though the underlying mechanisms are not well understood (Romero et al. 2017).

2.2 Spectral features

AGNs are too small to be spatially resolved, therefore our main source of information about these objects are their spectra. Here I go through the main spectral features associated with AGNs.

The **continuum emission** is emitted from the accretion disk, and determines the overall shape of the AGN spectrum. The shape of the continuum emission – whether it is steep or flat in the UV – depends on the mass accretion rate of the SMBH. A spectrum that is steeper in the UV is associated with a higher mass accretion rate (Winter 2008).

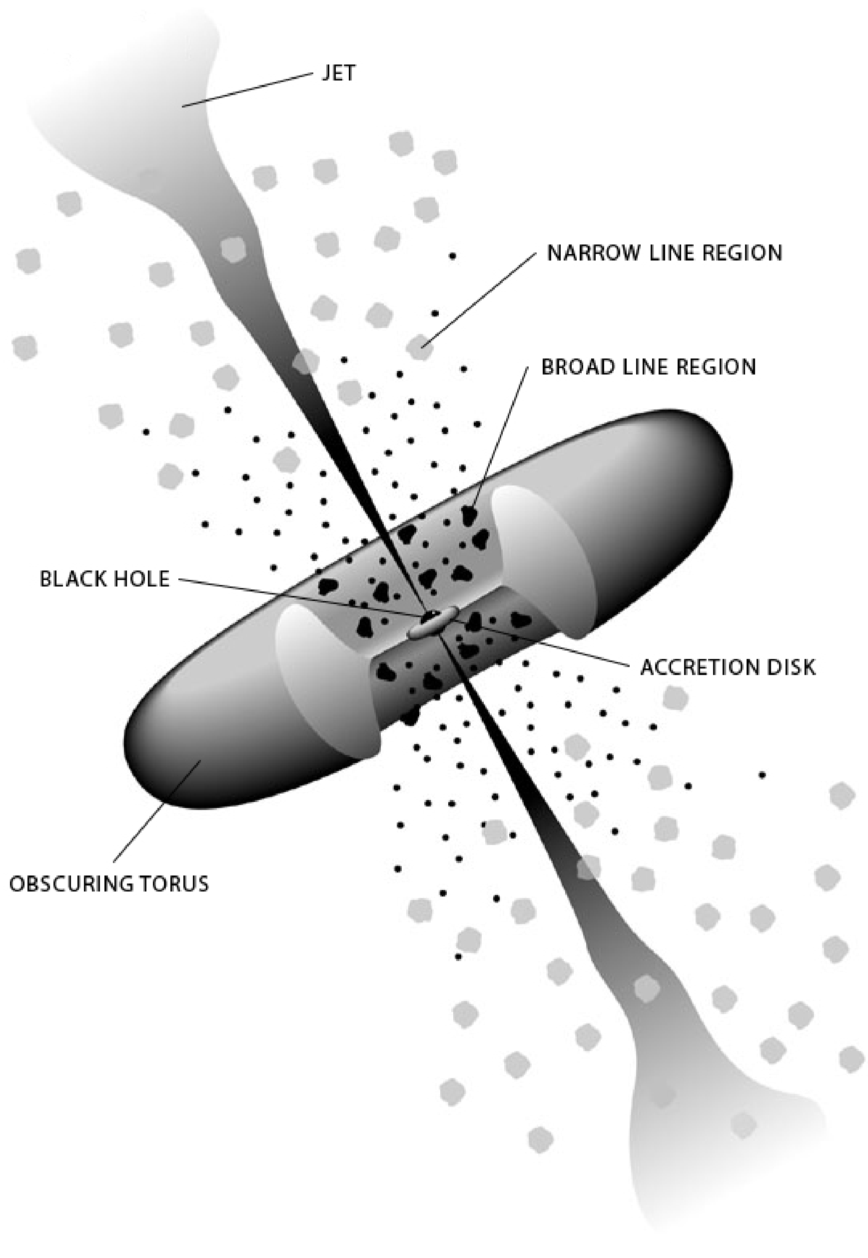


Figure 2.1: The anatomy of an AGN. At the core is the supermassive black hole with a surrounding accretion disk. Further out are gas regions: the broad line region and the narrow line region. Surrounding the accretion disk and the broad line region is an obscuring torus of dense gas. The host galaxy is located outside the torus. The AGN can have relativistic jets that ejects plasma from the central region into the intergalactic medium. Illustration from Urry & Padovani (1995), annotation by me.

Emission lines are often prominent features in AGN spectra. The width of the emission lines reflect the velocities of the line emitting gas regions. If observed emission lines are found to have very different linewidths, this is an indication that they are created in different gas regions. The AGN emission lines are divided into two categories based on their full linewidth at half of the maximum intensity (FWHM): Broad emission lines and narrow emission lines. Figure 2.2 illustrates how the broad and narrow emission lines can appear in AGN spectra, though many of the most prominent lines are found blueward of the spectral ranges shown here. More specifically (based on Peterson 1997 and Krolik 1999):

- *Broad emission lines* are produced in the innermost emission-line region of the AGN where the gas densities and velocities are high, known as the broad line region (BLR). Broad emission lines have linewidths in the range of $5{,}000 \lesssim \Delta v_{\mathrm{FWHM}} \lesssim 10{,}000\,\mathrm{km\,s^{-1}}$ and are typically the most dominant features in AGN spectra. The close proximity of the BLR to the central SMBH makes the BLR important in the study of AGNs, as the SMBH itself cannot be seen directly. The motions of the BLR gas is believed to be determined by the SMBH, and the intensity of the broad emission lines can vary strongly with the variations of the continuum luminosity. As I will discuss in Section 2.4.2, the broad emission lines may therefore be used to estimate the SMBH mass.

- *The narrow emission lines* are emitted from extended regions with lower gas densities and velocities, known as the narrow line region (NLR). Narrow emission lines have linewidths in the range $200 \lesssim \Delta v_{\mathrm{FWHM}} \lesssim 900\ \mathrm{km\,s^{-1}}$ and are most prominent in low-luminosity AGNs. This is because the strength of the narrow lines appear to decrease with increasing continuum luminosity, unlike broad emission lines, making them very difficult to detect in high-luminosity AGNs. The lack of narrow lines in higher-luminosity AGNs can also be the result of the lines being shifted outside the observing window at high redshifts, which is where high-luminosity AGNs often are observed.

Collisional broadening of spectral lines mainly takes place in high density gases, therefore narrow lines are typically of the forbidden type, whilst broad lines are of the permitted type. Forbidden lines are not strictly forbidden, simply much less probable than permitted lines. The emission lines that typically are present in AGN spectra, include (Krolik 1999):

- *Broad emission lines:* Lyα, Hα and Hβ; He II 1640, He II 4686, He I 5876 and He I 10830; C III] 1909 and C IV 1548, 1551; N V 1239, 1243; O VI 1032, 1038, O IV] 1400, O I 1305 and O I 8446; Si IV 1394, 1403; Mg II 2796, 2804; and several clusters of Fe II lines. The one-sided brackets indicate that these lines are of the semi-forbidden type.

- *Narrow emission lines:* [O III] 4959, 5007, [O II] 3727 and [O I] 6300; [N II] 6548, 6583; [S II] 6716, 6731. The two-sided brackets indicate that these lines are of the forbidden type.

The width and intensity of emission lines, as well as the strength of the continuum emission depends on the type of AGN observed and the particular object observed. I discuss the different types of AGNs in Section 2.3.

AGN spectra may exhibit **absorption lines** in addition to the emission lines. Given that AGNs often are observed at great distances, there will inevitably be some gas along our line of sight creating absorption lines in the AGN spectra that are not a characteristic of the AGN itself. But there can also be absorption features associated with the AGN itself.

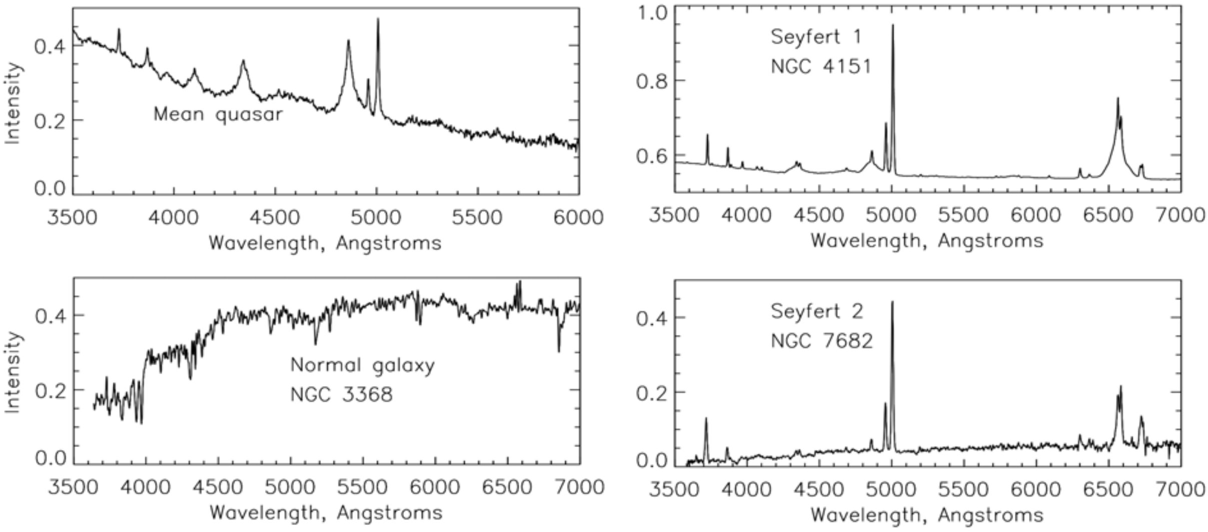


Figure 2.2: Examples of spectra for different AGN types for comparison, as well as the spectrum of a normal galaxy. The quasar spectrum is steep towards the UV, indicating a high SMBH mass accretion rate. The Seyfert spectra are close to flat, indicating a much smaller SMBH mass accretion rate. The pronounced broad emission line around 6500 Å is Hα. The two narrow emission lines around 5000 Å is the [O III] doublet. The quasar spectrum also has a strong Hβ line, directly left of the [O III] doublet. The galaxy spectrum shows clear absorption features. The spectra were assembled by W. Keel (2012; reproduced with permission).

These absorption features tend to be weak and unresolved, but there are also examples of AGNs with strong broad absorption lines (BALs). BALs are always blue-shifted, indicating that the gas where the lines are emitted from must be moving away from the AGN. This suggests that the BAL features might be a result of outflows or winds. BAL AGNs are relatively rare (Peterson 1997).

There will also be a contribution from the **host galaxy emission** in AGN spectra. This is because the long, wide slit that is used to take the spectroscopic measurement will not just cover the central AGN, but will also include a part of the galaxy. This contribution will be most pronounced in nearby low-luminosity AGNs. The host galaxy contribution will be difficult to observe in distant AGNs because the surface brightness of the galaxy dims rapidly with redshift as $(1+z)^{-4}$ (Sparke & Gallagher 2007). AGNs are not affected by this because they are point sources. The host galaxy contribution will also be difficult to detect in high-luminosity AGNs, as the host galaxy will tend to be outshined by the high-luminosity AGNs.

2.3 Types of AGNs

There are several types of AGNs, which are distinguished based on their spectral properties. For my study, which will be focused on the optical emission, the most important AGN types are Seyfert galaxies and quasars. Figure 2.2 shows examples of spectra for these different types of AGNs.

Seyfert galaxies are lower-luminosity AGNs with a clearly detectable host galaxy surrounding the AGN. Seyfert galaxies are classified as type 1 and 2. The spectra of Seyfert 1 galaxies have both broad lines and forbidden narrow lines in their spectra, while the spectra of Seyfert 2 galaxies only have narrow lines, both of the permitted and forbidden kind. Absorption features from the stars in the host galaxy may also be seen in the spectra of Seyfert galaxies. Most, if not all, Seyferts occur in spiral galaxies. Seyfert galaxies are the most common type of AGN, but are rare compared to normal, inactive galaxies (Sparke & Gallagher 2007).

Quasars are the most luminous type of AGNs. They look like point sources (i.e., spatially unresolved) with no clearly detectable host galaxy surrounding them as they are often observed a high redshifts. Both broad and narrow emission lines appear in quasar spectra.

2.4 The relationship between galaxies and SMBHs

Much attention has been given to SMBHs in the early Universe after the discovery of quasars existing at redshift $z > 6$ (Fan et al. 2001). These quasars emitted this light at a time when the Universe was less than a billion years old and harbour SMBHs that have been able to grow very massive at a very early time in the Universe. It is not yet understood how SMBHs form or how they are able to grow so rapidly (Latif et al. 2013). By studying AGNs and their SMBHs at different cosmological times, we hope to learn about the formation, evolution and relationship of the SMBHs and galaxies that host them.

Two relations has been discovered that are of particular importance, which I briefly summarise here.

2.4.1 The mass-velocity dispersion relation

The mass of the SMBH has been found to be tightly correlated with the velocity dispersion of the stars orbiting in the galactic bulge, known as the $M_\bullet$–σ_* relation (Ferrarese & Merritt 2000; Gebhardt et al. 2000). Here $M_\bullet$ is the mass of the central SMBH and σ_* is the velocity dispersion of the stars surrounding the SMBH.

Even if the SMBH is extremely massive, its gravitational reach – the so-called 'sphere of influence' – is very limited compared to the full extent of the host galaxy, and σ_* is measured well outside this region. It was therefore surprising to discover the $M_\bullet$–σ_* relation. The tightness of this relation indicates that there is a close relationship between the formation and growth of the SMBH and the host galaxy, pointing at some form of co-evolution (Kormendy & Ho 2013). However, the origin of the $M_\bullet$–σ_* relation is not yet determined. The tightness of the relationship has been taken as an indicator that most, if not all galaxies are harbouring a SMBH in their centres.

The $M_\bullet$–σ_* relation was originally determined by looking at elliptical galaxies and the bulges of spiral galaxies where there was believed to be SMBHs. In recent years, it has been shown that AGNs are consistent with the $M_\bullet$–σ_* relation as well (Xiao et al. 2011).

In order to better understand the cosmic evolution of SMBHs and possibly the connection to their host galaxy, the determination of SMBH masses at different redshifts is needed. The power of the $M_\bullet$–σ_* relation is that it can be used to measure the mass of SMBHs by measuring the velocity dispersion σ_* – if the relation becomes tight enough. At the moment there is a significant amount of scatter, depending on the objects included when the slope is calculated, illustrated in Figure 2.3. Also, σ_* can only be measured in nearby galaxies where the stellar and gas dynamics can be spatially resolved in observations. In galaxies that are further away – at redshifts above $z \sim 0.03$ ($d \sim 130$ Mpc) (Ferrarese & Ford 2005) – other methods are needed.

2.4.2 The radius-luminosity relation

Quiescent (non-active) galaxies become difficult to observe at large distances because their surface brightness dims rapidly with redshift, which will not affect the AGN, as mentioned previously. AGNs can however be extremely bright, ranging from having a luminosity that is 1 % of a typical galaxy to $\sim 10{,}000$ times greater (Krolik 1999). AGNs therefore offer an opportunity to study SMBHs at large cosmological distances.

The study of AGNs has lead to the discovery of the radius-luminosity relation. This

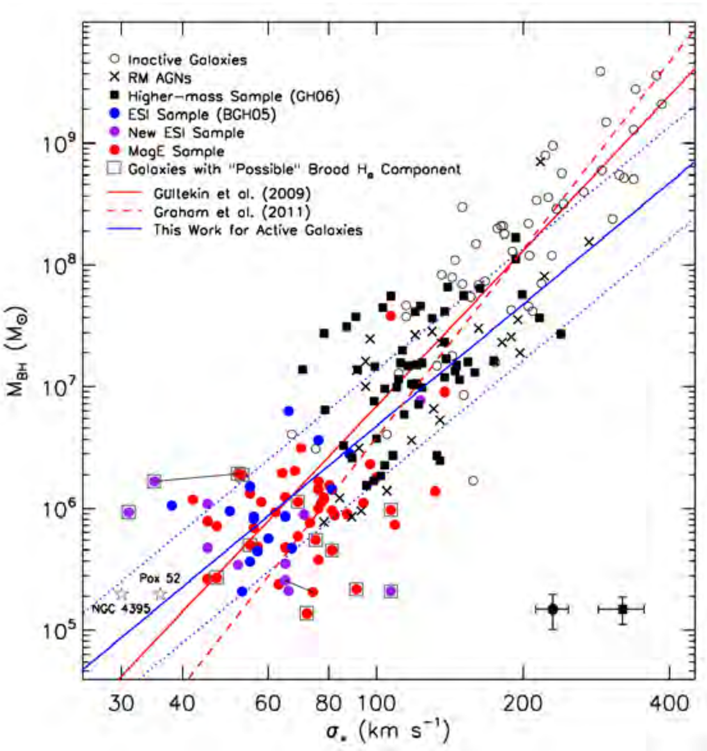


Figure 2.3: The $M_\bullet$–σ_* relation for massive black holes (Xiao et al. 2011). The lines represent the derived $M_\bullet$–σ_* relation based on various subsets of data. The blue line, representing the best fit for AGNs, has $\alpha = 7.68 \pm 0.08$ and $\beta = 3.32 \pm 0.22$ for the line log $M_\bullet = \alpha + \beta$ log $(\sigma_*/200$ km s$^{-1})$.

relation connects the radius of the BLR region to the AGN continuum luminosity as $R_{\rm BLR} \propto L^\alpha$ (Kaspi et al. 2000). The current best estimate for α is $\alpha = 0.519^{+0.063}_{-0.066}$ (Bentz et al. 2009), which is consistent with photoionisation arguments. The relation is illustrated in Figure 2.4. The $R_{\rm BLR}$–L relationship has been established using the technique of *reverberation mapping* (Blandford & McKee 1982; Peterson 1993). The idea behind this method is as follows: The emission from the AGN is not constant in time – it will change in response to the material from the accretion disk falling onto the SMBH. Because the BLR is photoionised by the emission from the central accretion disk, changes in the strength of the AGN continuum will be followed by changes in the strength of the broad emission lines. The change in the strength of the broad emission lines will have a time lag due to the light-travel time from the centre of the continuum emitting regions in the disk out to the gas region of the BLR where the broad emission lines are emitted. The BLR can be said to 'reverberate' in response to the changes in the continuum flux. The time-lag can be used as an estimate of the radius of the BLR. In reality, the time-lag will measure the distance out to the gas that responds most strongly to the continuum variations.

The width of the broad emission lines $\Delta v_{\rm FWHM}$ can then be used to estimate the virial mass enclosed by the BLR, which will be dominated by the mass of the SMBH, through

$$M \approx \frac{R_{\rm BLR} \Delta v^2_{\rm FWHM}}{2G}. \tag{2.1}$$

However, using reverberation mapping to determine SMBH masses is a time-consuming method as it requires observations over time scales of days, weeks, months or years, depending on the rate at which the observed AGN varies in luminosity. This can be done in theory, but it is not an efficient way to determining the SMBH mass. Instead, the already established $R_{\rm BLR}$–L relation can be used directly.

With the $R_{\rm BLR}$–L relation established, this relation may be used to estimate the virial

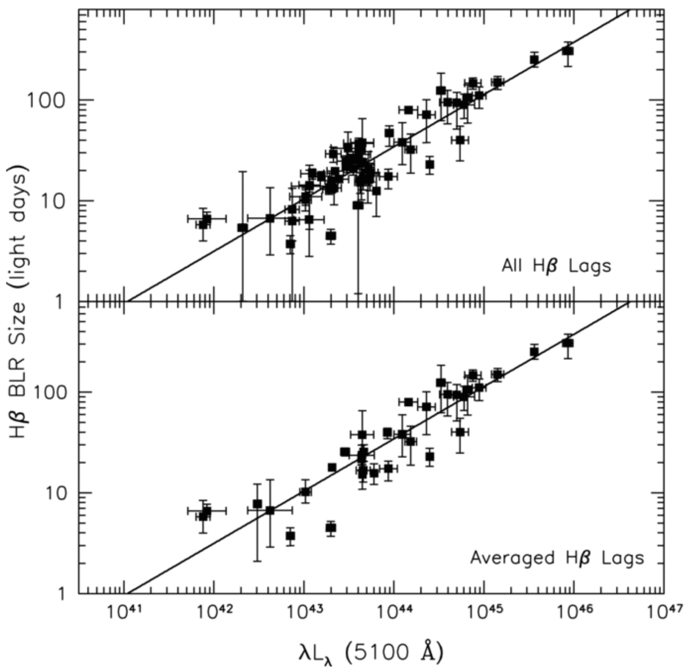


Figure 2.4: The Hβ $R_{\rm BLR}$–L relation measured from 35 AGNs after correcting the AGN luminosities for the contribution from host-galaxy emission (Bentz et al. 2009). The top panel shows each separate measurement as a data point, while the bottom panel shows the weighted mean of multiple measurements for each individual object. The solid lines are the best fit to the data and has a slope of $\alpha = 0.519$.

mass by simply making two measurements from a single optical spectrum of an AGN: 1) the width of the broad Hβ emission line, and 2) λL_λ at λ = 5100 Å. The Hβ line width is a measure of the BLR velocity dispersion, while the λL_{5100} is a measure of the BLR radius. The SE virial mass $M_{\rm SE}$ is then given by (Bentz et al. 2009)

$$\log\left(\frac{M_{\rm SE}}{M_\odot}\right) = -22.0 + 0.519\log\left(\frac{\lambda L_{5100}}{\rm erg\,s^{-1}}\right) + 2\log\left(\frac{V_{\rm H\beta}}{\rm km\,s^{-1}}\right), \tag{2.2}$$

where λL_{5100} is the luminosity at rest frame wavelength 5100 Å, and $V_{\rm H\beta}$ is the line width of the broad Hβ emission line.

Equation 2.2 is currently being used to estimate SMBH masses in distant AGNs. The simplicity of the measurements necessary to use the this equation makes it very powerful, as it allows for quick estimates of SMBH masses in large samples of AGNs at various redshifts – all from a single spectrum of each object.

In order to use the $R_{\rm BLR}$–L relation to estimate the SMBH mass through Equation 2.2, the luminosity L_{5100} must be measured from the AGN continuum flux, and *only* the continuum flux. The reality is that the starlight from the host galaxy surrounding the AGN will contaminate the luminosity measurements, making the AGN appear more luminous than it actually is. If the contribution from the starlight can be successfully determined and removed, correct measurement of the luminosity of the BLR can be obtained, and the $R_{\rm BLR}$–L relation will provide a simple and effective tool for estimating SMBH masses. If the host galaxy contribution is not taken into account and removed, this will lead to an overestimation of the BLR radius, which in turn will result in an overestimation of the SMBH mass (Denney et al. 2009). This effect will be more prominent for low-luminosity AGNs such as Seyfert galaxies where the host galaxy is relatively strong compared to the AGN, and less so for quasars, where the AGN is very strong compared to the host galaxy.

2.5 Current methods for determining the host galaxy contribution

In order to measure SMBH masses using the R_{BLR}–L relation, the host galaxy contribution to the AGN luminosity must be determined. Here I review some of the current methods.

2.5.1 Image decomposition

This method for determining the host galaxy contribution to AGN spectra requires high-resolution images of the AGN in question and its host galaxy. The surface brightness profile of the host galaxy is modelled using a 2D-image decomposition software code such as `GALFIT` (Peng et al. 2002). This type of program models the surface brightness profile of the galaxy by fitting analytic functions to the bulge and disk components of a galaxy. A central point source representing the AGN is described by an additional component, a point-spread function (PSF). The PSF is characteristic of the telescope and the detector being used to take the image. Once the central point source has been modelled, it can be subtracted from the galaxy, creating an AGN-free image. This image can be used to determine the host galaxy contribution by measuring the observed flux at a rest-frame wavelength, typically 5100 Å. The final step is to subtract this host galaxy flux from the original flux measured in the observations, which leaves only the flux contribution from the AGN itself.

A typical issue with this type of method is the degeneracy between galaxy components and between the parameters within the components. In order to try to remedy such issues, Bentz et al. (2009) suggests a number of consistency checks (see paper for details). The method is also highly limited by angular resolution. The method is not applicable on AGNs at high redshifts, as high-resolution images are needed in order to separate the PSF from the host galaxy light.

Bentz et al. (2006) attempted to use the image-decomposition method with ground-based telescopes and found that the images taken were more dependent on the initial parameters given to GALFIT, making the need for space-based telescopes. Specifically, the *Hubble Space Telescope (HST)* is needed because it is an optical telescope, allowing for a determination of the host galaxy contribution at 5100 Å. However, as *HST* is approaching the end of its lifetime, and there are no plans for a replacement operating in the optical, other methods must be considered.

2.5.2 Spectral decomposition

Using AGN spectra is an alternative to using high-resolution images from *HST* to determine the host galaxy contribution to AGNs. Spectral observations are more easily obtained than high-resolution images, and there are already large catalogues of AGN spectra available, such as the *Sloane Digital Sky Survey (SDSS)*.

The idea that is being explored, is to determine SMBH masses from single-epoch spectra by using Equation 2.2 (Denney et al. 2009). For single-epoch spectra, the host galaxy contribution to the spectra may be determined using spectral decomposition methods. There are different ways of decomposing an AGN spectrum, but the general idea behind such methods is the same: One assumes that the AGN is the linear combination of a set of known spectral components, and fit these components to an observed spectrum using for example a least-squares fitting algorithm. Once a satisfying fit has been found, one will have information about how the various components contribute to the spectrum, where one of the components will be the host galaxy contribution.

The main difference between various spectral decomposition methods is the choice of spectral components that are fitted to the observed spectrum. There may also be a difference in wavelength coverage and the fitting algorithm used. A commonly used set of spectral components consists of (Wills et al. 1985): AGN continuum emission, Fe II emission, Balmer emission, emission lines and host galaxy emission. I use this approach when I create my catalogue of synthetic AGN spectra, and I describe the components in more detail in Chapter 3.

Spectral decomposition methods are already being used on real observational data (e.g., Dietrich et al. 2002a; Matsuoka et al. 2015; Barth et al. 2015). With real observational data we do not know the correct contribution from the different spectral components. There is the possibility of degeneracy between the various spectral components, and the result from the spectral modelling may depend on the model assumed in the fitting process. It is therefore important to study how well spectral decomposition methods actually works, and that is where I come in. In my thesis, I will test one particular spectral decomposition method on a catalogue of synthetic AGN spectra that I make myself, in order to better quantify to what accuracy and precision this method actually works.

3 Creating synthetic AGN spectra

In order to test how well AGN spectra can be decomposed using a spectral decomposition software code, I create a catalogue of synthetic AGN spectra. By using synthetic data rather than real observations, I know exactly what the spectral decomposition program is supposed to find. In this chapter I describe how I create the synthetic AGN spectra, while I describe the decomposition program in Chapter 4.

3.1 The ingredients

An AGN spectrum can be approximated as a combination of the following five components (Wills et al. 1985):

1. AGN continuum emission
2. Fe II emission
3. Balmer continuum emission
4. Additional broad and narrow emission lines
5. Host galaxy emission

I go through these components in more detail below, describing how I created the final synthetic AGN spectra used in the modelling.

3.1.1 Continuum emission

The continuum emission in AGN spectra is associated with the accretion disk surrounding the SMBH and determines the overall shape of the AGN spectrum. The continuum can to a lower-order approximation be described by a power-law function of the form

$$F_\lambda = C\lambda^{\alpha_\nu - 2}. \tag{3.1}$$

Here F_λ is the flux at a specific wavelength λ, the spectral index α_ν determines the slope of the continuum[1], and C is a normalisation constant. This is the formula I use when creating the AGN continuum emission for my synthetic spectra.

The value of the spectral index depends on the mass accretion rate of the central SMBH in the AGN. A higher mass accretion rate leads to a steeper spectrum in the UV ($\alpha_\nu < 0$) and a lower mass accretion rate leads to a flatter spectrum in the UV ($\alpha_\nu > 0$). The spectral index is typically found to fall in the range $0 \lesssim \alpha_\nu \lesssim 1$ when fitting the power-law function to quasar spectra over a large range of frequencies (Peterson 1997).

When creating my catalogue of synthetic spectra, I want to create spectra with various amounts of SMBH mass accretion. I therefore choose to make spectra with three different values for α_ν, namely $\alpha_\nu = -0.5$ (steep in the UV), 0.5 (intermediate) and 1.0 (flatter in the UV). The difference in the appearance of these three values are shown in Figure 3.1.

[1] I use the spectral index in terms of frequency ν rather than wavelength λ ($\alpha_\lambda = \alpha_\nu - 2$) because the spectral index is usually quoted in terms of α_ν in the literature.

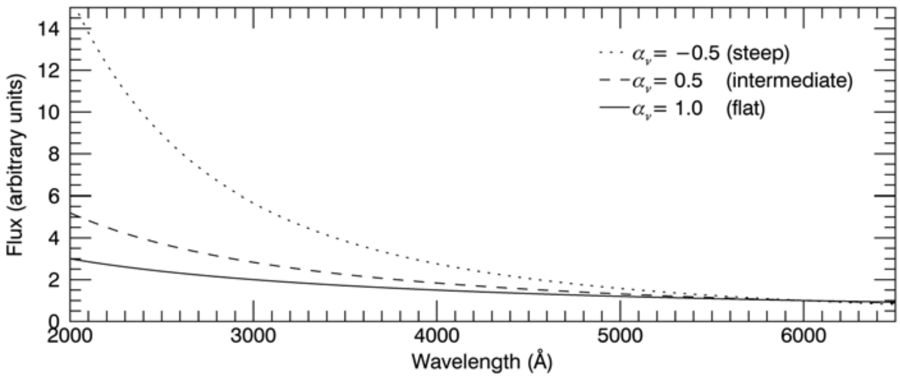


Figure 3.1: The power-law continuum for three different power-law indices α_ν. Lower values of α_ν results in continuum emission that is steeper in UV.

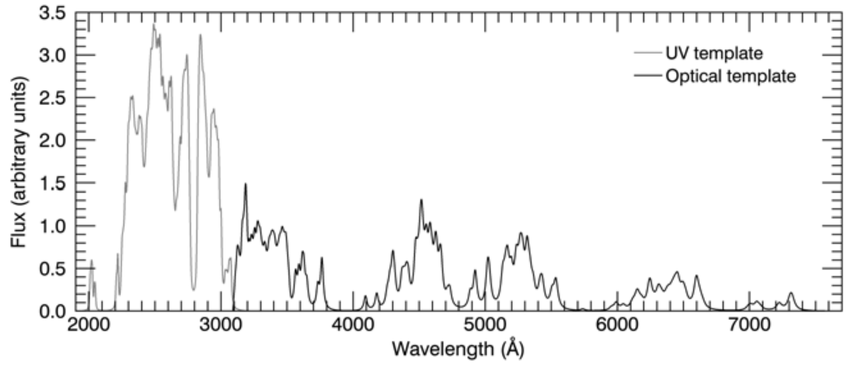
3.1.2 Fe II emission

The Fe II template I use in my synthetic spectra consist of an optical template and a UV template, as illustrated in Figure 3.2a. The optical template covers a wavelength range of 3090–7530 Å (I call it 'optical' even though it includes some UV wavelengths as well), while the UV template covers 1925–3090 Å. The optical template for 3535–7530 Å was created empirically by Véron-Cetty et al. (2004) based on the spectrum of the narrow line Seyfert 1 galaxy known as I Zwicky 1. This galaxy was chosen because of its very strong, rich and narrow iron emission, which makes it possible to separate Fe II emission from the other emission in the spectrum. The optical template for 3090–3535 Å is from B. Wills (2000, private communication). The UV template was provided by my supervisor.

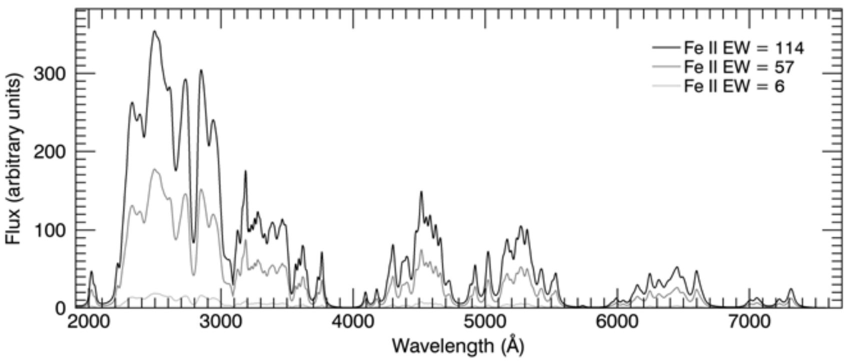
I velocity-broaden the Fe II optical template to 2000 km s^{-1} by convolving it with a Lorentzian. When creating my catalogue of AGN spectra I also make spectra where the optical template is broadened to 4000 km s^{-1} to test how the smoothness of the iron emission will impact the ability to recover the host galaxy emission. The linewidth of the UV emission relative to the optical emission is not known. However, the UV template is often observed to be relatively smooth, suggesting that the linewidth of the UV template might be broader than that of the optical template. I therefore choose to make the UV template broader than the optical template. Specifically, I broaden the Fe II UV template to be twice the width of the optical Fe II template. Choosing a factor of two is arbitrary. Exactly how the UV template looks is not likely to make a significant impact on the ability of the spectral decomposition software to recover the host galaxy emission in the modelling process, as the host galaxy contribution is very weak in the UV compared to the optical emission anyway.

Because the iron emission contributes in some of the same wavelength ranges as the host galaxy emission, it is important to test how different amounts of iron in the spectrum may affect the ability of the spectral decomposition program to recover the amount of galaxy in the spectrum. I have chosen to look at three different values: a small value, and intermediate value and a high value. The values I have chosen are based on the measurements of Fe II equivalent widths (EW) made by Borosen & Green (1992; hereafter referenced as BG92). They measured the Fe II EW from the line multiplet between λ4434 Å and λ4684 Å from the spectra of 87 different quasars. They found values ranging from zero to 114 Å, with 57 Å as the intermediate value. When creating my catalogue of synthetic spectra, I make spectra with no iron emission, intermediate iron emission, and maximum iron emission.

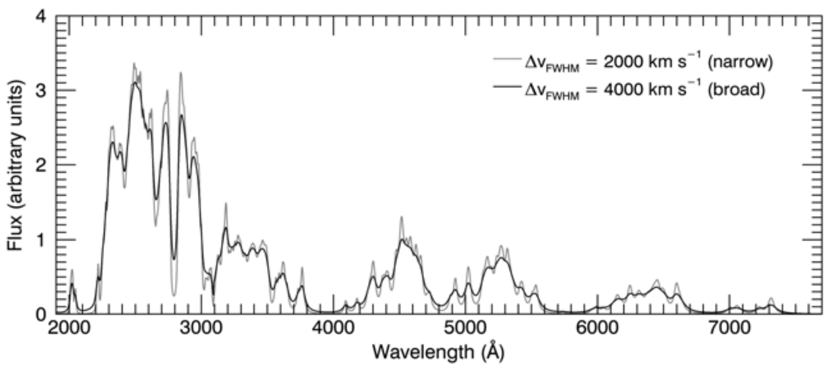
After I have velocity-broadened the iron templates, I measure the Fe II EW of the iron template over the same wavelength range as BG92, namely λ4434 Å and λ4684 Å. Then I



(a) Fe II templates



(b) Fe II EW



(c) Fe II linewidths

Figure 3.2: The Fe II emission. *Top:* The Fe II templates I use in my synthetic AGN spectra, consisting of an optical component and a UV component, velocity broadened to 2000 km s^{-1} and 4000 km s^{-1}, respectively. *Middle:* A visualisation of how the Fe II emission changes in strength with the Fe II EW's used in my catalogue of synthetic AGN spectra. *Bottom:* A visualisation of the difference in spectral features of a narrow and broad iron template. The broadening of the template, washes out the narrow features.

can scale the iron templates to the value I want for the Fe II EW before adding it to the synthetic spectrum.

3.1.3 Balmer lines & continuum emission

Emission lines produced by de-excitation of bound electrons to the $n = 2$ energy level of hydrogen are collectively known as *Balmer lines.* These lines are found in the wavelength range of $\lambda\lambda 3646$–6563 Å, ranging from UV to optical. The lower wavelength limit of $\lambda 3646$ Å is known as the *Balmer limit.* This marks the limit between the low energies that give rise to emission lines and the energies that are high enough for the hydrogen atom to become ionised, which gives rise to continuum emission. The *Balmer continuum emission* is the result of the recombination of *free* electrons being captured into the $n = 2$ energy state.

The IDL program and template I used to model the Balmer continuum and the related emission lines was provided by my supervisor. The template was created by Dietrich et al. (2002b) based on the study of 744 AGNs, building on the work of Grandi (1982). In addition to the Balmer continuum, the Balmer template also includes a large number of hydrogen lines and a range of helium lines (He I and He II), creating a template collectively referred to as *nebular emission.* This is because spectra of nebulae are dominated by hydrogen and helium lines. I velocity-broaden the template to 3500 km s^{-1} at FWHM, a typical value for the linewidth of Hβ (BG92), by convolving it with a Lorentzian. The strength of the hydrogen lines are given by atomic theory.

The Paschen continuum emission is associated with transitions to energy level $n = 3$ found shortward of $\lambda 8204$ Å (infrared). This emission should also have been included in the synthetic AGN spectra to better mimic real observations (Korista & Goad 2001). However, the strength of the Paschen continuum emission is not well constrained as it tends to be diffuse and difficult to observe compared to the Balmer continuum as a result of atmospheric extinction. It has therefore not been included in my synthetic spectra.

I use the equivalent width of Hβ to scale the Balmer template to the underlying power-law continuum. BG92 measured values of Hβ EW to be in the range 23–230 Å for the 87 quasars they studied. I use the median value in their set of measurements of Hβ EW $= 93$ Å as a reference point in my synthetic spectra.

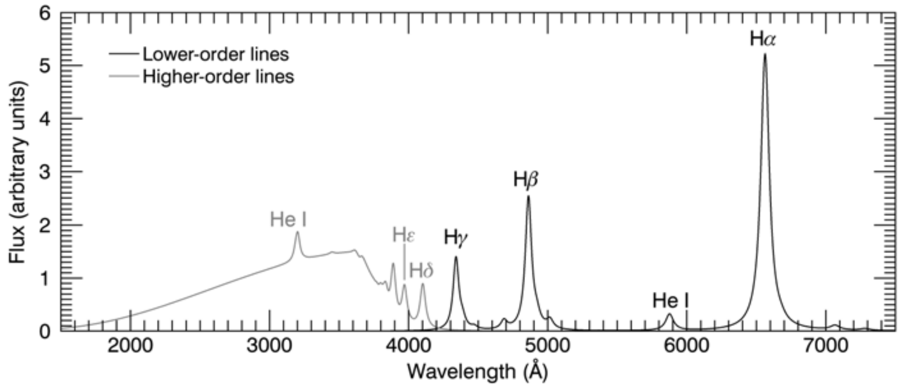


Figure 3.3: The Balmer continuum template (the big bump in the 2000–4000 Å range), as well as a selection of hydrogen and helium emission lines after being velocity-broadened to 3500 km s^{-1} (Grandi 1982; Dietrich et al. 2002b). I have separated the template into lower and higher order Balmer lines, where the continuum is part of the template containing higher order lines. Prominent lines are marked by name.

Table 3.1: The broad emission lines added to the synthetic spectra. Based on Francis et al. (1991).

Identification	Line centre (Å)	Start (Å)	End (Å)	Flux relative to Hβ
Si IV	1396	1353	1454	0.86
C IV	1549	1452	1602	2.86
C III]	1909	1828	1976	1.32
Mg II	2798	2650	2916	1.55
Hβ	4861	4704	5112	1.0

3.1.4 Additional emission lines

In addition to the emission lines added to the spectrum by the Balmer and Fe II emission templates mentioned above, there are other prominent emission lines that must be added that are characteristic of AGN spectra.

The broad emission lines I have included in my synthetic spectra are listed in Table 3.1 and shown in Figure 3.4. I model these lines using one Lorentzian profile for each line, characterised by the line centre and line width. I broaden the UV lines to $\Delta v_{\rm FWHM} = 5000$ km s^{-1} and broaden the optical line Mg II $\lambda2800$ to $\Delta v_{\rm FWHM} = 4500$ km s^{-1}. These are average linewidths for broad emission lines (Peterson 1997). The UV lines tend to be broader than the optical lines, which is why I set them to be a bit broader than the optical emission line. This small difference in linewidth for these two groups of emission lines is not likely to be significant for the results of the spectral decomposition.

It is a simplification to create the emission lines using one Lorentzian for each line. Observed spectral lines often have both a broad and narrow component, and are not necessarily symmetric around the line centre. Thus it is common to use multiple Gaussian profiles with different line widths and wavelength shifts in order to reproduce the sometimes complex profiles of the broad emission lines. Because the main goal of my thesis is to estimate the host galaxy contribution to AGN spectra rather than study the emission lines themselves in detail, my focus is getting the continuum level right. This is because the continuum level can affect the ability to recover the host galaxy contribution. The continuum level is mainly influenced by the lines wings of the emission lines. Because of this, I can get away with using one Lorentzian for each line, which has broader line wings than a Gaussian profile, and which can therefore account for the shape of the wings that I would otherwise need multiple Gaussian profiles to produce.

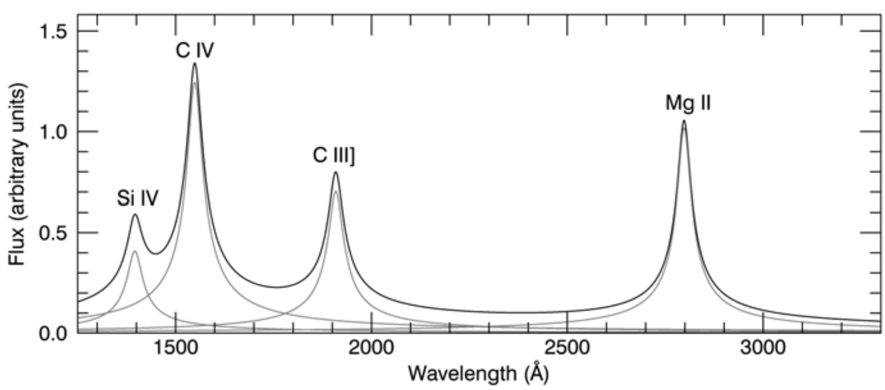


Figure 3.4: The four broad emission lines added to the mock spectrum, which are modelled with Lorentzian profiles. The UV lines are broadened to $\Delta v_{\rm FWHM} = 5000$ km s^{-1}, while the optical line Mg II $\lambda2800$ is broadened to $\Delta v_{\rm FWHM} = 4500$ km s^{-1}. The line fluxes are calculated based on the values in Table 3.1.

The narrow emission lines [O III] λ5007 Å and [O III] λ4959 Å are prominent features in many AGN spectra, located right next to Hβ, as can be seen in Figure 2.2. I did however decide not to include any narrow emission lines in my spectra. These lines are typically very narrow compared to the broad emission lines and are usually modelled with Gaussian profiles, which do not contribute to the continuum in a noticeable amount due to the weak line wings. They may therefore be excluded from my investigation in order to save computing time.

3.1.5 Host galaxy emission

A galaxy spectrum is simply the sum of the spectra of all the stars and H II regions inhabiting the galaxy. Because different galaxy types are composed of different types of stars, their spectra also differ. Elliptical galaxies (E) tends to be dominated by old, red stars. Spiral galaxies (Sa, b, c) consists of a mixture of stars, with young, hot bluer stars dominating in the spiral arms and older stars in the central bulge. The bulge can therefore resemble an elliptical galaxy. Sa galaxies has a bright bulge and tightly wrapped spiral arms, while the bulge is less bright and the arms more loosely wrapped around Sb and Sc galaxies. Lenticular galaxies (S0) are similar to spiral galaxies in that they have a bright central bulge surrounded by a disk, but their disk do not have a visible spiral structure or as active star formation as seen in spiral galaxies.

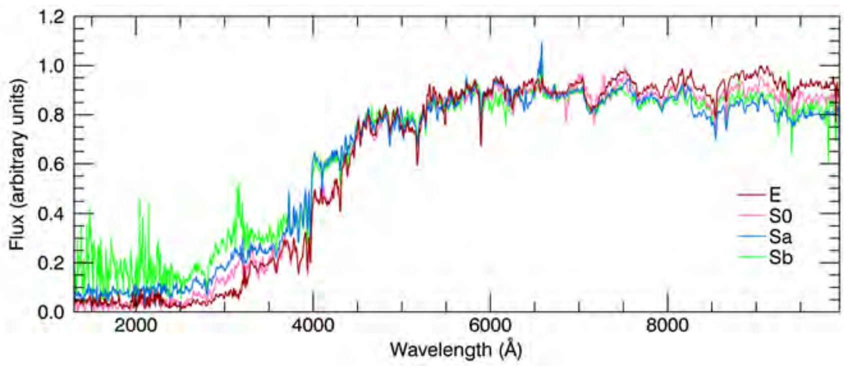
I add the emission coming from the host galaxy surrounding the AGN to my synthetic spectra by using galaxy templates from Kinney et al. (1996). The templates were created using UV and optical spectra taken from a number of quiescent galaxies within each morphological group, resulting in one galaxy template for each morphology type, shown in Figure 3.5a.

The galaxy emission is strongest for $\lambda > 4000$ Å. An important feature of galaxy spectra is the increase in flux around 4000 Å, known as the Ca II break. This break will be most prominent in elliptical galaxies as they are dominated by older, red stars emitting light in the redder wavelengths, and quite weak in spiral galaxies which consists of mixed young-old stellar populations.

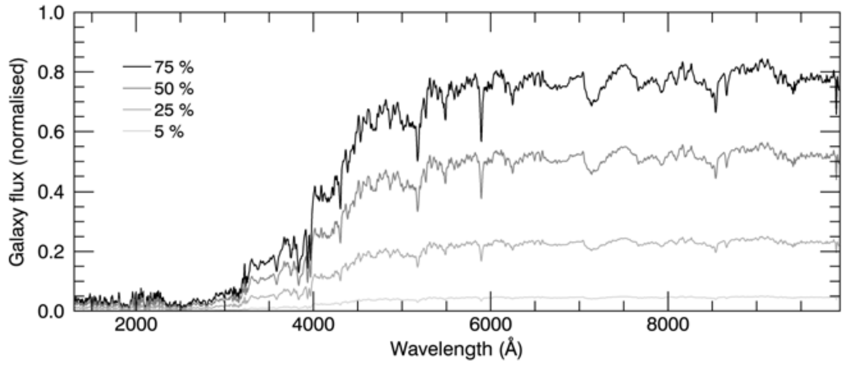
Galaxy spectra have absorption features due to the absorption of atoms and molecules in the atmospheres of the stars in the galaxy. These features can also come from cold gas in the interstellar medium which absorbs energy from photons passing through. Absorption lines are associated with metals, and are therefore more typical of spectra from galaxies containing older stellar populations, where metals have had time to form, such as elliptical galaxies. There are also emission lines present in galaxy spectra which are caused by gas being ionised and re-emitting radiation at certain wavelengths. Emission lines are typically seen in spectra of galaxies containing younger stellar populations, as young stars ionise the gas clouds from within which they are born (Sparke & Gallagher 2007).

In my catalogue of synthetic AGN spectra I have used the galaxy template corresponding to an elliptical galaxy. Considering the similarities between the galaxy templates in Figure 3.5a, it will probably not affect the result significantly if I choose one template or the other. I chose to use the elliptical galaxy template because it has the strongest Ca II break of the galaxy templates (i.e., largest relative difference in flux between the red and the blue wavelengths), meaning it will make a slightly stronger imprint on the synthetic spectra compared to the other templates. The elliptical galaxy template is also representative of the galactic bulge, which is where the AGN is located.

When creating my catalogue of synthetic AGN spectra, I want to have spectra with different amounts of galaxy emission. This is because I want to test how the amount of galaxy emission will affect the ability of the spectral decomposition program to accurately estimate the galaxy contribution in the spectra. In order to quantify the amount of galaxy emission in a given spectrum, I calculate the ratio of galaxy flux to total flux



(a) Galaxy template spectra



(b) Galaxy ratio

Figure 3.5: *Top:* Various galaxy spectral templates. Data from Kinney et al. (1996). The templates covers a wavelength range of 1235–9945 Å (ultraviolet to near-infrared). Sb appears to have quite a bit of noise at shorter wavelengths. Elliptical and S0 (lenticular) galaxies have the reddest spectra, a result of having the largest increase in flux going from UV to optical. *Bottom:* Visualisation of how the size of the galaxy ratio changes the strength of the host galaxy emission for the elliptical template.

over a small wavelength range. I choose to do the measurement over the wavelength range 6000 Å ± 20 Å, which is a part of the galaxy template where the slope is close to flat and the galaxy emission is more or less at its strongest. This choice of normalisation wavelength is arbitrary, but choosing a part of the spectrum where the slope of the galaxy emission is flat and the galaxy emission is strong, makes it easy to make an assessment about how large the galaxy ratio is at this wavelength by simply looking at the final synthetic spectrum. I will normalise the final spectrum at 6000 Å for this purpose.

In order to scale the galaxy template to a specific input galaxy ratio, I first have to calculate the total flux around 6000 Å. This is done by

$$F_{\mathrm{tot}}(6000\,\text{Å}) = \frac{F_{\mathrm{cont}}(6000\,\text{Å}) + F_{\mathrm{Fe\,II}}(6000\,\text{Å}) + F_{\mathrm{Balmer}}(6000\,\text{Å}) + F_{\mathrm{lines}}(6000\,\text{Å})}{1 - \text{galaxy ratio}}. \tag{3.2}$$

Here F_{cont} is the AGN continuum flux, $F_{\mathrm{Fe\,II}}$ is the iron template flux, F_{Balmer} is the Balmer template flux, F_{lines} is the flux contribution from the broad emission lines. After having calculated the total flux, I can calculate what the desired galaxy flux at 6000 Å is as

$$F_{\mathrm{gal}}(6000\,\text{Å}) = \text{galaxy ratio} \cdot F_{\mathrm{tot}}(6000\,). \tag{3.3}$$

I measure the value of $F_{\mathrm{gal}}(6000\,\text{Å})$ from the template and then scale the template to the result from Equation 3.3, before adding the galaxy emission to the spectrum.

3.1.6 Noise

When collecting data from astronomical objects there will always be some noise in the measurements. *Random noise* is caused by statistical fluctuations in the measurements due to the limitations of the measuring device. This will put a limit on the precision that is possible to achieve. *Systematic noise* is caused by a flaw in the design of the experiment or the instrument used (i.e., calibration issues, using the instrument incorrectly), and produce consistent errors in each measurement. If systematic noise is present, it will affect the accuracy of the measurements by shifting all the measurements in a particular direction. The spectral decomposition program may introduce systematic errors.

I add random noise to my synthetic spectra as though they were measured by an instrument, to make the spectra more realistic. Random noise typically follows a normal (Gaussian) distribution with a mean of zero and a standard deviation of one. Such a distribution can be generated by using the IDL function `randomn`. Because I am building the spectra from scratch, I can decide exactly what the noise level should be.

The *signal-to-noise ratio* (S/N ratio or SNR) describes how much noise there is compared to the actual signal (here 'signal' is the synthetic spectrum prior to adding noise):

$$\mathrm{S/N} = \frac{\text{mock spectrum (noiseless)}}{\text{noise}}. \tag{3.4}$$

As an example, a S/N ratio of 100 means that the relative errors in the measurements are at 1 %. The S/N ratio is typically measured in units of 'per pixel' or 'per resolution element' when dealing with observational data. Because I am creating the spectra from scratch, the choice of unit is arbitrary. It would however make most sense to say that the S/N ratio is measured 'per pixel' in my case, where one 'pixel' is the size of the step between two elements in the wavelength array of the spectrum.

When creating a database of synthetic AGN spectra, I make spectra with a range of values for the S/N ratio. Specifically, I am considering values S/N = 5, 7, 10 and 50. This is because I want to see how noise affects the ability of the spectral decomposition program to decompose the spectra. When the spectrum becomes more noisy, important spectral features may be lost in the noise, making it more difficult to identify certain spectral components accurately. Components such as the host galaxy emission and the iron emission might suffer from this because of their narrow spectral features.

I use the chosen values for the S/N ratio to scale the amplitude of the computed Gaussian noise before I add the noise to the synthetic spectra. I do this by simply dividing the Gaussian noise distribution with the S/N ratio. For each spectrum I create, I draw a Gaussian distribution at random, so that the noise distribution will be slightly different each time.

3.2 The final spectrum

My synthetic AGN spectra cover the wavelength range of 1000 Å to 9940 Å, where the upper limit is given by the range of the galaxy template. This means that the spectra starts in the ultraviolet, covers the optical, and ends in the infrared, with a larger part of a spectrum being in the optical. The host galaxy mainly contributes from 4000 Å and up, but including spectral features at lower wavelengths will likely help to constrain the different spectral components in the fitting process. I only include the emission lines that are found at wavelengths larger than 1300 Å, in order not to have too many emission lines that I need to fit later, which would result in a significantly larger number of fitting parameters and would slow down the decomposition process.

The spectra I have made have a constant velocity resolution of 69 $\mathrm{km\,s^{-1}}$ per pixel, identical to the spectra from *Sloane Digital Sky Survey (SDSS)*. This means that each pixel has the same velocity interval. This makes more sense than having a constant wavelength interval, as spectra are often used to measure physical quantities such as linewidths, which are measured in $\mathrm{km\,s^{-1}}$.

The various templates I use to create the synthetic AGN spectra described in Section 3.1 are interpolated to correspond with the wavelength array created with the *SDSS* velocity resolution. If I were to use a larger velocity interval, this would change the visual appearance of the spectra in that the narrowest spectral features would be lost. This would likely affect the ability of the spectral decomposition algorithm to recover the host galaxy component, as the host galaxy (and the Fe II emission) are recognisable by their narrow features. I do not explore this issue in my thesis.

The final synthetic spectrum is normalised at 6000 Å. I normalise my spectra because it is not important for my investigation exactly what the flux is, just the relative strength of the various spectral components. By normalising the spectra I am avoiding numerical problems that could arise from the otherwise small spectral flux values.

3.2.1 Spectral variations

I want my catalogue of AGN spectra to span a range of known spectral properties of AGNs. The spectra of observed AGNs will not necessarily have the same overall shape, amount of iron, amount of host galaxy light, etc. I therefore vary these quantities when I am creating the spectra for my catalogue.

To do so, I create a parameter space that I loop over when creating the spectra, covering a range of known spectral properties. The parameter space I am using is summed up in Table 3.2. This parameter space has 288 possible combinations of parameter values. I run 150 realisations of each combination of parameter values, where each realisation will have different randomly generated Gaussian noise, resulting in a total of 43,200 spectra!

Below I sum up the spectral properties that is accounted for by varying the input parameters the way I do:

- *Power-law index α_ν:* By having different power-law indices I am mimicking different amounts of SMBH mass accretion. A higher accretion rate leads to a steeper (i.e., bluer) spectrum ($\alpha_\nu < 0$) and a lower accretion rate leads to a flatter spectrum ($\alpha_\nu > 0$). I therefore want to have synthetic spectra that are steep (I use $\alpha_\nu = -0.5$), intermediate (I use $\alpha_\nu = 0.5$) and flat (I use $\alpha_\nu = 1.0$). Figure 3.1 illustrates the different values of α_ν that I am considering.

- *Galaxy ratio:* By having different amounts of galaxy emission relative to the total spectral emission, I am simulating different types of AGNs. I am specifically considering galaxy ratios of 0.05, 0.25, 0.05 and 0.75, which is measured at 6000 Å in the spectrum. In spectra where the galaxy emission is low, the AGN will be dominating the spectrum, as is the case with quasars. In spectra where the galaxy emission is high, the AGN will appear weaker, as is the case with nearby Seyfert galaxies. The appearance of the galaxy template for the different galaxy ratios I am considering are illustrated in Figure 3.5b.

- *Fe* II *EW:* I have chosen to look at three different values: a small value, and intermediate value and a high value, as illustrated in Figure 3.2b. The values I have chosen for Fe II EW are based on the measurements of BG92, who measured the values to range from zero to 114 Å, with 57 Å as the intermediate value. I am using a lower value of 6 Å (corresponding to approximately 5 % of the maximum value) rather than zero with the later spectral modelling in mind: The iron fitting parameter will be constrained to have only positive values as a negative iron emission would be non-physical. So in order to differentiate between the instance of the fitting program hitting the lower limit of zero and the instance of it actually estimating the parameter value correctly, a non-zero lower limit should be used as the lowest limit when creating the spectra.

- *Fe* II *linewidth:* Some AGNs have been seen to have narrow iron emission (Laor et al. 1997), while others have been seen to have broad iron emission with linewidths similar to that of the broad emission lines (Véron-Cetty et al. 2004). In addition to varying the equivalent width of the iron template, I therefore also vary the linewidth of the iron templates. I test both $\Delta v_{\mathrm{FWHM}} = 2000\ \mathrm{km\,s^{-1}}$ (narrow) and $4000\ \mathrm{km\,s^{-1}}$ (broad) for the optical template, and set the UV linewidth to be twice the width of the optical template in both cases, as discussed in Section 3.1.2. When the lines in iron template are narrow, its sharper spectral features resemble the features in the galaxy template, which might make it difficult for the decomposition algorithm to tell the two components apart. The spectral features will be smoothed out as the template is broadened, as illustrated in Figure 3.2c. I therefore choose to include some spectra with a narrow iron template and some with a broad iron template in my catalogue of spectra in order to test how this could influence the ability to recover the galaxy ratio.

Some examples of final spectra are shown in Figure 3.6 (galaxy ratio = 0.05) and Figure 3.7 (galaxy ratio = 0.75).

Table 3.2: The parameter space used to generate synthetic AGN spectra.

Parameter	**Adopted values**			
S/N ratio	5	7	10	50
Galaxy ratio	0.05	0.25	0.50	0.75
Power-law index, α_ν	−0.5	0.5	1.0	
Fe II EW (Å)	6	57	114	
Fe II linewidth ($\mathrm{km\,s^{-1}}$)	2000	4000		

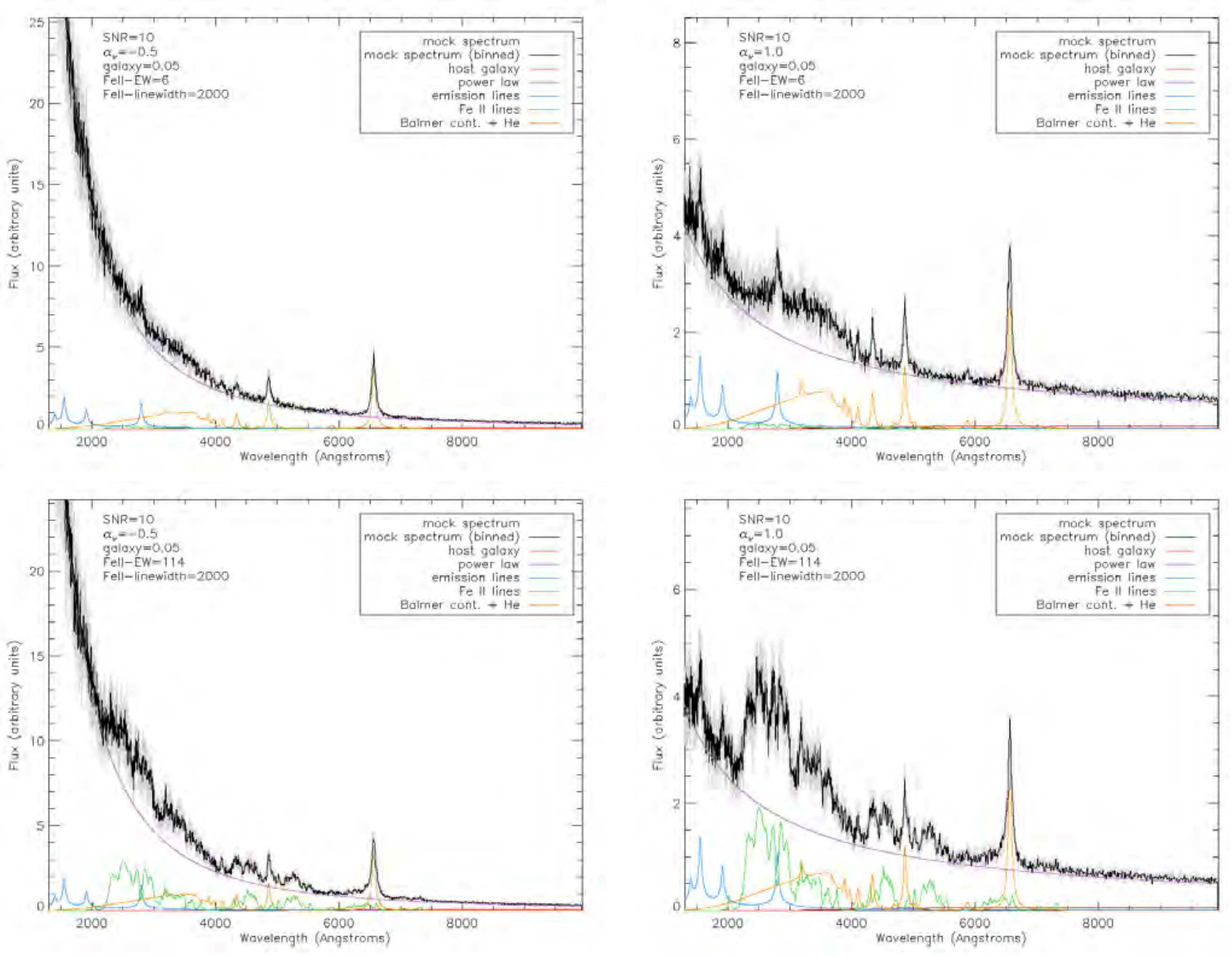


Figure 3.6: Examples of synthetic AGN spectra that I will perform a spectral decomposition of. Shown here are examples of how the power-law slope α_ν and Fe II EW influence the shape of the spectrum when the galaxy ratio is 0.05, measured at 6000 Å. Fe II linewidth = 2000 km s^{-1} and S/N = 10 in all the figures. It is impossible to see the galaxy component in any of the figures. *Left:* $\alpha_\nu = -0.50$ (steep). The spectrum closely follows the shape of the power-law. *Right:* $\alpha_\nu = 1.0$ (flat). The other spectral components becomes more visible when the spectrum is flatter. *Top:* Fe II EW = 6 Å. *Bottom:* Fe II EW = 114 Å.

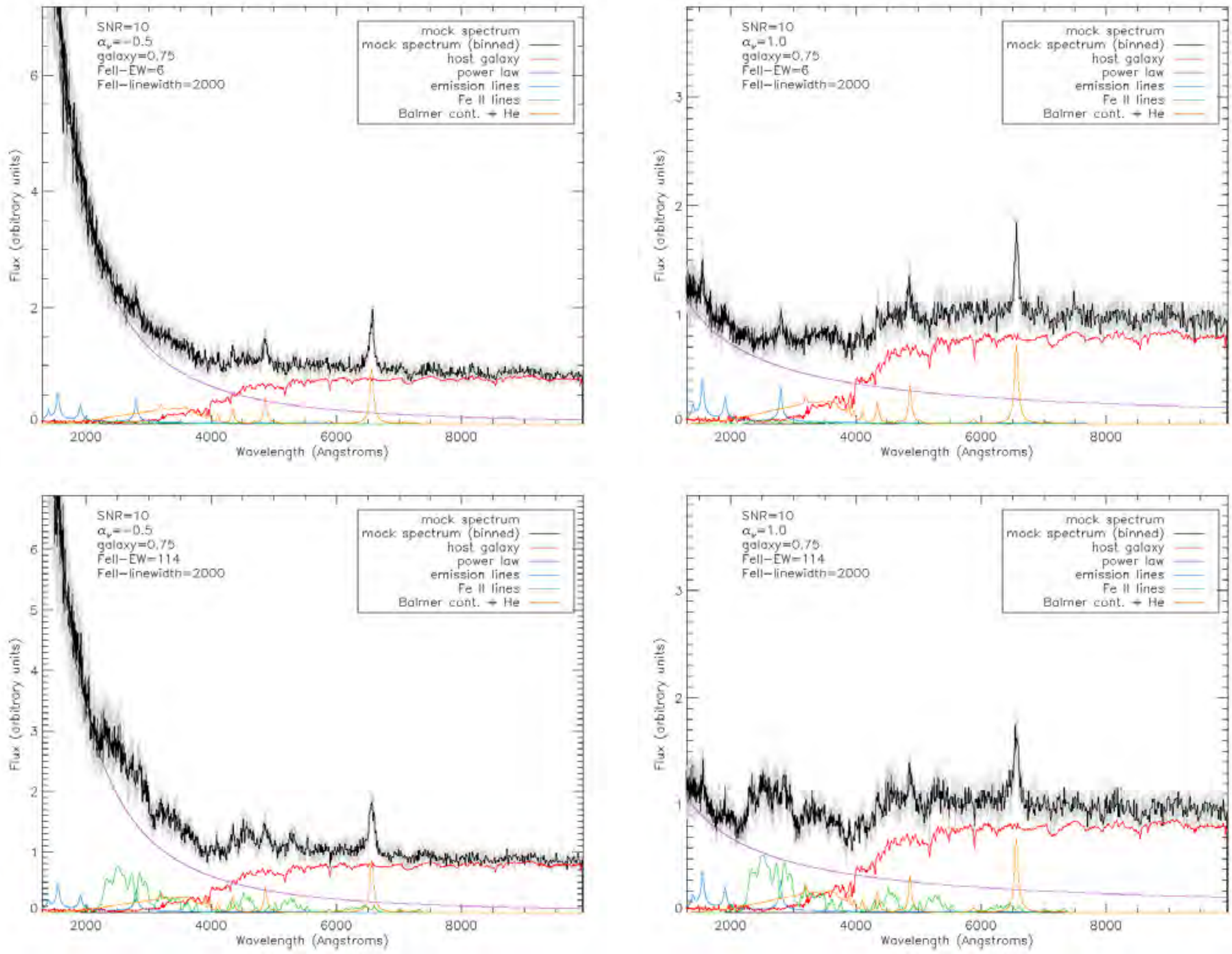


Figure 3.7: Examples of synthetic AGN spectra that I will perform a spectral decomposition of. Shown here are examples of how the power-law slope α_ν and Fe II EW influence the shape of the spectrum when the galaxy ratio is 0.75, measured at 6000 Å. Fe II linewidth = 2000 km s^{-1} and S/N = 10 in all the figures. It is possible to see the contribution from the host galaxy here, as opposed to Figure 3.6. *Left:* $\alpha_\nu = -0.50$ (steep). The spectrum closely follows the shape of the power-law. *Right:* $\alpha_\nu = 1.0$ (flat). The other spectral components becomes more visible when the spectrum is flatter. Due to the strength of the galaxy emission in these spectra, this particular spectrum almost looks entirely flat. *Top:* Fe II EW = 6 Å. *Bottom:* Fe II EW = 114 Å.

4 Modelling AGN spectra

In Chapter 3 I created a database of synthetic AGN spectra. Next I will attempt to recover the parameter values I used to create the synthetic spectra by applying a spectral decomposition method to the spectra. This will reveal to what extent it is possible to decompose the spectra accurately and precisely, and ultimately recover the true parameter values. Here I describe the method I am using to model the spectra, before I present the results in Chapter 5.

4.1 The spectral decomposition program

I have written my own spectral decomposition program using the IDL procedure `MPFITFUN` (Markwardt 2009) to model the synthetic AGN spectra. `MPFITFUN` is a non-linear least-squares curve fitting algorithm based on the Levenberg-Marquardt algorithm (Moré 1978). `MPFITFUN` will fit a user-supplied model function to user-supplied data points by adjusting a set of parameters. In my case the data points are the spectral flux value for each wavelength in a given synthetic spectrum in my catalogue. An array of 1σ uncertainties values must also be supplied by the user, which in my case is simply the array of flux values divided by the S/N ratio of the given spectrum.

A particular model must be assumed for `MPFITFUN` to be able to fit a function to the spectral data I have created. I use a model of the same functional form and recipe as I used to create the synthetic spectra. By doing this I am effectively performing a zeroth-order test of the spectral decomposition program I have written. In other words, I am testing how well the synthetic AGN spectra can be decomposed in the scenario where we know exactly how to describe the spectral components.

It might seem trivial to fit the spectra in my catalogue by the same recipe as I created them. One would expect the results to be very good and that this study would be a waste of time. However, I am testing how the noise (S/N ratio) affects the ability of the spectral decomposition program to recover the true parameter values. As will be seen in Chapter 5, the true parameter values are not necessarily well-recovered. Even if the program were to do a good job recovering the true parameter values in all cases, it would still have been important to have done the test I am doing, if only to make sure that it actually works they way people expect it to.

The model I use to decompose my spectra is a combination of the following functions and templates:

- *Continuum emission:* The continuum is fitted by a power-law function. This is the same functional form that I used to create the continuum in the synthetic spectra, but the parameters describing the shape of the power-law function will be determined in the modelling process.

- *Fe* II *emission:* The optical and UV iron emission, as was illustrated in Figure 3.2a, are fitted separately by adjusting the two Fe II templates I used to create the spectra to the input spectra. The two templates are fitted separately because the strength of the various multiplets of Fe II can vary in strength in different observed AGNs (Vestergaard & Wilkes 2001), in case I want to test the spectral decomposition method on real AGN spectra in the future.

- *Balmer (+ He) emission:* For the lower-order lines, the decomposition program will fit a Lorentzian function to each of the prominent Balmer lines Hα, Hβ and Hγ. This means that there are some minor spectral features for $\lambda > 4200$ Å that are not fitted by the model. The higher-order Balmer lines and the Balmer continuum are fitted as one template, see Figure 3.3 for illustration, which is the same template as I used when creating the spectra.

- *Broad emission lines:* The four broad emission lines I added to the synthetic spectra are fitted by individual Lorentzian functions, which is the same functional form that I used to create the lines in the spectra.

- *Host galaxy emission:* The galaxy contribution is fitted by scaling the same galaxy template I used to create the synthetic spectra.

These spectral components are fitted to the data by the spectral decomposition program by adjusting several parameters, see Section 4.2 for a detailed list. All these parameters are fitted to the synthetic spectrum simultaneously by the decomposition program. The value of the parameters are varied slightly from one iteration to the next in order to minimise the difference between the model function and the input data. This difference is quantified by the χ^2 value, which is defined as

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{y_i - y(x_i, p)}{\sigma_i} \right]^2, \tag{4.1}$$

where (x_i, y_i) are the data points of the input spectrum with a standard deviation of σ_i. In my case x_i will be the i'th wavelength and y_i will be the flux value at that wavelength. The model function is described by $y(x_i, p)$, where p is the array of fitting parameters. χ^2 provides a measure of the difference between the fitted model function and the input spectrum, and can therefore be used to evaluate the goodness of fit. Because the standard deviation σ appears in the denominator of Equation 4.1, the value of χ^2 will tend to be lower for spectra with more noise (i.e., larger standard deviation).

A more useful measurement of the goodness of fit, is the so-called *reduced chi-squared* χ^2_ν, which is defined as χ^2 per degree of freedom. It has the useful property that $\chi^2_\nu \lesssim 1$ indicates that the model is describing the data well, while $\chi^2 > 1$ indicates that the model does not fully capture the data. $\chi^2 < 1$ indicates that the model is 'over-fitting' the data and $\chi^2 \gg 1$ indicates that the model poorly describes the data.

The software codes I have written for this thesis are listed in Appendix A with short descriptions. I ran the software codes on multiple servers on a supercomputer belonging to the High Performance Computing Centre at the University of Copenhagen over many weeks.

4.2 The fitting parameters

When writing the model function, I must decide which parameters the program should be free to fit when modelling the data. The parameters I have chosen to be fitted by the program are:

- `p[0]`: Power-law index, α_ν
- `p[1]`: Power-law normalisation constant, C
- `p[2]`: Fe II optical template flux scaling constant

- `p[3]`: Fe II optical linewidth
- `p[4]`: Fe II UV template flux scaling constant
- `p[5]`: Fe II UV linewidth
- `p[6]`: Hα line flux scaling constant
- `p[7]`: Hα linewidth
- `p[8]`: Hβ line flux scaling constant
- `p[9]`: Hβ linewidth
- `p[10]`: Hγ line flux scaling constant
- `p[11]`: Hγ linewidth
- `p[12]`: Balmer continuum + He emission lines template scaling constant
- `p[13]`: Balmer continuum + He emission lines linewidth
- `p[14]`: Si IV line flux scaling constant
- `p[15]`: Si IV linewidth
- `p[16]`: C IV line flux scaling constant
- `p[17]`: C IV linewidth
- `p[18]`: C III line flux scaling constant
- `p[19]`: C III linewidth
- `p[20]`: Mg II line flux scaling constant
- `p[21]`: Mg II linewidth
- `p[22]`: Host galaxy template scaling constant

In addition to the parameters mentioned above, I also keep track of: the S/N ratio which is estimated from the input spectrum across four different wavelength ranges (see Section 5.1.2 for details), the value of χ^2_ν, the number of iterations and attempts needed to achieve a fit. The host galaxy ratio and the Fe II EW are both measured from the final fitted model and its components.

The residuals between the data and the model are calculated as: residuals = synthetic spectrum − fitted model.

4.2.1 Parameter constraints

`MPFITFUN` allows the user to set constraints limiting the range of possible values for the fitting parameters. The limits can be either a lower or an upper limit, or both. I do not want to constrain the parameter values too much, as it can keep `MPFITFUN` from being able to find a fit at all. I also do not want to use any prior knowledge I have about the real parameter values to force the program to achieve a good fit. I simply add enough constraints to keep the parameters within the realm of what can be deemed physical for an AGN spectrum, while giving the parameters plenty of wiggle room.

I started by only setting lower limits for the parameters, but after running the decomposition program on a set of synthetic spectra it became evident that I needed upper limits as well in several cases, especially for the linewidths. Introducing an upper limit helped the fitted models from settling on poor fits.

The limits I have imposed on the parameters are:

- *Power-law index α_ν:* I set a lower limit of -2 and an upper limit of 3. AGNs typically have $0 \lesssim \alpha_\nu \lesssim 1$ (Peterson 1997), and I am applying the decomposition program to spectra I have made with $\alpha_\nu = -0.5$, 0.5 and 1.0, so the limits should provide sufficient wiggle room.

- *Broad emission lines:* All the broad emission lines are limited to fall between 1,500–10,000 $\mathrm{km\,s^{-1}}$. Broad emission lines average around 5,000 $\mathrm{km\,s^{-1}}$ in real AGN spectra (Peterson 1997).

- *Balmer template:* The linewidths of the Balmer template are also set to fall in the range 1,500–10,000 $\mathrm{km\,s^{-1}}$. The Balmer lines are typically found to be around 3,500 $\mathrm{km\,s^{-1}}$ (Borosen & Green 1992).

- *Fe II template:* The linewidth of both the optical and the UV Fe II templates are set to have a lower limit of 1000 $\mathrm{km\,s^{-1}}$. This is the intrinsic linewidth of the Fe II templates, thus it is not possible to have the iron templates being fit with narrower linewidths than that.

I have set all the remaining fitting parameters to have a lower limit of zero, as negative values would be unphysical.

It is also possible to tie fitting parameters to each other. The Balmer linewidths are the only parameters I tie, in that I am forcing these linewidths to be identical. Specifically, the linewidth of Hβ, Hγ and the broadening of the Balmer continuum are tied to the width of Hα. Hα is the most prominent feature in the Balmer template, thus it should be easiest to estimate the linewidth using that line. These linewidths can be tied to each other because all the Balmer features are produced by the same gas, thus they will be broadened by the same amount.

4.2.2 Starting parameters

I must set starting values for all the fitting parameters in order to get the spectral decomposition code started. The starting parameters acts as rough initial guesses of what I think the values should end up being. In order for the fitted results to avoid being biased by my choice of values for the starting parameters, I let the values of the starting parameters be drawn randomly from a list of predefined choices. For each parameter I want to fit, I create a list of three possible start values for the program to choose from. One of these values is below the true value, one is above, and one is close to it.

The program takes the starting parameters and creates an initial spectrum. Then the program tweaks the different components more and more for each iteration in order to get a better overall agreement between the data and the model. If the relative error or the reduction in the sum of squares between two consecutive iterations is less than 10^{-10} (this is the limit is set by `MPFITFUN`, but it can be adjusted by the user), the program is satisfied with the model and I store the model in a file.

I allow the starting parameters to be redrawn up to 9 times if the program is not able to find a fit within the maximum limit of 300 iterations (limit set by me) for the current starting parameters. This gives the program a total of 10 attempts to find a fit. I have also implemented an upper limit of $\chi^2_\nu = 20$, which forces the program to try again with a new set of starting parameters if the fit found by the program exceeds this limit. This limit is quite loose, and is implemented in order to avoid ending up with models that are too poor.

4.3 Examples of fitted spectra

After the entire catalogue of spectra has been modelled, χ^2_ν is found to be between 1.037 and 19.82 when considering all the fitted spectra. The lowest value is very close to 1, indicating a good fit. The higher value is close to the upper limit of $\chi^2_\nu = 20$, indicating that setting a limit had an effect. Such high values are too high to be considered good. However, only 0.2 % of the spectra have $\chi^2_\nu > 5$, so it is not a widespread problem that χ^2_ν is high.

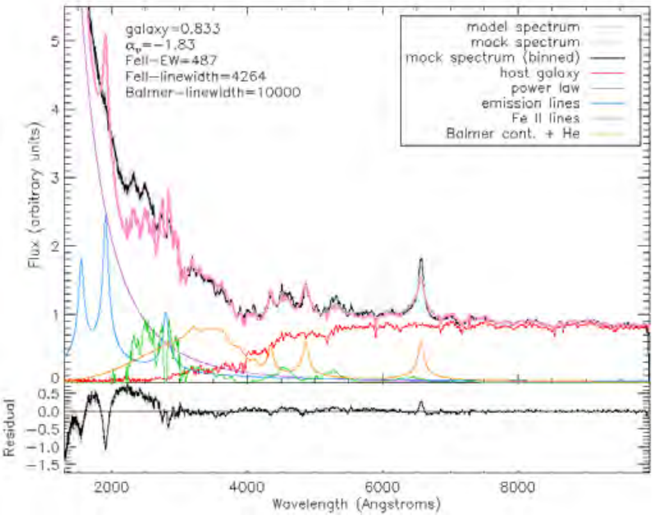
The higher values of χ^2_ν tends to be more common for $\alpha_\nu = -0.5$ (steep) than the other values of α_ν. This could be due to the fact that a steeper spectrum will wash out the other spectral components found in the UV when they are weak, as illustrated in the left panels of Figure 3.6 and Figure 3.7. This can make it difficult for the program to constrain the different spectral components, which will affect the goodness of fit.

The higher values of χ^2_ν do not seem to depend on the values of other input parameters in a specific way, in that a high value of χ^2_ν can occur for any galaxy ratio, S/N ratio and iron emission. Two examples of fitted spectra with some of the highest values of χ^2_ν are shown in Figure 4.1. Both examples are considered to be successfully modelled by the decomposition program (exit status = 2 in `MPFITFUN`, meaning the relative error between two consecutive iterations is less than 10^{-10}). These fits are obviously not good enough to be used to estimate the galaxy ratio from, which is the main parameter I am interested in estimating. But given that most of the fitted models have $\chi^2_\nu < 5$, fitted models with higher values of χ^2_ν than this can easily be avoided in future studies by imposing a stricter upper limit for χ^2_ν. For now, this limited number of poor fits will not affect the results significantly as I am using statistics that are robust to outliers (see Section 5.1).

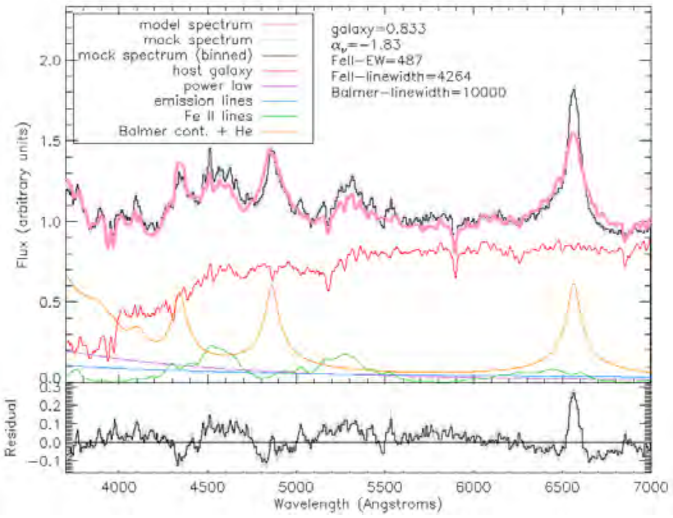
Some examples of successfully fitted spectra with a low χ^2_ν are shown in Figure 4.2 for two different S/N ratios. It is easier to find a model to describe spectra with low S/N ratio because there is more wiggle room when the uncertainty is large. Then the line representing the fitted model does not necessarily fall in the centre of the noise as the true signal does, but may be shifted to higher or lower values, as seen in Figure 4.2a-b. For spectra with high S/N ratios ($\sim$50), it is easier to see which areas of the spectrum the fitted model can struggle with capturing. An example is shown in Figure 4.2c-d. A feature in the spectra that the fitting algorithm most often struggles to get right, is the Hα peak. Because this line is so pronounced, it will easily be seen in the residuals if the linewidth or line flux is slightly off. Another part of the spectrum that can be challenging to fit, is the area around 3200 Å, where the Balmer, Fe II, and galaxy emission are blended.

In a total of 43,200 spectra, only 162 of them failed to be fitted, which is about 0.4 % of the spectra in the catalogue. There does not appear to be any patterns in which combinations of spectral parameters fails to be fitted. The results from the models that failed to find a fit to the spectra within the maximum number of iterations and attempts, are not included in the analysis.

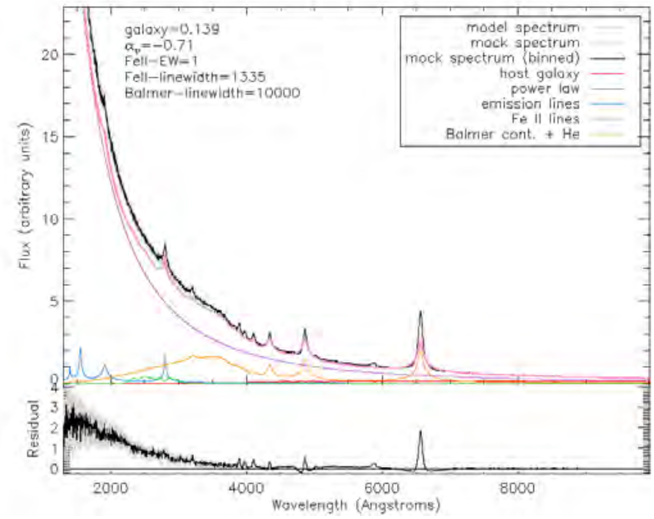
Despite the fact that I am using the same recipe to recover the parameter values that I used to create the synthetic spectra, the spectral decomposition program is not always able to recover the true parameter values in the modelling. Furthermore, the scatter in estimated parameter values can be large. I discuss these issues in-depth when I present the results from the modelling in Chapter 5.



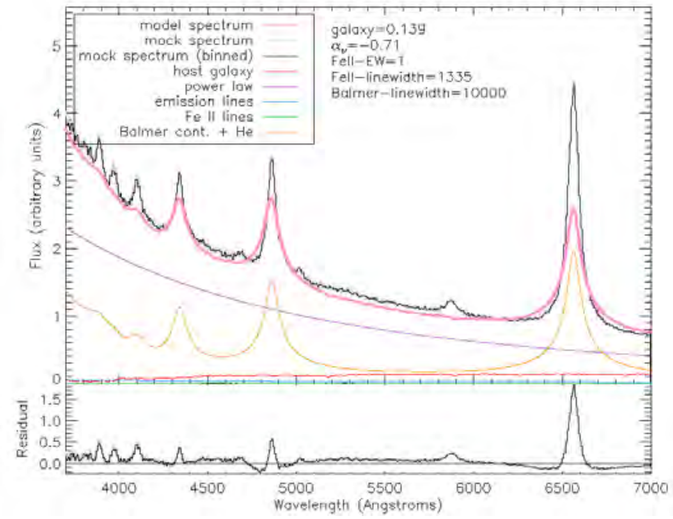
(a) $\chi^2_\nu = 19.50$



(b) $\chi^2_\nu = 19.50$ (zoomed in)

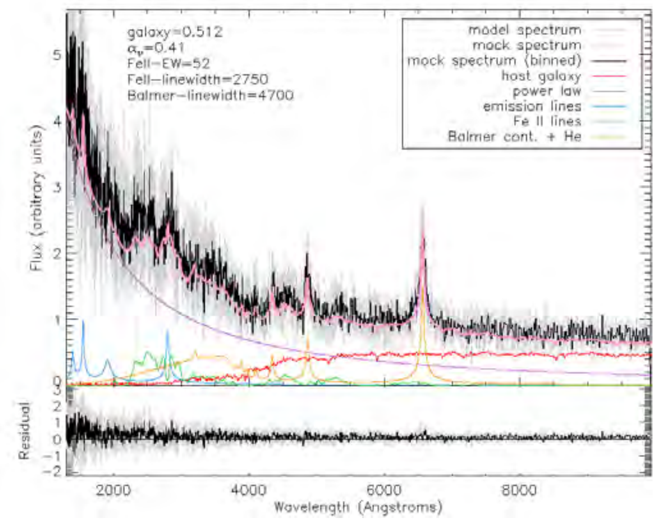


(c) $\chi^2_\nu = 16.08$

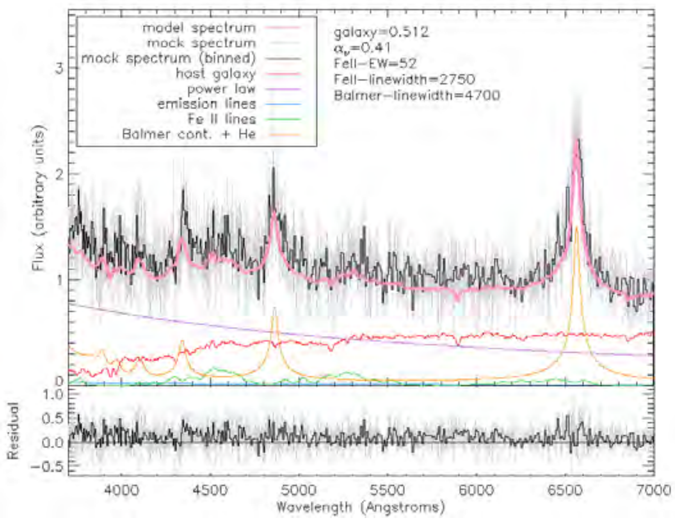


(d) $\chi^2_\nu = 16.08$ (zoomed in)

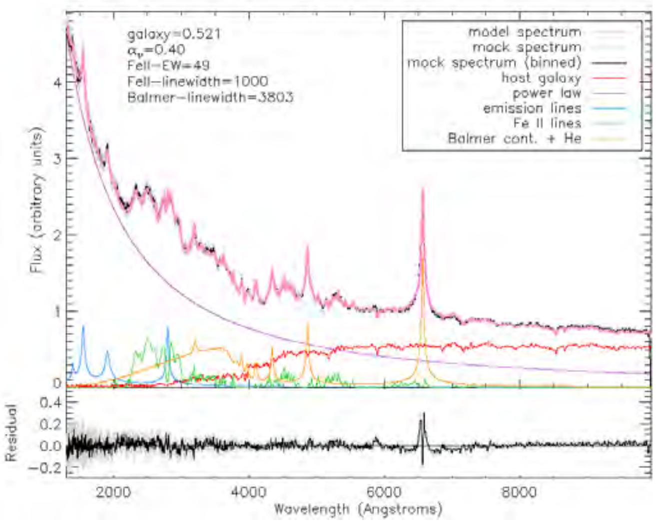
Figure 4.1: Examples of poorly fitted spectra. They have both been accepted as successful models by the program (exit status = 2 in `MPFITFUN`, meaning the relative error between two consecutive iterations is less than 10^{-10}). The right panels shows zoomed-in versions of the spectra in the left panels. The two spectra shown here have some of the highest values of χ^2_ν in the catalogue.
Top: Large amount of galaxy emission (galaxy ratio = 0.75). The broad emission lines are highly overestimated and the Balmer lines are much too broad, thus the peaks are not fitted properly. The other components are not fitted well either. The parameter values describing the input spectrum that was modelled were: galaxy ratio = 0.75, α_ν = –0.50, Fe II EW = 114 Å, Fe II linewidth = 2000 km s^{-1}, S/N = 50.
Bottom: Small amount of galaxy emission (galaxy ratio = 0.05). Again, the linewidth of the Balmer lines are much too broad, and the other components are off as well. The parameter values describing the input spectrum that was modelled were: galaxy ratio = 0.05, α_ν = –0.50, Fe II EW = 6 Å, Fe II linewidth = 4000 km s^{-1}, S/N = 50.



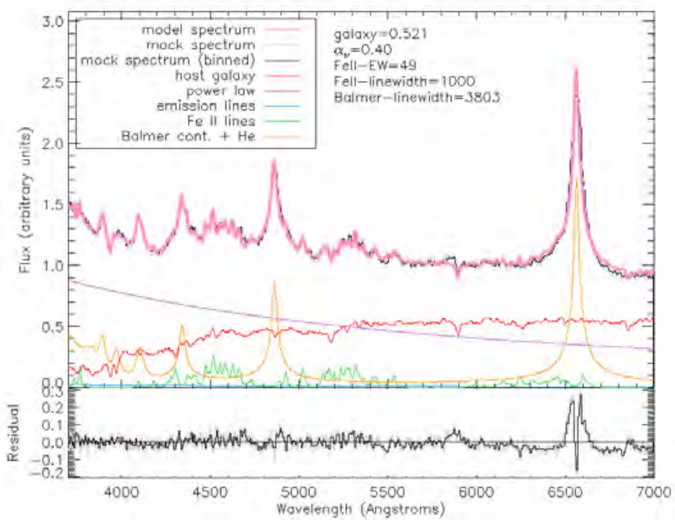
(a) $\chi^2_\nu = 1.387$



(b) $\chi^2_\nu = 1.387$ (zoomed in)



(c) $\chi^2_\nu = 2.081$



(d) $\chi^2_\nu = 2.081$ (zoomed in)

Figure 4.2: Examples of spectra with a low χ^2_ν, indicating a good fit. The right panels shows zoomed-in versions of the spectra in the left panels. The parameter values describing the input spectra that was modelled here are: galaxy ratio $= 0.50$, $\alpha_\nu = 0.50$, Fe II EW $= 57$ Å, Fe II linewidth $= 2000$ km s^{-1}. *Top:* Low S/N ratio (S/N $= 5$). The fitted spectrum (pink line) is slightly shifted to lower values compared to the centre of the noise fluctuations. *Bottom:* High S/N ratio (S/N $= 50$). The spectrum is mostly well-fitted, but clearly struggles to get the Hα peak right, which can be seen in the plot of the residuals in the bottom of the figure.

5 Results

In Chapter 3 I described how I have created a catalogue of synthetic AGN spectra from a variety of known spectral features. The spectra in the catalogue vary in signal-to-noise ratio, continuum emission, Fe II emission, and host galaxy emission. By varying these quantities, I am effectively simulating different types of AGNs. I then used the spectral decomposition program outlined in Chapter 4 to model the entire catalogue of synthetic AGN spectra in order to recover the known parameter values that describe the different contributions to the spectra. In this chapter I present the results from the modelling, before discussing their implications in Chapter 6.

5.1 Parameter estimation

The goal of the spectral decomposition program is to estimate the value of 23 spectral fitting parameters, listed in Section 4.1, for a large number of AGN spectra, with special emphasis on the accuracy and precision in recovering the host galaxy contribution to the input spectra. The spectra are created from different combinations of input parameter values, listed in Table 3.2.

I generate synthetic spectra for each combination of input parameter values (there are 288 possible combinations in total). I make 150 *realisations* of each spectrum by adding Gaussian noise to the spectra. The Gaussian noise is distributed differently and randomly for each realisation, but with the same S/N ratio per pixel for all the 150 realisations. I then apply the spectral decomposition method to all the realisations. When the decomposition program has finished, it has calculated 150 estimated values for each fitting parameter for each combination of input parameters. This gives me a distribution of estimated values for each input parameter, which allows me to do a statistical study of how accurately and precisely the parameters are estimated.

I use the median m to describe the accuracy of a set of estimated parameter values rather than the mean μ. I do this because the distributions of estimated parameter values calculated by the spectral decomposition program sometimes are quite asymmetric around the mean. I use an interpercentile range IPR to describe the precision rather than the standard deviation σ. This is because the distributions of estimated parameter values can have many outliers. The median and IPR are robust to outliers, which means that the inclusion or exclusion of a few outliers will not change the values of the statistics significantly. I define the interpercentile range as $IPR = P_2 - P_1$, where I use a lower percentile of 15.9 % (P_1) and an upper percentile of 84.1 % (P_2). This mean that my interpercentile range will contain 68.2 % of the central observations in the distributions. The distribution of a set of estimated parameter values is then summarised by $m \pm IPR/2$. This will be analogous to $\mu \pm \sigma$ whenever the distribution is close to a normal distribution.

When I discuss the results of the modelling, I will often talk about the size of the *offset* between the median of a distribution of estimated parameter values relative to the true parameter value. The size of the offset describes to what degree the distribution is shifted relative to the true parameter value.

I will also quote my results in terms of the *percent error* (or *percent difference*) between the median and the true parameter value. This number describes by how large a percentage the median of a distribution of estimated parameter values are overestimated (positive value) or underestimated (negative value) relative to the true value. The percent error is calculated by

$$\%\,\text{error} = 100\,\% \cdot \frac{\text{observed value} - \text{true value}}{\text{true value}}\,. \tag{5.1}$$

Similarly, the precision can also be given as a percentage:

$$\%\,\text{of true value} = 100\,\% \cdot \frac{IPR/2}{\text{true value}}\,. \tag{5.2}$$

I define outliers to be those estimated parameter values that are either less than $P_1 - 3 \cdot IPR$ or greater than $P_2 + 3 \cdot IPR$.

5.1.1 Main result: Estimated galaxy ratio

I am first and foremost interested in estimating the ratio of host galaxy emission to the spectral emission (measured at 6000 Å), which is the topic of this thesis. The ability of the spectral decomposition program to estimate the galaxy contribution will depend on the value of the power-law index α_ν, Fe II EW, Fe II linewidth, S/N ratio, and the galaxy ratio of the input spectrum that is being modelled by the spectral decomposition program.

A quick overview

Table 5.1 gives an overview of the values of the estimated galaxy ratios (outliers included) when considering the results from *all* the fitted models in my catalogue. For each combination of input parameters, I locate the lowest estimated galaxy ratio in the distribution, the median of the distribution and the highest estimated galaxy ratio in the distribution. I do this for all the possible combinations of input parameter values. This gives me a distribution of the lowest estimated galaxy ratios in my catalogue (Column 1), a distribution of the median values of the estimated galaxy ratios in my catalogue (Column 2), and a distribution of the highest estimated galaxy ratios in my catalogue (Column 3). Each column gives the minimum (a), the median with a standard deviation (b), and the maximum (c) of each of these distributions.

Column 1a shows that the galaxy ratio can be estimated to be zero for *all* input galaxy ratios, even when the input galaxy ratio is as high as 0.75. Comparing the values in Column 1c with the input galaxy ratio, shows that there are some combinations of input parameter values where the lowest value in the distribution of estimated galaxy ratios will be *larger* than the true value. This means that it will be impossible to recover the true galaxy ratio in some cases. This can happen for all the input galaxy ratios. However, it

Table 5.1: Span of values for the estimated galaxy ratio for the catalogue of synthetic spectra.

Input galaxy ratio	**Lowest estimates** (1)			**Median estimates** (2)			**Highest estimates** (3)		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
0.05	0.000	0.033 ± 0.033	0.122	0.048	0.100 ± 0.027	0.136	0.064	0.335 ± 0.220	0.729
0.25	0.000	0.227 ± 0.079	0.304	0.248	0.285 ± 0.021	0.324	0.261	0.507 ± 0.169	0.853
0.50	0.000	0.471 ± 0.153	0.529	0.497	0.519 ± 0.012	0.538	0.509	0.672 ± 0.110	0.915
0.75	0.000	0.722 ± 0.160	0.760	0.729	0.754 ± 0.008	0.763	0.746	0.855 ± 0.061	0.996

is more common that the lowest estimated galaxy ratio in a distribution is lower than the true value, which can be seen by comparing the values in Column 1b with the input galaxy ratio. Thus it will in many cases be possible to recover the true value.

Column 2 shows the span of median value of the estimated galaxy ratios in the catalogue. The median value of a distribution of estimated galaxy ratios is a measure of the accuracy of the distribution. If the median value is close to the input galaxy ratio, the estimated values are said to be accurate. If the median is significantly larger or smaller than in value compared to the input galaxy ratio, the estimated values are said to be inaccurate, because the distribution will be shifted away from the true value. The values in Column 2b shows that the distributions of estimated galaxy ratios tend to *over*estimate the galaxy ratio, because the values are larger than the input galaxy ratio. The exception is when the input galaxy ratio is 0.75 where the galaxy ratio tends to be *under*estimated. Column 2b also shows that the scatter *de*creases when *in*creasing the input galaxy ratio from 0.05 to 0.75. This indicates that the galaxy ratio is more precisely determined for the higher galaxy ratios.

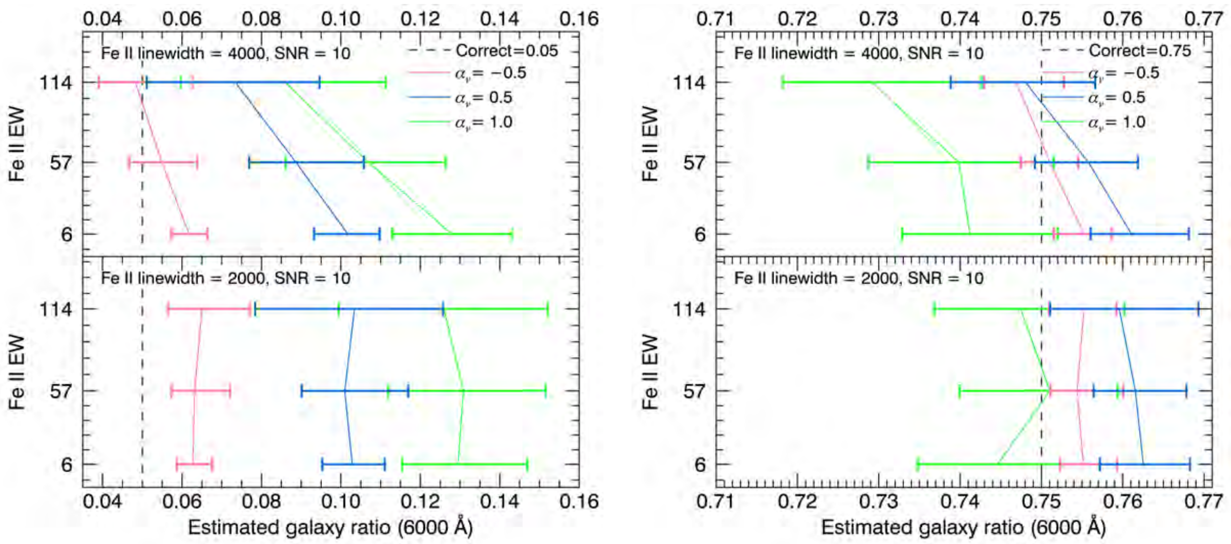
Column 3c shows how much the galaxy ratio can be overestimated in the most extreme cases. It is surprising to see how large the estimated values can become, even for the smallest galaxy ratio of 0.05. Column 3a shows that for an input galaxy ratio of 0.75, the highest estimated galaxy ratio in the distributions can be *lower* than the true value. This indicates that there are distributions of estimated galaxy ratios are underestimated to such a large degree that the true galaxy ratio cannot be recovered.

Recovering the true value

The goal of the spectral decomposition program is ultimately to recover the true galaxy ratio. The values in Table 5.1 have already revealed that this cannot always be done. By taking a closer look at the result for different combinations of parameters, it is possible to map out for which combinations of input parameters this works well, and for which combinations it does not.

Figure 5.1 shows how the accuracy of the distributions of estimated galaxy ratios depend on the values of the various input parameters, shown for input galaxy ratios of 0.05 and 0.75 (the result for input galaxy ratios of 0.25 and 0.50 are shown in Figure B.2 and Figure B.3 in Appendix B.1, respectively). The distributions of estimated galaxy ratios can be seen to be shifted relative to the true value. In most cases the distributions are shifted towards *higher* values, which means that *the galaxy ratio tends to be <u>over</u>estimated* for most of the combinations of input parameter values. I go into more details below. This means that the host galaxy contribution to the AGN spectrum tends to be estimated to be a larger part of the total AGN flux than it actually is.

Comparing Figure 5.1a and b shows that *the estimated galaxy ratio is closest to the true value (i.e., more accurate) when there is a large amount of galaxy emission in the spectra*. That is, the galaxy ratio is more accurately estimated when the input galaxy ratio is 0.75 compared to when it is 0.05. For an input galaxy ratio of 0.05, when considering all combinations of input parameter values in my catalogue, I find the offsets in estimated values to be within $[-0.002, +0.080]$ of the true value (meaning that the galaxy ratios are under- or overestimated to be $[-6, 160]\,\%$ that of the true value). Thus the positive offset of the distributions can be large ($+0.080$). In contrast, for an input galaxy ratio of 0.75, the estimated galaxy ratios are offset to within $[-0.021, +0.005]$ of the true value (meaning that the galaxy ratios are over- or underestimated to be $[-3, 2]\,\%$ that of the true value). This means that the negative offset can be large (-0.021) for a galaxy ratio of 0.75, but the absolute offset is only one quarter of the offset for a galaxy ratio of 0.05. This shows that the galaxy ratio is more accurately determined for a galaxy ratio of 0.75 compared to that of 0.05.



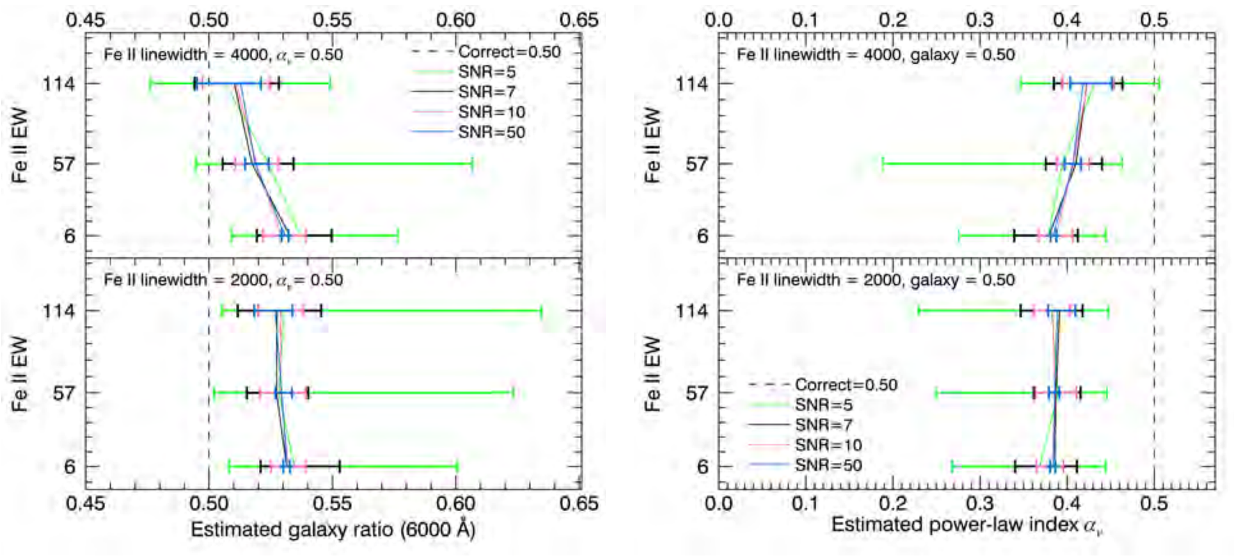
(a) Galaxy ratio = 0.05 **(b)** Galaxy ratio = 0.75

Figure 5.1: Distributions of estimated values for the galaxy ratio as a function of Fe II EW for different values of α_ν, shown for input galaxy ratios of 0.05 (left) and 0.75 (right). Here S/N = 10. The dashed line represents the true parameter value. The galaxy ratio is almost always overestimated when the input galaxy ratio is low, while it can be over- or underestimated when the input galaxy ratio is high, depending on the combination of parameters. The estimated galaxy ratios are offset by [–0.002 +0.080] (left; estimated to be [–6, 160] % that of the true value) and offset by [–0.003, +0.005] (right; estimated to be [–2.8, 1.7] % that of the true value) when considering all combinations of parameters shown here. In the right panels the true value is within 1–2σ of the estimated galaxy ratios, while in the left figure it can be up to 8σ. The estimated galaxy ratio is highly dependant on the input value of α_ν, and more so for lower galaxy ratios. α_ν = –0.5 (steeper in the UV) tends to give the most accurately estimated galaxy ratios relative to the true value. The precision tends to become poorer with increasing values of α_ν and Fe II EW, likely due to degeneracies. The estimated values are close to constant with increasing Fe II EW when Fe II linewidth = 2000 km s^{-1}, while it is decreasing when Fe II linewidth = 4000 km s^{-1}. Appendix B.1 shows similar figures for the other input galaxy ratios.

It is as expected that the galaxy ratio is more accurately determined for increased input galaxy ratios, given that the host galaxy emission makes a stronger impression on the spectrum then. Similarly, it is expected that the galaxy ratio is more difficult to determine when the input galaxy ratio is lower. Comparing the appearance of the galaxy emission in the synthetic spectra in Figure 3.6 (galaxy ratio = 0.05) and Figure 3.7 (galaxy ratio = 0.75) shows that the host galaxy emission is clearly visible in the spectrum when the galaxy ratio is 0.75, while it is impossible to see the contribution when the galaxy ratio is 0.05.

Similarly, I would expect the scatter in estimated galaxy ratios to be larger for a galaxy ratio of 0.05 compared to a galaxy ratio of 0.75, simply because of the ambiguity of the host galaxy emission when the galaxy ratio is low. That is also what the results show. When the galaxy ratio is 0.75, the scatter is in the range of $[+0.001, +0.057]$ when considering all combinations of input parameter values. For comparison, when the galaxy ratio is 0.05, the scatter is in the range of $[+0.001, +0.144]$, which is up to 2.5 times larger than the scatter for a galaxy ratio of 0.75.

The amount of scatter will be affected by the amount of noise in the spectrum being considered. How the distributions of estimated galaxy ratios change with the S/N ratio can be seen in Figure 5.2a for a few combinations of input parameter values for a galaxy ratio of 0.50. The S/N ratios considered in this thesis are S/N = 5, 7, 10 and 50. The offset of the distributions can be seen to be of approximately the same size independent of the S/N ratio. The same tendency can be seen in the figures showing how the estimated galaxy



(a) Galaxy ratio **(b)** Power-law index α_ν

Figure 5.2: Examples of distributions of estimated values for the galaxy ratio (left) and power-law index α_ν (right) as a function of Fe II EW, Fe II linewidth and S/N ratio. Here galaxy ratio = 0.50 and $\alpha_\nu = 0.50$. The dashed line represents the true parameter value. The value of the S/N ratio does not influence the accuracy of the estimated galaxy ratio, but affects the precision. The spread is significantly larger for S/N = 5 than the other S/N ratios.
The shift in the accuracy of the estimated values remain more or less constant with increasing Fe II EW for Fe II linewidth of 2000 km s^{-1} (bottom), while it is changing with increasing Fe II EW for Fe II linewidth of 4000 km s^{-1} (top), and becoming more accurate relative to the true parameter value. The estimated galaxy ratio and α_ν are clearly seen to be degenerate when Fe II linewidth is 4000 km s^{-1}, as α_ν increases while the galaxy ratio decreases in a similar fashion for higher values of Fe II EW. The estimated galaxy ratios are offset by $[+0.007, +0.037]$ relative to the true value when considering all the different sets of parameters in the above figures (the galaxy ratios are overestimated by $[2, 6]$ %). The estimated values of α_ν are offset by $[-0.132, -0.068]$ (the value of α_ν is underestimated by $[-26, -14]$ %).

ratios change with the S/N ratio for other input galaxy ratios, shown in Appendix B.2. This means that *the accuracy of the estimated galaxy ratios does not depend on the S/N ratios.*

The precision (i.e., the amount of scatter) will be significantly different for the various distributions when considering the four S/N ratios I have used for the spectra in the catalogue. The amount of scatter will affect the ability to recover the true galaxy ratio in many cases. Figure 5.2a shows that the scatter in estimated galaxy ratios is significantly larger for S/N = 5 than the other S/N ratios. I added S/N = 7 to the parameter space later in the process in order to determine the limit where the precision starts changing between S/N = 5 and S/N = 10, as the amount of scatter is very different for these two S/N ratios. In Figure 5.2a the amount of scatter can be approximately 2–6 times larger for S/N = 5 compared to S/N = 10, depending on the combination of input parameter values. I find that S/N = 7 typically has an amount of scatter in the estimated galaxy ratios that are closer to that of S/N = 10 than that of S/N = 5. This indicates that there is much to gain in terms of precision by having spectra with S/N = 7 or higher, compared to S/N = 5.

Both Figure 5.1 and Figure 5.2a show that the true value of the galaxy ratio is within the uncertainties for some combinations of input parameter values. But the galaxy ratio is often too highly overestimated for the true value to be recovered for many of the parameter combinations, as the distributions are shifted to values well above the true galaxy ratio. Specifically, when the input galaxy ratio is 0.75, the accuracy of the estimated galaxy ratios is mostly within a few percent of the true value. The true value is within 1σ for S/N = 5, but can be as much as 6σ for S/N = 50 (Figure B.11). This is because a low S/N ratio will allow for a larger range of estimated galaxy ratios to be considered a good fit to the data, resulting in a potentially very large scatter of estimated galaxy ratios. This means that the true value is often recovered for S/N = 5 due to the large scatter. For S/N = 50 the precision tends to be very high (i.e., small amount of scatter), and the distribution of estimated galaxy ratios does often not include the true value at all. When the input galaxy ratio is 0.05, the galaxy ratio tends to be highly overestimated. At the most, the estimated galaxy ratios are offset by +0.080 relative to the true value, meaning the galaxy ratio is overestimated by 160 % (Figure B.6). The true value is within 2σ for S/N = 5, while it can be as much as 30σ for S/N = 50.

Dependence on other input parameters

Figure 5.1 showed that the extent to which the galaxy ratio is overestimated, is highly dependant on the combination of input parameter values being considered. Here I go through the effects of the different values of the power-law index α_ν and the iron emission on the estimated galaxy ratios.

Comparing the distributions of estimated galaxy ratios for the different input values of α_ν in Figure 5.1, shows that *the estimated values for the galaxy ratio tends to be most accurate when α_ν = –0.5 (steep in the UV).* This is perhaps because it is easier to separate galaxy emission from the continuum emission when their slopes are very different. The synthetic spectra in Figure 3.7 illustrates that the power-law emission makes a clearly visible contribution to the synthetic spectrum when α_ν = –0.5 (steep in the UV; left panel), compared to the synthetic spectrum when α_ν = 1.0 (flatter in the UV; right panel), where the components appear more blended.

Comparing the distributions of estimated galaxy ratios in the top and bottom panels in Figure 5.1 or in Figure 5.2a shows that the offset of the distributions of estimated galaxy ratios relative to the true value remain close to constant with increasing Fe II EW when the Fe II linewidth is 2000 km s^{-1} (narrow). When the Fe II linewidth is 4000 km s^{-1} (broad), the offset tends to *de*crease when increasing Fe II EW from 6 Å to 114 Å. This means that *the estimated galaxy ratio tends to be more accurately estimated when the iron emission is stronger and broader.* The fact that the galaxy ratio is more overestimated for smaller amounts of iron, is likely because the weak iron emission is difficult to identify and can be mistaken for host galaxy emission. This is supported by the fact that the value of Fe II EW tends to be underestimated when the input iron emission is weak, as I will discuss in Section 5.1.2. When the iron emission is strong relative to other spectral components, it is easier to identify the iron and therefore also estimate the galaxy ratio more accurately – as long as the iron emission is broad. When the iron emission is narrow, it resembles the features in the host galaxy spectrum, thus it will be difficult for the spectral decomposition program to estimate the galaxy ratio correctly even for increasing values of Fe II EW.

Figure 5.1 and Figure 5.2a also shows that the precision of the estimated galaxy ratios tends to *de*crease with higher values of Fe II EW. This loss of precision for higher values of Fe II EW coincides with an increase in accuracy for the estimated galaxy ratio for higher values of Fe II EW. The fact that the precision goes down when the accuracy goes up, and vice versa, is perhaps because the iron emission tends to be attributed to the galaxy ratio for small values of Fe II EW. While for large values of Fe II EW, some of the iron emission may sometimes attributed to the galaxy, while some of the galaxy is sometimes attributed to the iron, creating a spread in estimated values. This indicates a stronger degree of degeneracy between these two parameters when the iron emission is strong.

In Appendix C I show histograms of the distributions of estimated galaxy ratios for most combinations of input parameter values I am considering in this thesis. For the lowest S/N ratios (S/N = 5–7), the distribution of estimated galaxy ratios can sometimes become *bi*modal, rather than *uni*modal. This can for example be seen in Figure C.1d and Figure C.2c, and in many other histograms as well. This additional peak in the distribution is a clear indication that these parameters are difficult to determine accurately. This additional peak in the distribution can also be seen in the distribution of estimated values for the power-law index, as well as other fitting parameters (histograms of these distributions are not shown in thesis).

In conclusion, the overall trend I find is that it is easier to accurately extract the host galaxy emission when the iron emission is strong (Fe II EW $\gtrsim 57\,\text{Å}$) compared to when it is weak, when the iron lines are broad (Fe II linewidth $\gtrsim 4000\ \mathrm{km\,s^{-1}}$) compared to when it is narrow, and the slope of the continuum emission is steep in the UV ($\alpha_\nu = -0.5$) compared to when it is flat. The relative strength of the various spectral components is important, and there will therefore be combinations of parameter values that deviate from the trend stated here.

5.1.2 Supplementary results

How successfully the galaxy ratio is estimated, will depend on how well the other parameters describing the spectrum are estimated, given that they are all modelled simultaneously. It is therefore important to look at the results for the other parameters as well to better understand what is going on. I do not go into as much detail for the other components as I did with the galaxy ratio.

Estimated power-law index (continuum emission)

The power-law function describes the AGN continuum emission. It covers the entire spectral range and determines the overall shape of the synthetic AGN spectra. I would expect that the other components would have to compensate for the poorness of fit that can be introduced by an inaccurately estimated power-law function. It is therefore important to get the power-law continuum model right. The main results from the modelling are:

- *The power-law index α_ν is almost always underestimated.* This means that the power-law is estimated to be steeper in the UV than it actually is, which means that the

SMBH is estimated to have a higher mass accretion rate than it actually has. α_ν is more underestimated (i.e., less inaccurate) for input synthetic spectra that are steep ($\alpha_\nu = -0.5$) compared to spectra that are flatter ($\alpha_\nu = 1.0$). This can be seen by comparing the results for synthetic spectra with an input $\alpha_\nu = -0.5$ in Figure D.1–D.4 in Appendix D with the results for the synthetic spectra with $\alpha_\nu = 1.0$ in Figure D.5–D.8. For the synthetic spectra with a steep power-law, the estimated values of α_ν are underestimated by $[-22, -10]$ % relative to the true value, when considering all combinations of input parameter values. For synthetic spectra with a flatter power-law, the estimated values of α_ν are under- or overestimated by $[-13, 2.5]$ %, respectively, relative to the true value. While the power-law index tends to be more accurately determined for spectra that are flatter in the UV, the estimated galaxy ratio tends to be more accurately determined for spectra that are steep (this is often also the case for Fe II EW). It should also be noted that the value of the power-law normalisation constant also tends to be underestimated by the spectral decomposition program, which means that the entire power-law function tends to be shifted to lower flux values than it originally has in the input spectra. So even if the power-law index is accurately determined, the entire power-law may be shifted to lower values, so that the galaxy and iron contribution would have to make up for that shift in order to achieve a good fit.

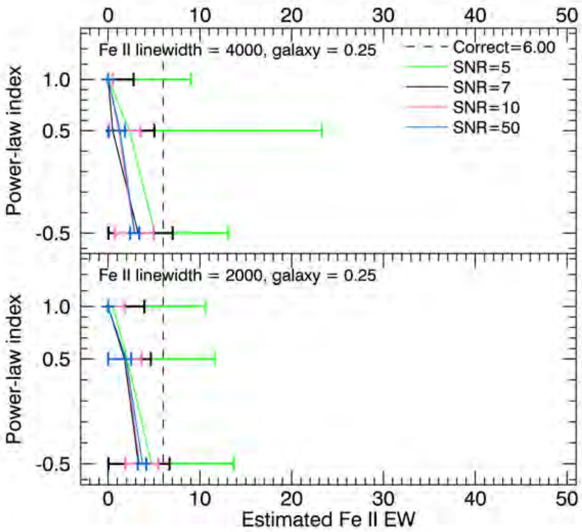
The accuracy of the estimated value of α_ν does not change significantly with input galaxy ratio for synthetic spectra that are steep. This can be seen in Figure D.1–D.4. Here the accuracy is within a range of $[-24, -10]$ % relative to the true value for all galaxy ratios. The accuracy tends to improve with increasing galaxy ratio for flatter spectra. In this case, the accuracy is within a range of $[-13, -8]$ % relative to the true value for an input galaxy ratio of 0.05 (Figure D.5), while it is within $[-6, 3]$ % for an input galaxy ratio of 0.75 (Figure D.8). This indicates that it is easier to estimate the value of α_ν for flat continuum spectra with a large amount of galaxy, which seems strange, considering what such a spectrum looks like, as seen in Figure 3.7 (right panels). It is hardly possible to see that there is a power-law present there.

- *The power-law index tends to be more accurately estimated for spectra with strong and broad iron emission.* When Fe II linewidth is 2000 $\mathrm{km\,s^{-1}}$ (narrow), the estimated values of α_ν do not change noticeably with increasing Fe II EW. When Fe II linewidth is 4000 $\mathrm{km\,s^{-1}}$ (broad), the estimated values of α_ν are *in*creasing (and becoming more accurate) with increasing Fe II EW. This can be seen by comparing the bottom and top panels in Figure 5.2b, respectively. The opposite was seen for the estimated galaxy ratio, for which the estimated values *de*creased with increasing Fe II EW. This indicates that the the galaxy ratio and power-law index are degenerate to some degree. Figures 5.2a–b illustrate how the estimated power-law index and the estimated galaxy ratio changes with the iron emission across the same part of the parameter space.

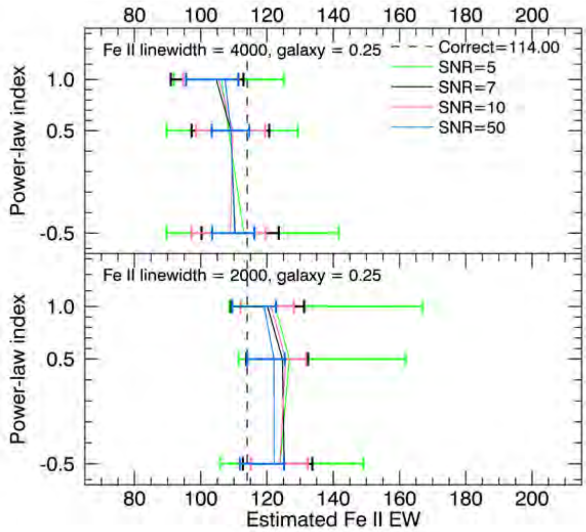
- *The true value of α_ν is sometimes within the uncertainties, but mostly the estimates are too far below the true value to be able to recover it.* As an example, for S/N ratio = 10 the estimated value can be up to $\sim 12\sigma$ away from the true value. The precision of the estimated values of α_ν tends to *de*crease with *in*creased Fe II EW, and more strongly for high S/N ratios. As discussed in Section 5.1.1, this indicates a larger degree of degeneracy for spectra with strong iron emission.

Estimated Fe II EW

The Fe II emission contributes to the synthetic spectrum in much of the same wavelength range as the galaxy. The results for the estimated galaxy ratio discussed above showed



(a) Fe II EW = 6 Å



(b) Fe II EW = 114 Å

Figure 5.3: Examples of distributions of estimated values for Fe II EW when the input value of Fe II EW is 6 Å (left) and 114 Å (right). Shown here as a function of α_ν for a galaxy ratio of 0.25 and for different S/N ratios. Note that the y-axis is showing the power-law index α_ν, and not Fe II EW as in the previous figures. *Left:* The value of Fe II EW is strongly underestimated, and often estimated to be zero. The true value of Fe II EW is always within the uncertainties for S/N = 5, and for S/N = 7 when the input spectrum is steep ($\alpha_\nu = -0.5$). This indicates that the noise is likely mistaken for iron in the fitting process. *Right:* The value of Fe II EW tends to be underestimated when the iron emission in the input spectrum is broad (top), while it tends to be overestimated when the iron emission is narrow (bottom) for the combinations of parameters shown here, indicating that it is easier to recover the narrow iron features in the spectra than the broad iron.

that the amount of Fe II emission – quantified by the Fe II EW – plays an important part in how accurately the galaxy ratio is estimated. The main results from the modelling are:

- *Fe* II *EW = 6 Å:* The iron content in the spectra is almost always *under*estimated, and more so for higher galaxy ratios, as can be seen in Figure 5.3a (and in Figures E.1–E.4 in Appendix E). It is almost impossible to recover the true value of Fe II EW when the input galaxy ratio is high – the estimated value of Fe II EW becomes zero for almost all these spectra. This is because the iron emission is so weak relative to the other components. The true value of Fe II EW is within the uncertainties in many cases when the S/N ratio is low (large amount of scatter), but otherwise not. The Fe II EW is most accurately estimated relative to the true value for the lowest S/N ratios, indicating that the decomposition program might be mistaking some of the noise for iron.

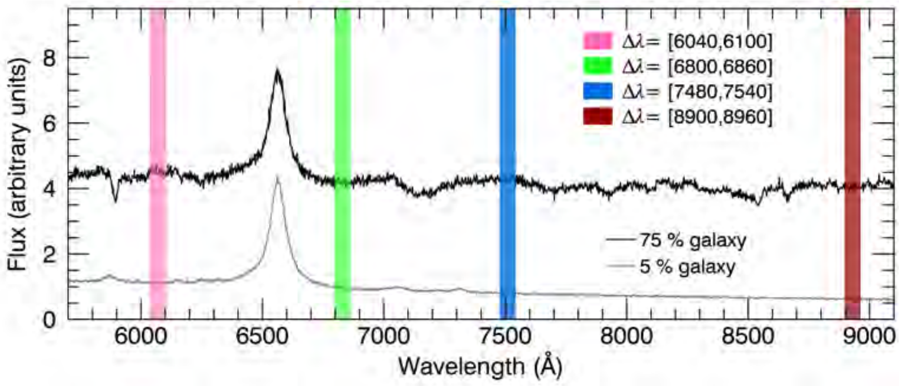
- *Fe* II *EW = 57 Å and 114 Å:* The iron content in the spectra tends to be *under*estimated for Fe II linewidth of 4000 km s^{-1} (broad). For Fe II linewidth of 2000 km s^{-1} (narrow) the iron content is *over*estimated when the galaxy ratio is low and *under*estimated when the galaxy ratio is high. An example of this can be seen by comparing the bottom panels in Figure E.6 and Figure E.8. The fact that the iron emission goes from being overestimated to underestimated with higher galaxy ratios indicates that the strength of the iron emission relative to the galaxy ratio is important for the results.

  Whether or not the true value of Fe II EW is within the uncertainties strongly depends on the combination of parameters being considered. See Figure E.5–E.12

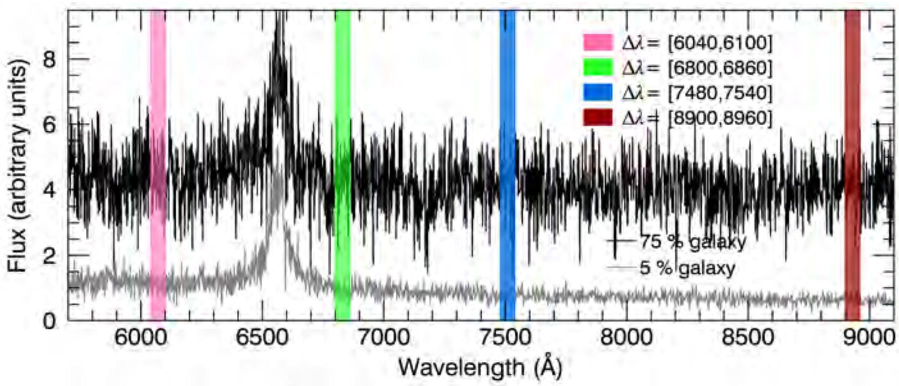
for the results for all the different input galaxy ratios. The precision is lower when the amount of iron is higher and improves with higher galaxy ratio, similar to what was seen for the estimated galaxy ratio and α_ν. The estimated values for Fe II EW are generally lower for flatter spectra (high values of α_ν), where the estimated galaxy ratio tends to be higher and more overestimated, indicating a degree of degeneracy between the two.

Estimated signal-to-noise ratio

The S/N ratio is estimated from the synthetic spectra directly, making it independent of the spectral fitting process. The S/N ratio is estimated by the simple relation $\mathrm{S/N} = \mu/\sigma$, where μ is the mean spectral flux and σ is the standard deviation, measured over a small wavelength interval. I have chosen to look at a wavelength interval of 60 Å. This interval is small enough for the shape of the spectrum not to change significantly within that range and to be able to avoid significantly strong spectral features (the galaxy contributes with many minor spectral features that must be avoided as they can be mistaken for noise). At the same time, the interval is large enough to capture many fluctuations in the noise. It would be interesting to see how a smaller or larger interval would influence the ability to measure the S/N ratios accurately and precisely in future studies.

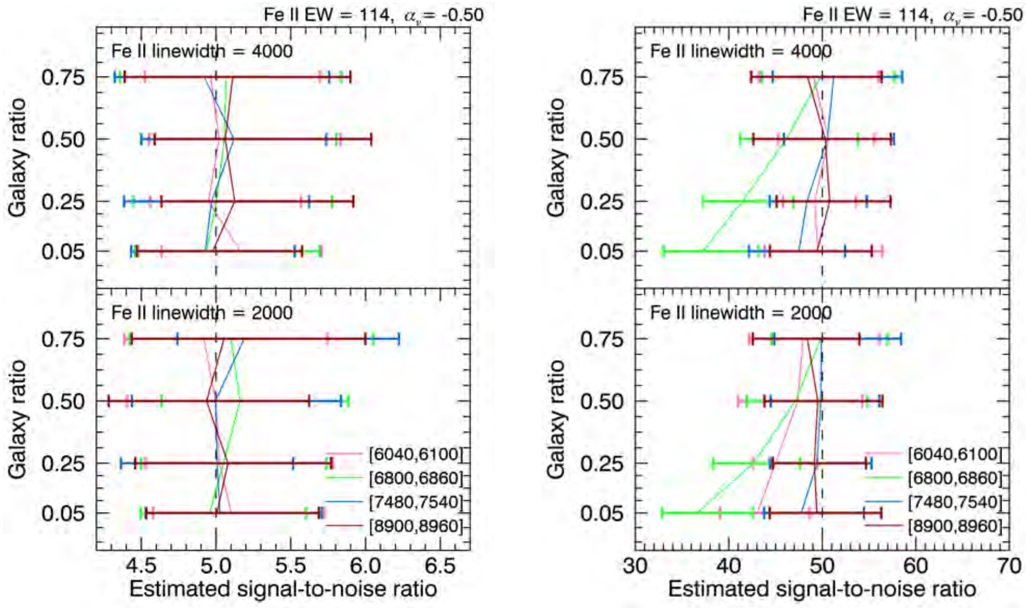


(a) $\mathrm{S/N} = 50$



(b) $\mathrm{S/N} = 5$

Figure 5.4: The S/N ratio was measured for four different wavelength ranges of the spectra where the flux is close to constant across the range. By comparing the black and grey curve for $\mathrm{S/N} = 50$ (top), representing a spectrum with 75 % and 5 % galaxy ratios, respectively, it is evident that the galaxy adds a lot of small-scale features to the spectrum. For $\mathrm{S/N} = 5$ (bottom) small spectral features will disappear in the noise. The pronounced peak in the spectrum is Hα, while the absorption features comes from the galaxy template. Here $\alpha_\nu = 0.5$, Fe II EW = 57 Å, Fe II linewidth = 2000 km s^{-1}.



(a) S/N=5 **(b)** S/N=50

Figure 5.5: These are some of the results from estimating the S/N ratio across four different ranges. When S/N = 5 (left) there tends not to be a difference between the accuracy of the estimated S/N ratio in the four ranges. This is because the noise will wash out any spectral features that could influence the noise measurements. For S/N = 50 (right) the green range – as well as the pink range when the iron is narrow – do noticeably poor for small galaxy ratios, likely due to the presence of weak spectral features. The red and blue range tends to do equally well in all cases, though the red range has slightly better precision. The distributions of estimated values for the S/N ratio for other combinations of parameters can be seen in Appendix F.

I estimate the S/N ratio over four different wavelength ranges where the flux appears to be more or less constant. The ranges are illustrated in Figure 5.4. I am specifically considering the ranges: $\Delta\lambda = [6040, 6100]$ Å(pink), $[6800, 6860]$ Å(green), $[7480, 7540]$ Å(blue) and $[8900, 8960]$ Å(red). By looking at how well the S/N ratio is estimated for the different wavelength ranges for different combinations of parameters it is possible to make recommendations for which wavelength ranges is preferable to use to estimate the S/N ratio of real AGN spectra.

When the S/N ratio is low (S/N < 50), it tends not to be a noticeable difference between how accurately the various wavelength ranges are reproducing the true value, see the results shown in Figure 5.5a. The strong appearance of the noise for low S/N ratios will effectively remove any signs of spectral features that could impact the results of the noise measurements.

It is only when the S/N ratio gets as high as S/N = 50 that there is an obvious difference in how the wavelength ranges are able to reproduce the true S/N ratio, see the results shown in Figure 5.5b. For S/N = 50 the green range does noticeably poorer for lower galaxy ratios, dramatically underestimating the S/N ratio. For galaxy ratios of 0.25 and below, the true value of the S/N ratio is not found within the uncertainties of the measurements. Specifically, the S/N ratio is underestimated by 2–3σ by the green range.

For high S/N ratios the spectral features are likely to make an impact on the noise measurements to some degree because the noise is so weak that it will not be able to 'hide' spectral features. This is an indication that the spectral flux is not as close to constant in the green range as I thought. By closer inspection of Figure 5.4a, the flux in the green range appears to be slightly higher for shorter wavelengths, due to its proximity to the pronounced Hα line. There also appears to be a bit more features for the shorter wavelengths in that range, though almost impossible to see. These factors might account for the poor result coming from the green range, though it is surprising that such weak features can impact the estimated S/N ratio so dramatically. This shows how crucial it is to choose the wavelength range wisely when estimating the S/N ratio for high-quality spectra.

The blue and red range tends to do best overall. These two ranges produces similar results when considering all the different combinations of parameter values, though the red range tends to have higher precision. The red range is therefore preferable. The red range is able to reproduce the true S/N ratio with an accuracy of $\pm 4\,\%$ with a precision that allows for estimates that are within $\pm 14\,\%$ for the most precise dataset and within $\pm 36\,\%$ for the least precise dataset, when considering the results from all combinations of parameter values in the catalogue.

The results for the estimated S/N ratio for other combinations of parameter values are shown in a number of figures in Appendix F.

5.2 Correlations

The results of the parameter estimations in the previous sections showed that there are strong hints of important associations between some of the parameters. The strength of these associations can be quantified by calculating correlation coefficients, which can provide a clearer picture of why the results are the way they are. There are no correlations between parameters inherently in the synthetic spectra I have created.

I use Spearman's correlation coefficient[1] r_s to quantify the association between the fitting parameters. r_s can take on values between –1 and +1, where the sign indicates whether one of the parameters tends to *in*crease when the another one *in*creases (positive association), or if one parameter tends to *in*crease when the other one is *de*creasing (negative association). If r_s is zero there is no association between parameters to speak of.

I compute the correlation matrix for all the AGN spectra in my catalogue in order to compare the estimated values for the parameters being fitted. The strength of the associations will depend on the spectrum at hand, but there are two parameters that are consistently highly correlated with the estimated galaxy ratio: The estimated power-law index α_ν and the estimated Fe II EW[2]. The main results from the correlation matrices are:

- *Galaxy ratio v. Power-law index:* The estimated galaxy ratio is highly *anti*-correlated with the estimated power-law index α_ν, typically in the range $-0.9 \gtrsim r_s \gtrsim -1.0$ (significant at a 1 % level). How the correlation coefficient depends on other parameters is shown in Figure 5.6a for S/N = 10. The results for the other S/N ratios are shown in Appendix G.1. The fact that these parameters are anti-correlated means that the galaxy ratio tends to be estimated to be high when α_ν is estimated to be low, or vice versa. This relationship was already visible in the results discussed previously where the galaxy ratio was almost always *over*estimated while the value of α_ν was almost always *under*estimated.

  The strength of the anti-correlation changes with the input parameters of the spectra in such a varied manner that it is difficult to identify clear trends across the parameter

[1] The Spearman's rank correlation coefficient r_s is suitable in this case as it describes how well the relationship between two parameters can be described by a *monotonic* function – a function that is steadily increasing or decreasing over a given range – rather than a *linear* function, the latter being the basis for the more commonly used Pearson's correlation coefficient. This is because I am interested in seeing if certain parameters tends to be overestimated while other are underestimated, and vice versa, rather than looking for linear relationships.

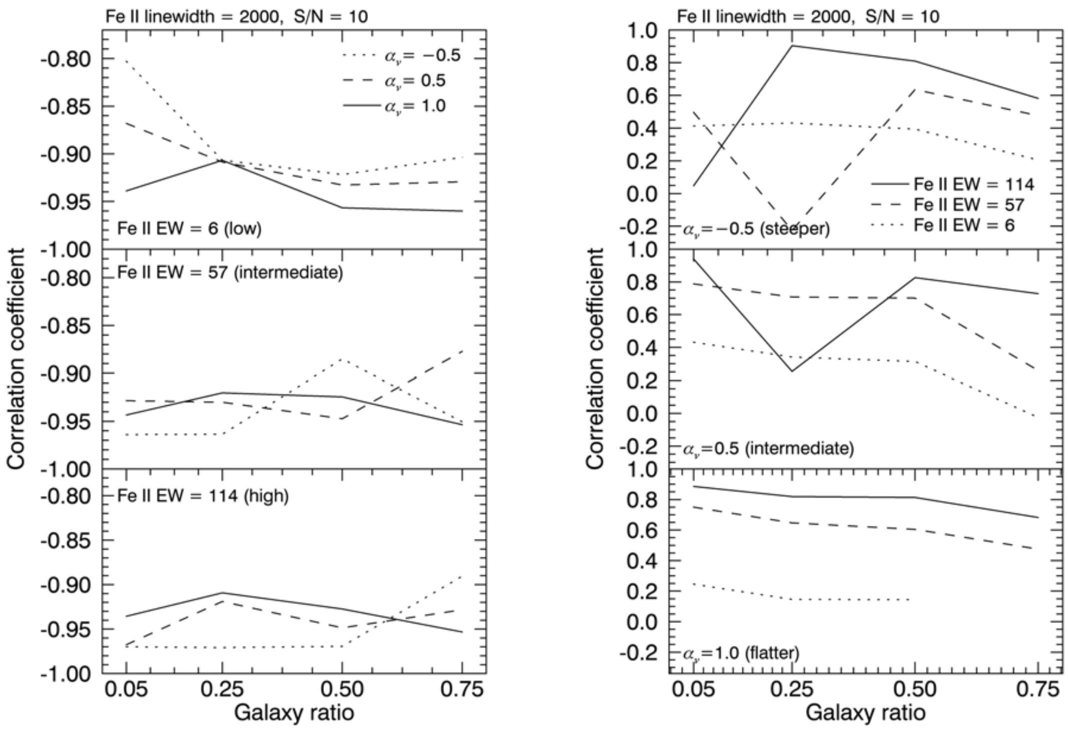
[2] The estimated power-law index α_ν is in turn correlated with the estimated power-law constant, and the estimated Fe II EW is in turn correlated with the estimated Fe II scaling flux constant, which are also correlated with the galaxy ratio. But the correlation between the estimated galaxy ratio and the estimated α_ν and Fe II EW are generally the strongest.

space. But even with some variations, the anti-correlation is always stronger than $r_s \lesssim -0.8$. See the figure captions in Appendix G.1 for comments on the individual cases.

- *Galaxy ratio v. Fe* II *EW:* The estimated Fe II EW is (often highly) correlated with the estimated galaxy ratio. How the correlation coefficient depends on other parameters is shown in Figure 5.6b for S/N = 10. The results for the other S/N ratios are shown in Appendix G.2. The fact that these parameters are correlated means that the galaxy ratio tends to be estimated to have high values when the Fe II EW is also estimated to be high, or be low when Fe II EW is also estimated to be low.

  The strength of the correlation tends to be stronger when the amount of iron in the spectra is higher, and is often $r_s > 0.8$ (mostly significant at a 1 % level). However, the strength for a particular combination of input parameter values strongly depends on the S/N ratio and the other input parameters. The correlation is often weak ($r_s \lesssim 0.6$) when Fe II EW = 6 Å. See the figure captions in Appendix G.2 for comments on the individual cases.

  The value of the correlation coefficient fluctuates strongly for different combinations of input parameter values when S/N = 10, while the coefficient is more consistent across the parameter space for the other S/N ratios. Figures showing the results for the other S/N ratios, as seen in the figures in Appendix G.2. The fluctuations is perhaps because the noise for S/N = 10 is at a level where the decomposition program



(a) Galaxy ratio v. power-law index α_ν **(b)** Galaxy ratio v. Fe II EW

Figure 5.6: The change in the correlation coefficient r_s with galaxy ratio, Fe II EW and α_ν for S/N = 10. The correlation coefficient for other combinations of input parameters can be seen in Appendix G. *Left:* The estimated galaxy ratio is anti-correlated with the estimated power-law index α_ν, typically in the range $-0.9 \gtrsim r_s \gtrsim -1.0$ (significant at a 1 % level). *Right:* The estimated galaxy ratio is correlated with the estimated Fe II EW. The strength of the correlation strongly depends on the combination of input parameters. The fluctuations in the strength of the correlation for a galaxy ratio of 0.25, indicates that the relative strength of the galaxy ratio and the iron emission is important.

gets confused about which spectral component is responsible for certain features, particularly mixing the galaxy, iron and the noise. Thus the components will not be estimated in a similar manner for every spectral realisation. E.g., sometimes the galaxy might be estimated to be high while the iron emission is estimated to be low, while other times both might be estimated to be high, and so on. This variation in how the parameters are estimated from one spectral realisation to another will weaken the strength of the correlation.

5.3 Stability of the parameter estimations

Having looked at the above results, two questions spring to mind: 1) Will the spectral decomposition program always find the *same* parameter values for a given realisation of a spectrum? And, 2) does it matter what *order* I put the fitting parameters in?

I investigated the first question by modelling the *same* realisation of a spectrum over and over again 150 times (identical noise each time; Case A) for a few different sets of parameters. I then compared the results with the original distributions shown in Appendix C where I modelled 150 *realisations* for the various combinations of parameters (the noise is randomly generated from a Gaussian distribution and will be different for each realisation; Case B).

The Levenberg-Marquardt algorithm used in the decomposition program finds *local* minima in the parameter space rather than a *global* minimum, and there can be many local minima. I would therefore expect there to be some scatter in the estimated galaxy ratios, even when modelling identical spectra. But it is hard to guess how large the scatter would be before actually doing a test.

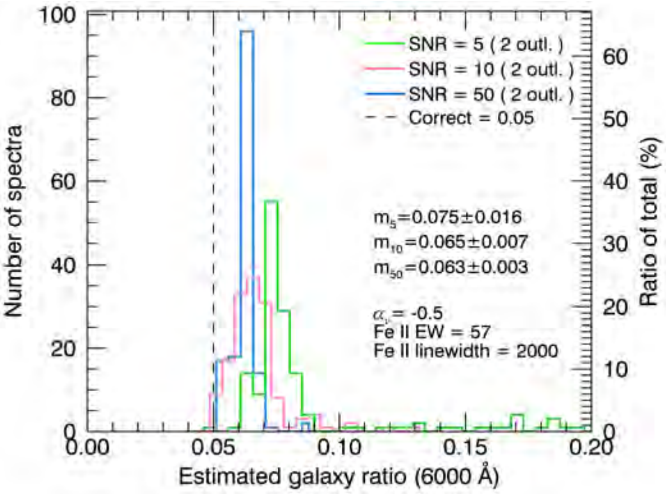
I chose to remodel four spectra created with some extremal sets of parameter values: low and high input galaxy ratio, steep and flat input spectrum. The result for Case A are shown in the left panels in Figure 5.7 (galaxy ratio = 0.05) and Figure 5.8 (galaxy ratio = 0.50), and compared with Case B in the right panels. I originally considered S/N = 5 and 50. S/N = 10 was added later to better constrain how the results change with S/N ratio. The estimated values for the galaxy ratio for S/N = 10 tends to follow the behaviour of S/N = 50 more closely than S/N = 5.

The main results from the remodelling are:

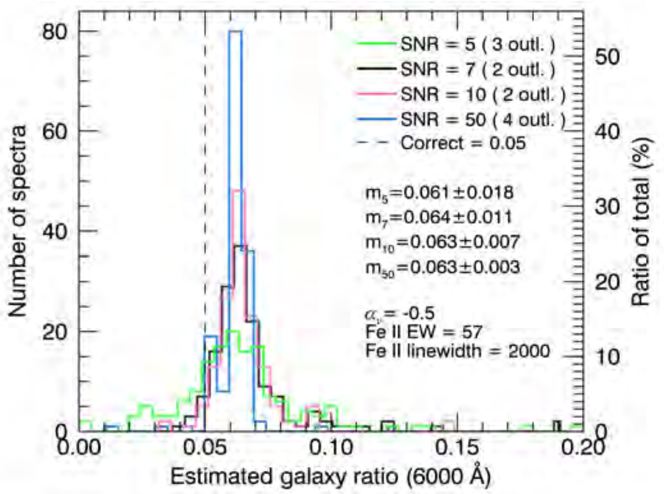
- *The fitting process itself introduces a scatter in the estimated parameter values.* This is in accordance with my expectations described above. This result can be seen by the fact that there is a scatter in the estimated parameter values for Case A. Specifically, for S/N = 5 there is a scatter in the estimated parameter values of $\pm\,[0.003, 0.016]$ ($\pm\,[0.6, 48]\,\%$ relative to the true galaxy ratio) for the combinations of input parameters considered here. For S/N = 50 there is a scatter of $\pm\,[0.003, 0.008]$ ($\pm\,[0.6, 16]\,\%$). This means that some of the scatter in the final results presented in this thesis (represented by Case B) are introduced by the fitting process itself (represented by Case A).

- *When the S/N ratio is low, the accuracy and precision for Case A and Case B can be very different.* For a the high input galaxy ratio of 0.050, the accuracy in Case B is only different from the accuracy in Case A with a few percent, when comparing the two cases in Figure 5.8. For the low input galaxy ratio of 0.05, the difference in offset relative to the true value between Case A and Case B can be as large as $[-0.065, +0.014]$ (19–95 %), as seen in Figure 5.7.

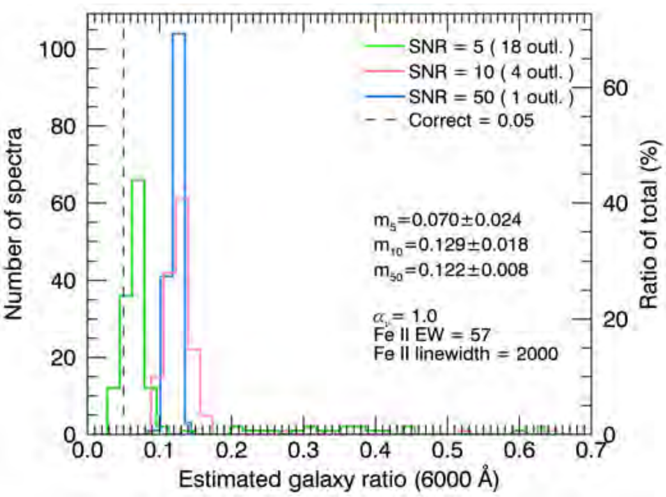
  The difference in results for a galaxy ratio of 0.05 and the lack thereof for a galaxy ratio of 0.50 is consistent with the fact that higher galaxy ratios are more accurately determined than lower galaxy ratios, as discussed in Section 5.1.1. The difference in the offset for the low galaxy ratios means that the exact noise distribution in the



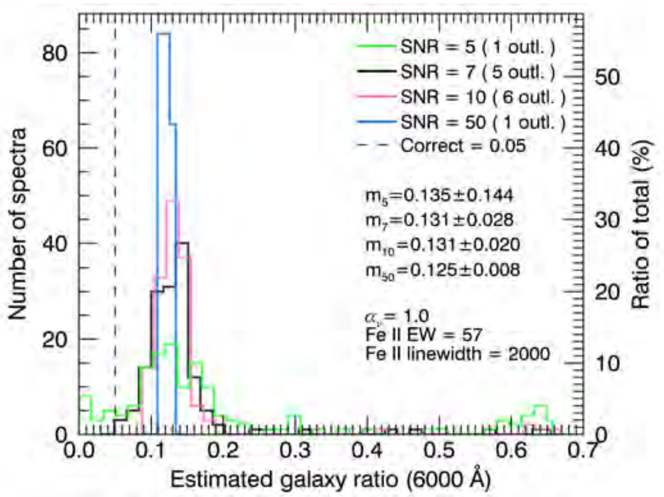
(a) Case A ($\alpha_\nu = -0.5$)



(b) Case B ($\alpha_\nu = -0.5$)



(c) Case A ($\alpha_\nu = 1.0$)

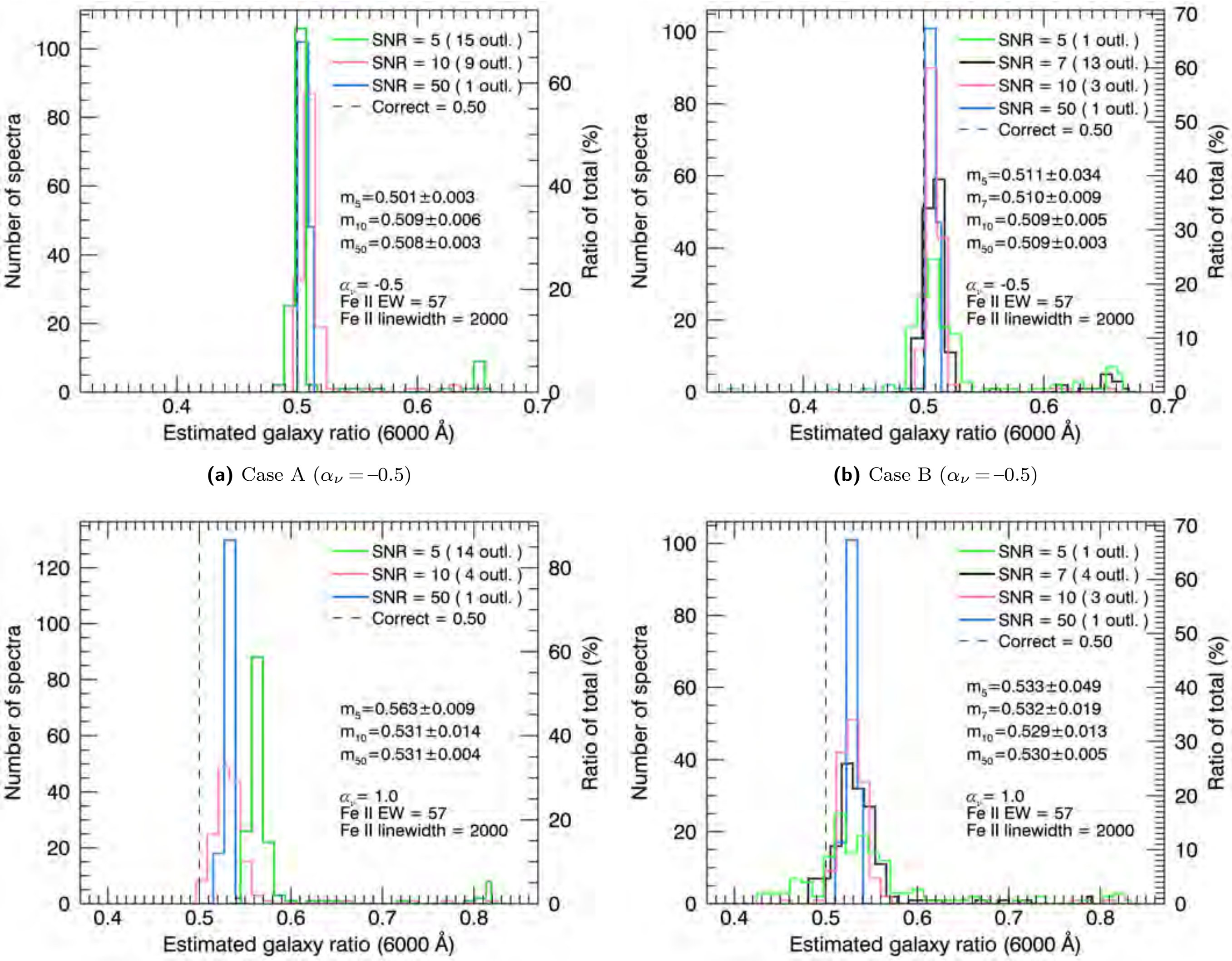


(d) Case B ($\alpha_\nu = 1.0$)

Figure 5.7: Here galaxy ratio = 0.05. Outliers are not included. *Left:* The result from fitting two spectra repeatedly 150 times each (Case A). *Right:* The original distributions when fitting 150 realisations of a spectrum, where each realisation has slightly different randomly generated Gaussian noise (Case B).

Top: $\alpha_\nu = -0.5$. When S/N = 50, the distribution in Case A is practically identical to the distribution in Case B. This indicates that the accuracy and precision is set by the decomposition program. When S/N = 5, the distributions in Case A and Case B have a similar spread, but the accuracy is very different – it is much poorer in Case A (left). This indicates that for low S/N ratios, the actual noise array of the spectrum is important for the accuracy of the result.

Bottom: $\alpha_\nu = 1.0$. The distributions are different in Case A and Case B. Both the accuracy and precision is better in Case A (left). The difference in accuracy and precision between the two cases is very large when S/N = 5. For S/N = 50 the precision is the same, and the difference in accuracy is very small.



(a) Case A ($\alpha_\nu = -0.5$)

(b) Case B ($\alpha_\nu = -0.5$)

(c) Case A ($\alpha_\nu = 1.0$)

(d) Case B ($\alpha_\nu = 1.0$)

Figure 5.8: Here galaxy ratio = 0.50. Outliers are not included. *Left:* The result from fitting two spectra repeatedly 150 times each (Case A). *Right:* The original distributions when fitting 150 realisations of a spectrum, where each realisation has slightly different randomly generated Gaussian noise (Case B).

Top: $\alpha_\nu = -0.5$. The distributions in Case A and Case B is practically identical when S/N = 50. This indicates that the accuracy and precision is set by the decomposition program. When S/N = 5 the result is more accurate and much more precise in Case A (left). Again, this indicates that the noise in the spectra is important for the accuracy of the result.

Bottom: $\alpha_\nu = 1.0$. The distributions in Case A and Case B is practically identical when S/N = 50. Again, this indicates that the accuracy and precision is set by the decomposition program. When S/N = 5 the result is less accurate in Case A (left), but has a much higher precision compared to Case B (right). Again, this indicates that the noise in the spectra is important for the accuracy of the result.

spectrum actually is important when estimating the galaxy ratio from spectra with a low S/N ratio. In practice, it means that when using the spectral decomposition method presented in this thesis on real AGN spectra with low S/N ratio, one can be lucky and have a spectrum with a noise distribution that allows for a rather accurate estimate for the galaxy ratio. Or one can be unlucky and have a spectrum with a noise distribution that will make the estimated galaxy ratio highly inaccurate. This makes the result for a low S/N ratio hard to trust.

The noise variations in the 150 realisations in Case B adds additional scatter to the estimated values for the galaxy ratio when S/N ratio is low ($S/N < 10$), compared to the scatter introduced by the fitting program in Case A. For $S/N = 5$ the increase in scatter for Case B is about 100–1100 % relative to the scatter in Case A, depending on the combination of parameters being considered. It is as expected that there is additional scatter for Case B, given that the noise distribution varies for each realisation.

- *When the S/N ratio is high, the accuracy and precision for Case A and Case B are practically identical.* This result indicates that the accuracy and precision of the estimated galaxy ratio tends to not be much affected by the noise when the S/N ratio is high. This is because there is so little noise in the spectra. In practice, this means that the precision and accuracy of the estimated galaxy ratio is determined and limited by the decomposition program itself in this particular case. This conclusion does not change with the input galaxy ratio, as can be seen by comparing Case A and Case B in both Figure 5.7 and Figure 5.8.

From the results mentioned here, it is evident that the Levenberg-Marquardt algorithm is not doing well at estimating the galaxy ratio in many cases when the S/N ratio is low. The biggest concern for spectra with a low S/N ratio is that the achieved accuracy appears to be strongly dependent on the specific noise distribution of the spectrum. This effect is especially important when the input galaxy ratio is low.

Now for the second question. Because I want to understand how the fitting process works and how it influences the results, I did a quick test to see if the *order* of the fitting parameters matters. The order of the fitting parameters that was used to produce the results I have presented in this chapter is stated in Chapter 4, with the power-law first on the list and the galaxy template last.

I tried a few different configurations of orders for the parameters. For each order I modelled 150 spectral realisations for one particular set of parameter values[3] in order to compare the results from the different configurations. The configurations I tested was: 1) power-law first, galaxy second, 2) galaxy first, power-law second, and 3) galaxy first, power-law last.

The achieved accuracy and precision for the estimated galaxy ratio was identical to the original configuration when testing the different configurations of parameter orders. The only exception was the third configuration where the spread was about 9 times larger than in the other cases. This simple test shows that it likely only matters where in the order the power-law is. It is good to know that there is such good agreement between the different orders of parameters. If the results had not been in agreement, there would be a need for trying all possible orders of the parameters to find the order that best reproduce the true parameter values. It is however strange that the spread of estimated galaxy ratios would be different with the power-law last. This does indicate that the order of the parameters is important to some degree. It could be, given that the power-law describes the overall shape of the spectrum, that saving the power-law for last makes it slightly more difficult

[3]Galaxy ratio = 0.05, S/N = 50, $\alpha_\nu = 1.0$, Fe II EW = 57 Å, Fe II linewidth=2000 $\mathrm{km\,s^{-1}}$

for the decomposition program to estimate the different components, but in a way that it only affects the precision and not the accuracy.

5.4 Explaining the results

I am not merely interested in reporting the results, I also hope to understand why the results are the way they are. The tests I have performed in addition to the original parameter estimations was done to better understand how the fitting algorithm interprets the spectra. The situation is however complex and detangling the various effects on the different parameters and determine their causes is not straightforward because of the dimensionality of the parameter space. Also, it is not enough to consider what the parameter values are. One must also consider the difference in the relative strength of various components when going through the parameter space, as well as differentiate between the input parameter values from the estimated parameter values of the fitted spectra when trying to assess how one parameter influences another. Not to mention all the possible ways of comparing the results and parameters with each other.

The results from the spectral fitting process showed that the galaxy ratio tends to be *over*estimated. It is strongly anti-correlated with the power-law index α_ν, which in turn tends to be *under*estimated. It would be ideal to be able to explain why that is, but it does not seem possible to draw conclusions on *why* the galaxy ratio tends to be overestimated and the power-law index tends to be underestimated – and not the other way around or in another way – by looking at the results. The strength of the anti-correlation between the two parameters says that these two components are in fact degenerate, meaning that they are compensating for each other in order to achieve a good fit. But it appears to not be a way of telling which component is dictating the other or if there is any dictating behaviour in the fitting process at all, given that it does not make a difference for the accuracy of the estimated parameter values what the order of the fitting parameters is.

I would expect the power-law to be very accurately determined when the input galaxy ratio is low and the input spectrum is steep. In that case the synthetic spectrum does not deviate much from the shape of a pure power-law function, as can be seen in Figure 3.6 (left panels). However, this is not what happens. The decomposition program is in fact doing better at estimating the power-law index for flat spectra with a large contribution from the galaxy, perhaps because the various components are more clearly visible when the spectrum is flatter. This indicates that it might be beneficial to fit the power-law before fitting in a different way. I discuss some options in Section 6.4.1.

It could be of interest to record the starting parameters that were responsible for the final fit for every spectrum in order to see if there is a pattern in which starting parameters are able to achieve a fit. If this happens to be the case, the values of the starting parameters could be responsible for pushing the estimated values of the various spectral components in certain directions. Then it would be possible to get a better understanding of why the results are the way they are. However, if this were to be the case, it is not a problem that one could actually fix.

6 Discussion

The goal of this thesis has been to test how well the spectral decomposition software code I have written, described in Chapter 4, can estimate the host galaxy contribution to AGN spectra. The decomposition program is based on the Levenberg-Marquardt algorithm (Moré 1978), a non-linear least-squares curve fitting algorithm. I have tested the decomposition program on a catalogue of 43,200 synthetic AGN spectra that I have created 'from scratch' using a variety of templates that represent known spectral contributions, as described in Chapter 3. This catalogue covers a range of observed spectral properties of AGNs – various continuum emission, iron emission and host galaxy emission – quantified by a number of parameters that can take on different values, as listed in Table 3.2, with a total of 288 possible combinations of parameter values. By varying the values of these parameters, I am effectively simulating different types of AGNs. For each combination of parameter values, I make 150 spectral realisations with randomly generated Gaussian noise, for four different S/N ratios ranging from 5 to 50, to check how the noise will affects the ability to recover the host galaxy emission.

In this chapter I will discuss the implications of the results I presented in Chapter 5 and discuss possible improvements for how to estimate the host galaxy contribution to AGN spectra.

When discussing the result of this thesis, it is important to remember that I have explored a simple model for AGN spectra. A real observed AGN spectrum will have more and/or different features and complexities for the spectral components than the ones that are described my the templates used in my model, such as narrow emission lines, dust extinction, Doppler shifts, line asymmetries, etc. This means that the accuracies and precisions that are presented in this thesis can only be regarded as *best-case scenarios*. Thus the result presented here really sets the theoretical limit of the accuracy and precision that can be achieved using this type of decomposition method. When using the method on real spectra, the accuracy and precision is therefore likely to be worse.

6.1 Recovering the true host galaxy contribution

How successfully the spectral decomposition method used in this thesis can be said to estimate the host galaxy contribution, depends on the accuracy and precision one is requiring. The results discussed in Section 5.1.1 showed that host galaxy contribution is almost always *over*estimated. This means that the host galaxy emission is found to account for a larger part of the spectral emission that it actually does. When the input galaxy ratio is high ($\gtrsim 0.75$; measured at 6000 Å), the galaxy ratio will be estimated with high accuracy with an offset of ± 0.005, meaning that it is under- or overestimated by $< 3\,\%$ relative to the true value. The precision is also high, in that the scatter is small: it is in the range of $[+0.001, +0.057]$ when considering all combinations of parameters. The host galaxy contribution is relatively more overestimated compared to the true value the weaker the galaxy component of the input spectra is. When the input galaxy ratio is as low as 0.05, the estimated galaxy ratios can be offset by up to $+0.080$ relative to the true value, meaning it is overestimated by as much as $160\,\%$. The fact that the galaxy ratio can be overestimated by this amount when the input galaxy ratio is close to zero means that the decomposition program is likely to find a galaxy contribution even in spectra where there might not be any. The scatter (or uncertainty) in estimated galaxy ratios can be large for an input galaxy ratio of 0.05, in the range of $[+0.001, +0.144]$ (up to 2.5 times larger than for a galaxy ratio of 0.75).

The accuracy to which the galaxy ratio is estimated will depend on the input values of the other spectral components. When I study how the different parameter values affect the ability to recover the host galaxy contribution, I find that it is easier to accurately extract the host galaxy emission when the iron emission is strong (Fe II EW $\gtrsim 57\,\text{Å}$) compared to when it is weak, when the iron lines are broad (Fe II linewidth $\gtrsim 4000\ \mathrm{km\,s^{-1}}$) compared to when it is narrow, and the slope of the continuum emission is steep in the UV ($\alpha_\nu = -0.5$) compared to when it is flat. I discuss the impact of the S/N ratio on the accuracy and the precision below.

6.1.1 The impact of the signal-to-noise ratio

In my catalogue of synthetic AGN spectra I have made spectra with S/N ratios of 5, 7, 10 and 50 in order to check what quality of spectra is needed to acquire a desired accuracy and precision for the host galaxy emission.

I found that increasing the value of the S/N ratio from 5 to 50 (i.e., increasing the quality of the spectra) did not affect the *accuracy* of the estimated galaxy ratios. For a given combination of parameters, the offset of the distribution of estimated galaxy ratios relative to the true value is practically independent of the S/N ratio (there can be minor differences in the offset when comparing the result for the various S/N ratios in some cases). This could be seen in Figure 5.2a for a galaxy ratio of 0.50 (the result for the other galaxy ratios are shown in Appendix B.2).

The S/N ratio does however significantly affect the *precision* of the measurements, in that increasing the S/N ratio from 5 to 50 will lead to higher precision (i.e., less scatter). For example, when the input galaxy ratio is 0.75, the scatter in the estimated galaxy ratio will be in the range of $[+0.017, +0.057]$ when $\mathrm{S/N} = 5$, while it will be in the range of $[+0.001, +0.008]$ when $\mathrm{S/N} = 50$.

Because the scatter is so different when $\mathrm{S/N} = 5$ compared to $\mathrm{S/N} = 50$, this means that if I was to make *one* measurement of the galaxy ratio of a spectrum, the probability of recovering the true value for the galaxy ratio would be very different for the different S/N ratios. In terms of recovering the true galaxy ratio, it is also important to keep in mind that the estimated galaxy ratio tends to be shifted away from the true value by a significant amount. This means that because the scatter increases with decreasing S/N ratio, there is actually a larger probability of recovering the input galaxy ratio for lower S/N ratios. Specifically, for $\mathrm{S/N} = 5$ the true value is always within 0–2σ. The histograms in Appendix C shows that there is typically a 5–15 % probability of recovering the true galaxy ratio for $\mathrm{S/N} = 5$. In comparison, for $\mathrm{S/N} = 50$ there is almost always a 0 % probability of recovering the true value because the scatter in estimated values are smaller than the offset of the distribution. For $\mathrm{S/N} = 50$ the true galaxy ratio is within 0–30σ, and almost always more than 1σ.

For $\mathrm{S/N} = 5$ the scatter in values for the estimated galaxy ratio is often so large that it will include the input galaxy ratio, even if the distribution of estimated galaxy ratio tends to be shifted to higher values. But the scatter also extends into the opposite direction of the input galaxy ratio, resulting in the possibility of estimating the galaxy ratio to be much higher than the input galaxy ratio. This means that even though I have a higher probability of recovering the true value with $\mathrm{S/N} = 5$ compared to $\mathrm{S/N} = 50$, I will also have a higher probability of estimating the galaxy ratio to be more inaccurate, compared to $\mathrm{S/N} = 50$. To take an extreme: In the case of galaxy ratio $= 0.05$, $\alpha_\nu = 1.0$, Fe II EW $=$

57 Å, Fe II linewidth $= 4000$ km s^{-1} (Figure C.2d), I have approximately a 10 % probability of recovering the true value for S/N = 5. But I also have a 3 % probability of recovering a value that is 1200 % larger than the true galaxy ratio. In the same situation for S/N = 50 there is zero probability of recovering the true value because the distribution of estimated galaxy ratios is offset from the true value. There is a 65 % probability of estimating the galaxy ratio to be 100 % larger than the true value, and the galaxy ratio will never be estimated to be larger than approximately 200 % in the particular case I am considering here.

The amount of scatter in a distribution of estimated galaxy ratios will be affected by the following interrelated factors:

1. the amount of noise in the spectrum (quantified by the S/N ratio),
2. the variation in noise distribution from one spectral realisation to another, and
3. the limitations of the fitting algorithm.

The last point is highly important, as it can significantly influence both the accuracy and the precision of the estimated galaxy ratios. Because several of the spectral parameters can be degenerate, the decomposition program will not necessarily identify the same value for the estimated galaxy ratio each time I model a spectrum, even if I am in fact modelling the exact same spectral realisation multiple times (i.e., the noise distribution is identical each time).

As I discussed in Section 5.3, for low S/N ratios ($\lesssim 10$) the accuracy that is possible to achieve will be highly dependent on the specific noise distribution of a spectrum. This could be seen by comparing the distributions for Case A (modelling the same spectral realisation 150 times) with Case B (modelling 150 different realisations of a spectrum). Here Case A represents the limitations of the fitting algorithm. The fact that the distributions of estimated galaxy ratios tended to be significantly different in the two cases for low S/N ratios shows the importance of the noise distribution for what the estimated galaxy ratios becomes. The specific noise distribution can make it impossible to recover the true value, which was seen by comparing the distributions for S/N = 5 in Figures 5.8c–d. Or the specific noise distribution can make it possible to recover the true value (with a 25 % probability for the combination of parameters considered here), which was seen by comparing the distributions for S/N = 5 in Figures 5.7c–d. This results in an unquantifiable uncertainty in the estimated parameter values when using this decomposition method to study a single spectrum with a low S/N ratio. This fact must be considered in addition to the above discussion about using low S/N ratios. When S/N = 50, the accuracy and the precision of the estimated galaxy ratios is not significantly affected by the noise, but it is mainly the result of the fitting algorithm finding slightly different combinations of parameter values each time due to degeneracies between spectral components.

To summarise, the estimated galaxy ratios are inaccurate (relative to the true value) with high precision for S/N = 50, while the estimated values tend to be accurate with low precision for S/N = 5. Ultimately the question becomes: Is it better to be wrong with a high degree of certainty, or right with a low degree of certainty? As discussed previously, this will be an issue that one has to consider for small to intermediate galaxy ratios ($\lesssim 0.50$) and for low S/N ratios ($\lesssim 10$).

6.2 Effect on SMBH mass measurements

The main motivation for finding a method for successfully determining the host galaxy contribution in AGN spectra, is to be able to use the R–L relation (discussed in Section 2.4.2) to easily estimate the mass of supermassive black holes (SMBHs) from single-epoch spectra.

Denney et al. (2009; hereafter referenced as D09) studied the systematic uncertainties associated with determining the SMBH mass from single-epoch spectra by using the many epochs available for the Seyfert 1 galaxy NGC 5548 and the quasar PG1229+204. One of the issues they examined is what happens to the accuracy and precision of the estimated SMBH masses when the host galaxy contribution is subtracted from the AGN spectra, compared to when it is not. They measured the host galaxy contribution by using 2D-image decomposition (following the approach of Bentz et al. 2009), as well as using two different spectral decomposition methods. Their spectral decomposition method A is very similar to my approach, in that they used similar spectral components and the same fitting algorithm as I have done, though it is unclear how many fitting parameters they have used to do the modelling.

D09 used the result of the 2D-image decomposition to subtract the host galaxy component from the two AGN spectra rather than the results from the spectral decomposition. They do however use the spectral decomposition method as a consistency check. They find that their method A and the image decomposition method gives highly consistent results for the estimated host galaxy contribution, with an average relative difference in the flux of the host galaxy contribution at 5100 Å of –6.5 % when comparing the two methods, though with approximately twice the scatter for the spectral decomposition method compared to the image decomposition. This consistency check was only done for the Seyfert galaxy NGC 5548, where the host galaxy contribution is likely to be stronger than for the quasar PG1229+204.

D09 then estimated the SMBH mass using Equation 2.2 from the multi-epoch spectra for the two AGNs, before and after they had estimated and subtracted the host galaxy contribution from the AGN spectra. The resulting SMBH masses was then compared to SMBH masses measured from reverberation mapping (RM), in order to see how the host galaxy contribution would affect the SMBH mass estimates. When D09 compares their results with RM measurements they find that failing to subtract the host galaxy contribution from the AGN spectra leads to a shift of the entire SMBH mass distribution towards higher values (i.e., the SMBH masses are being *over*estimated). This is because the luminosity will be larger when the host galaxy contribution has not been subtracted, compared to when it has. The SMBH mass is proportional to the luminosity, as can be seen in Equation 2.2. When the galaxy *has* been subtracted, they find that there is an average *under*estimation of SMBH masses compared to the RM masses. In this scenario too much luminosity has been subtracted from the AGN spectra. This indicates that the host galaxy contribution was not accurately determined. Specifically, it indicates that the galaxy contribution is *over*estimated, consistent with my findings.

Based on their figure showing the result of the spectral decomposition for NGC 5548 (Figure 2 in D09), it looks like the galaxy ratio is estimated to be around 20–30 % at 6000 Å, the power-law index appears to be close to $\alpha_\nu = 0.5$ and the iron emission is very weak. In my studies, AGN spectra with an input galaxy ratio of 25 %, $\alpha_\nu = 0.5$ and very weak iron emission have an accuracy for the estimated galaxy ratio that can be overestimated by as much as 52 % relative to the true value (see Figure B.2). This means that the degree of overestimation for the host galaxy contribution of D09 can be 52 % or more. Their input galaxy ratio is likely to be lower than their estimated galaxy ratio, which will result in an even larger relative error than 52 %. It is therefore to be expected that subtracting the host galaxy contribution will result in an underestimation of the SMBH mass.

With the goal of estimating SMBH masses from spectra, I check how large the effect of subtracting host galaxy components is on the accuracy and precision of SMBH masses for different galaxy ratios. Then it is possible to determine for which galaxy ratios it starts to become important to accurately estimate and subtract the galaxy contribution to achieve a certain accuracy and precision for the SMBH masses.

In order to evaluate to which extent the host galaxy contribution affect the SMBH mass

Table 6.1: The effect on SMBH mass estimates from the host galaxy contribution

Input galaxy ratio	**Host galaxy contribution** (1)	**Δlog M** (2)	**Diff (present–removed)** (3)
0.05	Present	+0.017	
	Removed lowest estimate (0.048)	+0.001	+0.016
	Removed highest estimate (0.127)	–0.026	+0.043
0.25	Present	+0.096	
	Removed lowest estimate (0.248)	+0.001	+0.095
	Removed highest estimate (0.307)	–0.023	+0.119
0.50	Present	+0.249	
	Removed lowest estimate (0.497)	+0.002	+0.247
	Removed highest estimate (0.533)	–0.019	+0.268
0.75	Present	+0.545	
	Removed lowest estimate (0.729)	+0.022	+0.523
	Removed highest estimate (0.755)	–0.011	+0.556

estimates, I make some SMBH mass estimations with my own results using Equation 2.2. I calculate the SMBH mass offset (i.e., the additional mass resulting from the host galaxy contribution being present in the spectrum) by

$$\Delta\log M = 0.519 \cdot (\log L_{5100,\,\mathrm{spec}} - \log L_{5100,\,\mathrm{galaxy-free\,spec}}). \tag{6.1}$$

Here I calculate the difference in SMBH mass estimates measured from the full spectrum (represented by $L_{\mathrm{spec},\,5100}$) and the spectrum where the host galaxy contribution has been subtracted (represented by $L_{\mathrm{galaxy-free\,spec},\,5100}$). This allows me to quantify how much 'extra' SMBH mass is added by having the host galaxy contribution present when doing the mass measurement. To arrive at Equation 6.1 I have assumed that the linewidth of Hβ in Equation 2.2 is the same when measured on the spectrum with the host galaxy contribution present and the spectrum where the host galaxy contribution has been removed.

Because I consider the difference between two spectra in Equation 6.1, I may use the flux at 5100 Å in my spectra directly instead of the luminosity L_{5100}, as constant terms will cancel each other out. I measure the flux from the my spectra by taking the mean flux over 5100 Å $\pm$ 20 Å. I do this measurement for each of input galaxy ratios considered in this thesis, for one particular combination of the other input parameters[1], as an example of how the estimated galaxy ratios can affect the SMBH mass estimates.

The results of the SMBH mass estimates are shown in Table 6.1. There I list the estimated SMBH mass offset for three different cases for each of input galaxy ratios considered in this thesis: *Top row:* the host galaxy contribution was present in the spectrum when calculating the SMBH mass offset; *middle row:* the lowest median value of the estimated galaxy ratios in my catalogue was removed from the spectrum before calculating the SMBH mass offset; and *bottom row:* the highest median value of the estimated galaxy ratios was removed from the spectrum before calculating the SMBH mass offset. Column 2 in Table 6.1 shows the SMBH mass offset in these different cases. A value of zero in the 'Removed' cases would mean that the host galaxy ratio is accurately estimated and subtracted from the spectrum, and the SMBH mass can be accurately estimated. The SMBH mass offsets first of all show that having the galaxy contribution present in the spectrum (top row for each galaxy ratio) will lead to an increasingly overestimated SMBH mass

[1] S/N = 50, $\alpha_\nu = 0.5$, Fe II EW = 57 Å, Fe II linewidth = 4000 km s^{-1}

when increasing the galaxy ratio from 0.05 to 0.75. The SMBH mass offset goes from being +0.017 to +0.545 in these cases. This increase is as expected because a larger host galaxy contribution will lead to a larger luminosity at 5100 Å.

What happens when the host galaxy contribution is removed from the spectrum, based on how large I have estimated this contribution to be? The values in the 'Removed' cases in Column 2 reveals that for galaxy ratios that are 0.25 or higher, the absolute size of the SMBH mass offsets will be *smaller* when the galaxy contribution *has* been removed compared to when the host galaxy contribution is present. This means that there is a significant probability of obtaining a more accurate SMBH mass estimate when removing the host galaxy contribution from spectra where the input galaxy ratio is 0.25 and higher, compared to having it present. For an input galaxy ratio of 0.05, removing the lowest estimated galaxy contribution results in a highly accurate SMBH mass estimate. But when removing the highest estimated galaxy contribution results in a negative offset in the SMBH mass that is larger than the original positive offset from having the host galaxy contribution present in the spectra. This indicates that it might be better to leave the host galaxy contribution in the spectrum when estimating SMBH masses from AGN spectra where the host galaxy contribution is close to zero.

D09 sees an opposite effect than I do, in that their SMBH masses are typically as accurate or more accurate relative to RM mass measurements (i.e., the absolute value of $\Delta\log M$ is smaller) when the host galaxy contribution has *not* been subtracted. For this to be possible, their host galaxy contribution must be highly overestimated. This discrepancy might arise from the fact that they are considering real data while I am using synthetic data.

Column 3 shows the difference between having the host galaxy contribution being present in the spectrum and the host galaxy contribution being removed. These values indicate when the host galaxy emission starts to make a large impact on the SMBH mass measurements, and thus when the host galaxy contribution needs to be removed. Already at a galaxy ratio of 0.25 does the host galaxy contribution start to be important. At what point it becomes important, will of course depend on the accuracy needed.

I have not taken into account the scatter in the distributions of the estimated galaxy ratios in this analysis. If I were to include the scatter, the lowest and highest estimates listed in Column 1 would be even lower and higher, respectively, than the ones I have used. This would in turn make the mass difference in Column 3 larger. But I think it is more informative to do this simple analysis based on the median values of the distributions, as these values have the highest probability of being measured.

There is also a scatter in the R–L relation itself that will add to the uncertainties of the rough estimates I have made here for the SMBH mass offsets. In Equation 6.1 I have a factor of 0.519, but the uncertainties in the R–L relation allows this factor to be in the range of 0.45–0.59 (Bentz et al. 2009). How this affects the SMBH mass estimates will depend on the galaxy ratio being considered, and I do not go into this here.

6.3 Which types of AGNs can be studied with the spectral decomposition method?

In spectra where the host galaxy emission is relatively low, the AGN will be dominating the spectrum, as is the case with quasars. In spectra where the host galaxy emission is relatively high, the AGN will appear weaker, as is the case with Seyfert galaxies. The question becomes: For which types of AGNs can the host galaxy contribution be accurately determined using the spectral decomposition method studied in this thesis in order to get an accurate SMBH mass estimate? And for which types of AGN is it better to leave the host galaxy contribution in the spectrum without trying to remove it?

There are two possible issues:

1. Removing an overestimated host galaxy contribution from the total spectrum will lead to an underestimation of the AGN continuum flux as too much flux will be subtracted. This will in turn lead to an underestimated SMBH mass.

2. Leaving the host galaxy contribution in the spectrum would lead to an overestimation of the AGN continuum flux, and more so for stronger host galaxy contributions, as there will be more flux attributed to the AGN continuum flux than there should be. This will in turn lead to an overestimated SMBH mass.

In this thesis I have found that the host galaxy contribution is relatively more accurately and precisely determined when the host galaxy makes an increasingly large contribution to the spectra, going from a minimum galaxy ratio of 0.05 to a maximum galaxy ratio of 0.75. Because the host galaxy contribution can actually be estimated quite accurately and precisely when it is strong, as discussed in Section 6.1, the SMBH masses will therefore be accurately estimated when modelling and subtracting a visually strong host galaxy component. Large host galaxy contributions would cause a large shift in SMBH masses if not subtracted. This will be most relevant for low-luminosity Seyfert galaxies where the host galaxy contribution to the spectrum is strong relative to the AGN contribution. An AGN spectrum with a strong host galaxy contribution would be spectra where it is possibly to identify the Ca II break around 4000 Å. Figure 3.7 shows examples of synthetic spectra with the strongest host galaxy contribution in my catalogue for reference.

In quasars, the AGN is very bright compared to the surrounding host galaxy, and the effect of not subtracting the host galaxy will have a smaller impact on the SMBH mass estimates. If the estimated galaxy contribution in quasars is subtracted, it could lead to an underestimation of the SMBH mass for quasars, considering that the weak host galaxy contribution in quasars is likely to be highly overestimated. I would be cautious to use this decomposition method to estimate the galaxy ratio in spectra where it is unclear to what extent a galaxy contribution is present in the spectra one wishes to model, due to the large uncertainties associated with estimating weak host galaxy contributions. But again, this depends on the degree of uncertainties one is willing to accept.

There is a redshift aspect to this as well. The surface brightness of galaxies dims rapidly with increasing redshift z, specifically as $(1+z)^{-4}$. This does not affect the AGN because it is a point source rather than an extended object. This means that for quasars in the non-local Universe, the host galaxy contribution can be neglected when studying AGN properties. Shen et al. (2011) consider this to be the case for *SDSS* quasars at $z \gtrsim 0.5$. Because we do not necessarily know how large the host galaxy contribution is to quasar spectra in the local Universe, it is difficult to make a definite statement about how to treat the host galaxy contribution in these local cases.

6.4 Possible improvements to spectral host galaxy estimates

The estimated galaxy ratios in my study were often shifted to higher values relative to the true value, and the host galaxy emission tended to be degenerate with the continuum emission and the iron emission. This indicates that there is a need for an improved (or different) spectral decomposition method that can account for the spectral components more accurately. Below I go through three possible improvements and/or changes to the spectral decomposition method and discuss their possible impact on the ability to accurately estimate the host galaxy emission. I go through them in the order of increasing difficulty of implementation.

6.4.1 Continuum fitting

Because the power-law continuum covers the entire spectral range of my synthetic spectra, how well the continuum emission is modelled is likely to impact how accurately the other parameters are able to be modelled. The results showed that the power-law index α_ν was almost always underestimated (by as much as –24 % in the worst case) and that the estimated value of α_ν was highly anti-correlated with the estimated galaxy ratio ($-0.9 \gtrsim r_s \gtrsim -1.0$), indicating the need for a better way to accurately determine the value of α_ν, in hopes of more accurately estimate the galaxy ratio.

In my spectral decomposition program I fitted the power-law to the spectrum in a very simple manner, in that I fitted a power-law function to the entire spectrum. There are two other possible approaches:

1) Fit the power-law and subtract it from the spectrum before fitting the other spectral components. I am currently fitting all the spectral components simultaneously. Fitting the power-law first would be especially relevant for spectra that are steep in the UV ($\alpha_\nu = -0.5$) and where the other spectral components are weak, as illustrated in the left panel of Figure 3.6. In this case the power-law makes a prominent contribution to the spectra. However, it seems highly likely that the power-law index α_ν will be estimated to be steeper than it actually is as soon as there is a significant contribution from the host galaxy emission to the spectrum, as illustrated in the left panel of Figure 3.7. This would be consistent with the results I am already getting for α_ν. I also expect that fitting the power-law separately would be difficult for spectra that are flatter in the UV, as the power-law emission is almost impossible to identify visually, which is illustrated in the right panels of Figure 3.6 and Figure 3.7. I therefore doubt that this is a useful route to take.

2) Fit the power-law over a smaller wavelength range. Wills et al. (1985) proposed that the continuum could be measured in the wavelength ranges of 5400–6200 Å, 1430–1460 Å, and 1330–1380 Å, which are areas of AGN spectra where the contribution from Fe II and other emission lines is small. However, the two latter wavelength ranges will be very close to the broad Si IV emission line in my spectra, while the first wavelength range will be in a part of the spectrum where the power-law emission rarely is the dominating component due to host galaxy emission. By considering some of the synthetic AGN spectra I have made where there is a strong contribution from the host galaxy emission, illustrated in Figure 3.7, there is not a particular wavelength range that stands out as a suitable range to measure the power-law slope in across the parameter space I am considering.

Based on this discussion, it appears that the approach I am currently using for estimated the shape of the power-law emission might be the best one when considering such a large range of spectral properties as I am doing here.

6.4.2 Changing the fitting algorithm

I have used the IDL procedure `MPFITFUN` (Markwardt 2009) to model the synthetic AGN spectra in my thesis, which is a non-linear least-squares curve fitting algorithm based on the Levenberg-Marquardt algorithm (Moré 1978). This algorithm works by changing the values of the fitting parameters in order to reduce the value of χ^2 (the method is described in Section 4.1).

This approach is limited by the fact that χ^2 might have several possible minimum values, so-called *local* minima, in addition to the true global minimum value. This is due to degeneracies between the parameters being fitted. The *global* minimum is the minimum value of χ^2 that corresponds to having reached parameter values identical to the true parameter values that the algorithm are trying to recover. The fitting algorithm can risk being 'stuck' in a local minimum in the parameter space that is far away from the global minimum that it cannot get out of, which can result in a poorly fitted model.

In Chapter 5 it became clear that there are degeneracies between the fitting parameters.

It would therefore be beneficial to explore other fitting algorithms.

An alternative fitting algorithm to the one used in this thesis is the Markov chain Monte Carlo method (MCMC; Metropolis et al. 1953; Hastings 1970), which have started to become widely used in physics. There exists many varieties of MCMC methods. I will not go into details about the inner workings of the MCMC method as this is beyond the scope of this thesis.

MCMC methods are particularly useful for multi-dimensional problems. The main advantage is the ability to explore the parameter space more quickly than other fitting methods. The argument for this type of method is also that it gives a complete description of the uncertainties and degeneracies of the fitting parameters (e.g., Serra et al. 2011). The degeneracy of the fitting parameters are studied by looking at the probability distribution functions of the various fitting parameters. `MPFITFUN` can also give information about the uncertainties of the fitting parameters due to the noise in the spectrum (not studied in this thesis), but not about the degeneracies. I have however learned about the degeneracies in other ways, by for example studying correlations and looking at the histograms for the various distributions, but perhaps not in the same amount of detail that MCMC might provide.

The reason that I started to look into this alternative fitting algorithm was because I have gotten the impression that it is a superior method for sampling a multi-dimensional parameter space. I have however not been able to find studies that look at the difference in accuracy and precision of MCMC methods compared to non-linear least-squares curve fitting methods when fitting a large number of parameters to back up my initial impression. It is therefore difficult to quantify to what extent switching to the MCMC method can improve on the ability to accurately recover the host galaxy emission to the AGN spectrum. Because of this, and due to the complexity of the implementation of the method, it is not clear to me if this would be a route that is worthwhile to take.

6.4.3 Expanding the spectral range

How well the spectral decomposition method can recover the emission from the various spectral contributions is likely to depend on the wavelength range being considered, as different spectral features will be visible in different wavelength ranges. In my thesis I have considered a wavelength range consisting of the optical part of the AGN spectrum, including some ultraviolet (UV) and near-infrared (near-IR) emission, specifically 1000–9940 Å.

Optical emission has the benefit of being observable from the ground, which makes it much easier to obtain observational data compared to having to rely on space-based telescope. But considering that my spectral decomposition program struggled to determine the host galaxy contribution in many cases, larger parts of the full spectral energy distribution (SED) of AGNs might be needed to better constrain the host galaxy component.

Old, cool stars emit most of their light in the red part of the optical and near-infrared wavelengths because of their temperatures, with most of the light being emitted around the wavelength of $1\,\mu$m (10,000 Å). Young bright, hot stars emit radiation in the UV and the blue part of the optical, but the radiation is re-emitted in the far-infrared at wavelengths larger than $10\,\mu$m (100,000 Å) due to surrounding dust in their galaxy (Sparke & Gallagher 2007). These differences between the two stellar populations can be seen in the number of galactic SEDs shown in Figure 6.1. The emission from the elliptical template (dominated by old stars) quickly falls off after $1\,\mu$m, while the emission from the spiral galaxy templates (a mixture of young and old stars) peaks after $10\,\mu$m. Because the range of possible stellar temperatures is limited, the host galaxy will not be contributing significantly in the UV, which could be seen in the galaxy templates in Figure 3.5a. The AGN will therefore be dominating in the UV, and expanding the wavelength coverage

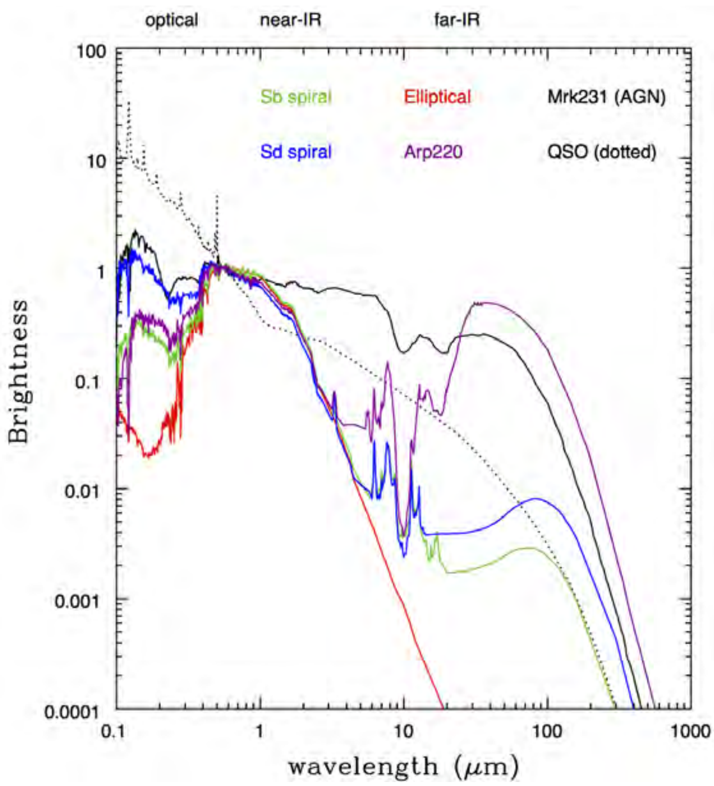


Figure 6.1: Template spectra of different types of galaxies: two spiral galaxies (green and blue), elliptical galaxy (red), starburst galaxy Arp 220 (purple), Seyfert 1 galaxy Mrk 231, and a QSO (black, dotted). In Ångstroms, the spectral range shown here is 1,000–10,000,000 Å. My spectra covers approximately the 0.1–1 μm range in this figure. The templates shown here are from Polletta et al. (2007), the figure is from the blog of the CANDELS Collaboration. The y-axis shows the flux as νF_ν in arbitrary units.

into the UV will not be helpful in trying to constrain the host galaxy contribution more successfully.

The SEDs of a Seyfert galaxy and a quasar are also shown in Figure 6.1. Because the host galaxy emission makes a stronger contribution to the spectrum of Seyfert galaxies, the Seyfert galaxy SED can be seen to have a similar shape to that of the galaxy SEDs, but at a higher brightness level. For quasars, where the host galaxy contribution tends to be weak, the SED is noticeably different from the galaxy SEDs. There is in general little overlap between the different SEDs in Figure 6.1, except for around 1 μm. This suggests that it would be a good idea to expand the spectral range into the IR to better constrain the host galaxy contribution.

The downside of considering a wider spectral range this is that observations with several instruments would be needed in order to obtain observations for the full desired spectral range. This might not be easily obtained due to limited amounts of observation time available, especially if the goal is to map SMBH masses for a large number of AGNs at different cosmological distances to get a better understanding of AGNs. The upcoming *James Webb Space Telescope (JWST)* will be operating in the 0.6–28.5 μm range, spanning orange to mid-infrared wavelengths. The *Spitzer Space Telescope* is already working in the infrared covering the 3.6–160 μm range. These two telescopes in combination with an optical telescope, would cover the wavelengths of interest in Figure 6.1 – if this entire range is needed. The question then becomes how large the spectral range actually needs to be in order to accurately extract the host galaxy contribution. This could be probed by doing spectral decomposition on synthetic AGN SEDs for different subsets of the wavelength range to check if there are certain spectral features that are more important for the spectral decomposition.

Ciesla et al. (2015) has done a study of how the AGN contribution to AGN SEDs can be constrained (rather than the host galaxy contribution – but if one of them is obtained, the other one ultimately follows) by modelling the full SED to learn more about the host galaxy in terms of star formation rates, stellar masses, etc. They find that it is difficult to constrain the fraction of the AGN contribution to the total infrared luminosity when the fraction is $< 20\,\%$, while the uncertainties are $\sim$5–30 % for higher fractions depending on the AGN type. The AGN fraction can be significantly *over*estimated in the case of weak AGN. They also explore the effect of including different wavelength ranges in terms of being able to determine various stellar properties. The wavelength range needed depends on stellar features they wish to recover. A wavelength range from UV to far-IR appears to be needed in order to constrain all the different parameters in their model. They do not comment on any degeneracies between the host galaxy and the AGN.

Another example of AGN SED fitting is the program `AGNfitter` (Calistro Riviera et al. 2016) that does a full SED fitting of AGN spectra using a Bayesian MCMC approach. They find a degeneracy between the host galaxy and the AGN contribution around $1\,\mu$m, which is where all the spectra are seen to overlap in Figure 6.1. They find this degeneracy to be more pronounced for low-luminosity AGNs. Because the exact determination of the host galaxy emission is not a focus of their study, but rather the topic of classifying AGNs, they do not discuss the importance of this degeneracy. They do however point out that degeneracies are bound to occur due to the number of fitting parameters being considered, and that the strength of the degeneracies will increase as the number of fitting parameters increases.

The two studies mentioned here have tried full SED fitting of AGN spectra. Because neither of them focus on determining the total host galaxy contribution in the way that I do in this thesis, there is not enough information to compare their approach with mine, or to make any definite conclusions on whether expanding the spectral range is worthwhile or not to better constrain the host galaxy contribution. This might therefore be a route that is worth exploring further in future studies.

7 Conclusion

The aim of this thesis has been to characterise how accurately and precisely the host galaxy contribution to optical AGN spectra can be determined using a spectral decomposition method. This type of method serves as an alternative to the current method which relies on high-resolution imaging with the *Hubble Space Telescope*, which will soon no longer be available.

In order to characterise the ability of the spectral decomposition method to determine the host galaxy contribution I have created a catalogue of 43,200 synthetic AGN spectra with various combinations of different amounts of continuum emission, iron emission and host galaxy emission, effectively simulating different types of AGNs. The spectra were also given various amounts of noise, so that I could study the effect of noise on the results.

The spectral decomposition software code I have written for this thesis is based on the Levenberg-Marquardt algorithm, implemented by the IDL package `MPFITFUN`. The spectral decomposition code closely follows the recipe that I used to create the synthetic spectra, thus I am in practice exploring the theoretical limit of the accuracy and precision that can be achieved with this method.

The results from the spectral decomposition indicate that the host galaxy emission may be accurately estimated when it is a relatively strong contributor to the spectrum. The galaxy ratio (measured at 6000 Å) is accurate to within 3 % of the true value when the input galaxy ratio is 0.75, which is the highest galaxy ratio in my catalogue. It becomes increasingly difficult to accurately determine the host galaxy emission as this contribution becomes relatively weaker, resulting in a significant *over*estimation of the host galaxy emission for most combinations of the spectral variations.

The main motivation for finding a method to accurately determine the host galaxy contribution to optical AGN spectra is to be able to use the $R_{\rm BLR}$–L relation to estimate SMBH masses from single-epoch spectra. When using the $R_{\rm BLR}$–L relation for this purpose, it is important that the spectral contribution from the host galaxy emission is either accurately determined and subtracted from the spectrum, or so small that it is negligible. Because the host galaxy emission tends to be highly overestimated by the spectral decomposition method when the host galaxy emission is relatively weak, it is best to neglect this contribution in the weakest cases rather than try to remove it. It will however be difficult to judge when the host galaxy contribution is sufficiently low for this to apply. When the host galaxy emission makes a relatively strong contribution to the spectrum, it is estimated with high accuracy and precision, and should therefore be modelled and removed from the AGN spectrum. In practice, this means that the spectral decomposition method studied in this thesis is mostly applicable for the study of low-luminosity Seyfert galaxies where the host galaxy makes a relatively large contribution to the spectrum. In non-local quasars the host galaxy emission can be neglected because it will be very faint. The problem is all the AGNs that fall in-between these two extremes. In these cases it ultimately becomes a question of what accuracy and precision is needed.

In this thesis I have considered synthetic AGN spectra with S/N ratios of 5, 7, 10 and 50. Increasing the S/N ratio from 5 to 50 does not affect the accuracy of the estimated galaxy ratios, but it improves the precision significantly. Due to the uncertainties being significantly larger and the degeneracies being stronger for low S/N ratios, I recommend only using the spectral decomposition method presented in this thesis for spectra with S/N $\gtrsim 10$ pixel^{-1}.

A possible improvement to the spectral decomposition method presented here, is expanding the spectral range into the infrared, where the host galaxy emission is relatively strong. This is however not trivial, as it would rely on data from several space-based telescopes, whereas the spectral decomposition method studied here only relies on wavelengths that can be observed with ground-based telescopes.

A List of software codes

Here I list the software codes I have written for my thesis with short descriptions. They are written in IDL.

- `spectrum_wrapper.pro`: A wrapper program that calls on the functions to create mock (synthetic) spectra, model the spectra and store information and plots.
- `filename_generator.pro`: Generates filenames and folder structures based on the values of the input parameters so that all the spectra are named and stored according to their input parameters.
- `plot_spectrum_new.pro`: Plots the synthetic spectrum (and the model spectrum when called upon during the modelling) with all its spectral components, and saves the plots to file.
- `mock_spectrum.pro`: Calls on all the different templates and functions to create the synthetic spectra by adding the components together. Writes a text file with the wavelength and flux array.
  - `powerlaw_continuum.pro`: Sets up the power-law continuum function.
  - `fe_emission.pro`: Code and template from supervisor to create the iron spectrum.
  - `balmer_continuum.pro`: Code and template from supervisor to create the Balmer spectrum and lines.
  - `equivalent_width.pro`: Calculates the equivalent width of a spectral line. Used to scale the iron and Balmer emission to the continuum.
  - `emission_lines.pro`: Sets up the broad emission lines I have chosen to include in the spectrum. Uses another function I have written, `lineflux.pro`, to make the individual lines and put them all together.
  - `hostgalaxy_spec.pro`: Retrieves a host galaxy spectrum from a directory of templates.
  - `noise.pro`: Creates a noise distribution with a given signal-to-noise ratio (S/N) for the spectrum.
- `spec_modeling.pro`: Models a given AGN spectrum using `MPFITFUN`. Saves the final fitting parameter values to file.
  - `set_parinfo.pro`: Picks the starting parameters for all the fitting parameters randomly from a list of three options for each fitting parameter.
  - `model_func.pro`: Model function to be fitted to an AGN spectrum. Calls on the same functions as listed under `mock_spectrum.pro:`.

In addition to the software codes listed here, I have also written several programs for plotting and statistical analysis.

B Figures: Estimated galaxy ratio

Here the results for the estimated galaxy ratio from the spectral decomposition are shown graphically for a range of different combinations of parameter values. The galaxy ratio is defined as the fraction of host galaxy emission in AGN spectra at 6000 Å. The plotted graphs show the median and an interpercentile that covers 34.1 % of the measurements on each side of the median (see Section 5.1 for details). For a detailed look at each distribution of estimated galaxy ratios, see the histograms presented in Appendix C.

B.1 Estimated galaxy ratio v. Fe II EW & Power-law index

The figures below show how the estimated value for the galaxy ratio changes with power-law index α_ν and Fe II EW for galaxy ratio = 0.05, 0.25, 0.50 and 0.75. The results show the estimated galaxy ratio as a function of Fe II EW for different α_ν for S/N = 10. For a look at how the results change with S/N ratio, see Section B.2. In the following figures the top panels show Fe II linewidth = 4000 km s^{-1}, and the bottom panels 2000 km s^{-1}.

A brief summary of the results: The galaxy ratio is almost always *over*estimated. There is a noticeable difference in how well the galaxy ratio is estimated for the three power-law indices α_ν. $\alpha_\nu = -0.5$ tends to provide the most accurately estimated galaxy ratios relative to the true value. The precision decreases with increasing α_ν and increasing Fe II EW.

Another common trend is that the values of the estimated galaxy ratios do not tend to change significantly with changes in Fe II EW when Fe II linewidth = 2000 km s^{-1} (narrow). But when Fe II linewidth = 4000 km s^{-1} (broad), there is a clear decrease in the value of the estimated galaxy ratio with increased Fe II EW. The estimated galaxy ratio is closest to the true value when there is more iron in the spectrum.

The figures begin on the following page.

Galaxy ratio = 0.05

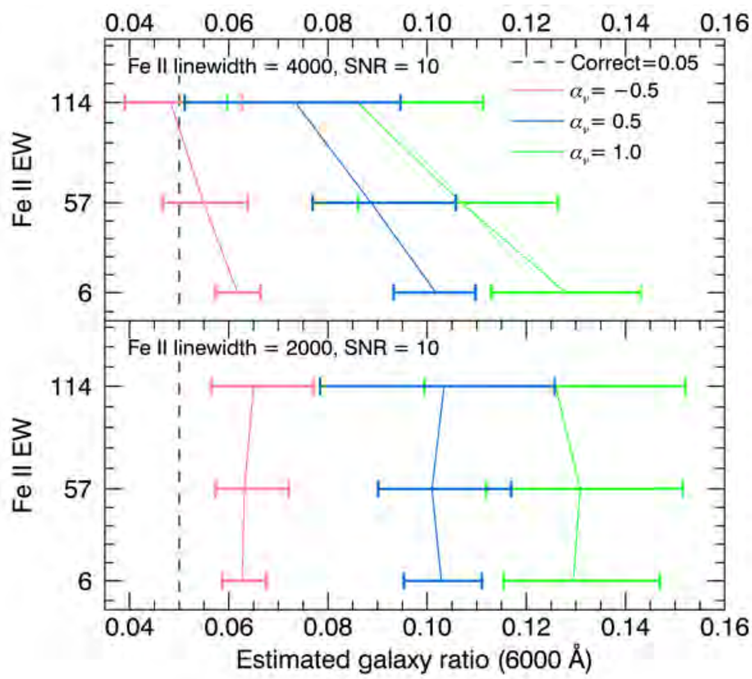


Figure B.1: *Top:* The galaxy ratio is overestimated for all combinations of input parameters, except when $\alpha_\nu = -0.5$ and Fe II EW = 114 Å. The estimates improve in accuracy with increasing Fe II EW. For $\alpha_\nu = -0.5$ the estimated galaxy ratios are offset by [−0.002, +0.012] (the galaxy ratio is estimated to be [−6, 24] % of the true value). For $\alpha_\nu = 1.0$ the estimated values are offset by [+0.036, +0.078] (overestimated by [70, 154] %).
Bottom: The estimated galaxy ratios are close to constant with changing Fe II EW, and always overestimated. For $\alpha_\nu = -0.5$ the estimated galaxy ratios are offset by [+0.013, +0.015] (overestimated by [24, 30] %). For $\alpha_\nu = 1.0$ the galaxy ratio is offset by [+0.076, +0.080] (overestimated by [150, 160] %).

Galaxy ratio = 0.25

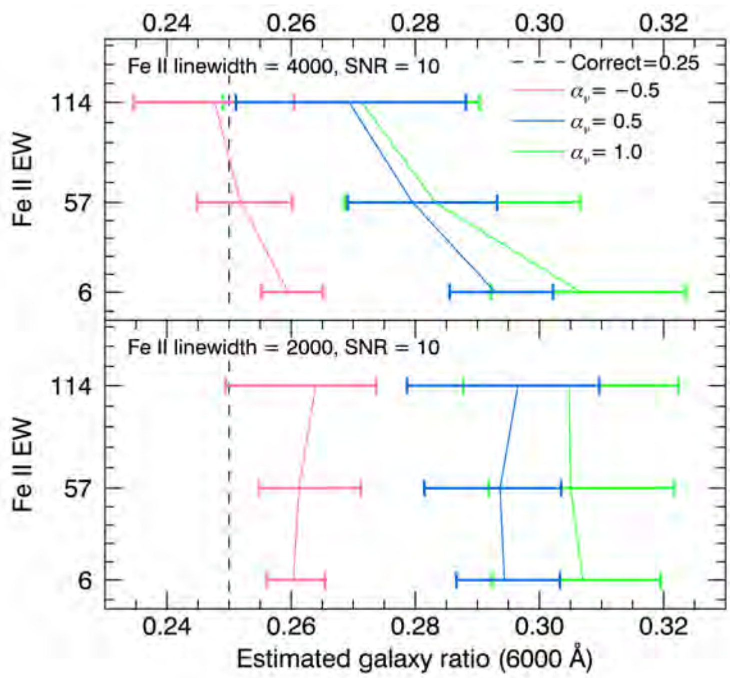


Figure B.2: *Top:* The galaxy ratio is overestimated in all cases except when $\alpha_\nu = -0.5$ and Fe II EW = 114 Å. The difference between $\alpha_\nu = 0.5$ and $\alpha_\nu = 1.0$ becomes smaller with larger amounts of iron in the spectra. For $\alpha_\nu = -0.5$ the galaxy ratio is offset by [−0.002, +0.01] (the galaxy ratio is estimated to be [−0.8, 4] % of the true value). For $\alpha_\nu = 1.0$ the estimated values are offset by [+0.021 +0.057] (the galaxy ratio is overestimated by [8, 24] %).
Bottom: The estimated galaxy ratios are close to constant with changing Fe II EW. For $\alpha_\nu = -0.5$ the estimated galaxy ratios are offset by [+0.01, +0.014] (overestimated by [4, 6] %). For $\alpha_\nu = 1.0$ the estimated values are offset by [+0.055, +0.057] (overestimated by [22, 24] %).

Galaxy ratio = 0.50

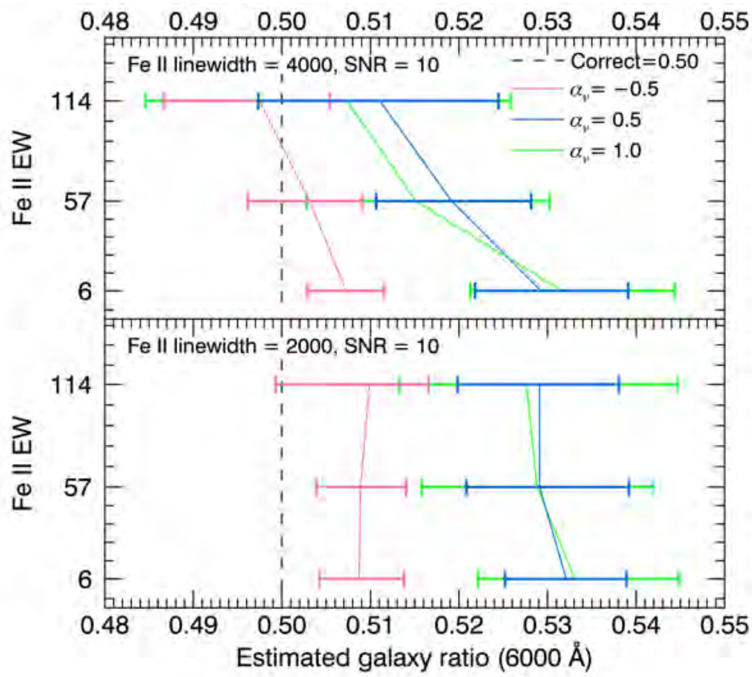


Figure B.3: *Top:* The galaxy ratio is overestimated in all cases except $\alpha_\nu = -0.5$ and Fe II EW $= 114$ Å. $\alpha_\nu = 0.5$ is better at estimating the true value than $\alpha_\nu = 1.0$ for higher Fe II EW, opposite of the above figures. When $\alpha_\nu = -0.5$ the offset in estimated galaxy ratios are $[-0.003, +0.007]$ (estimated to be $[-0.6, 1.6]$ % of the true value). For $\alpha_\nu = 1.0$ the offset is $[+0.007, +0.032]$ (overestimated by $[2.2, 5.8]$ %).
Bottom: The galaxy ratio estimates are close to constant and almost always overestimated with changing Fe II EW. The difference in accuracy between $\alpha_\nu = -0.5$ and 1.0 is negligible, but the precision is better for $\alpha_\nu = 0.5$. When $\alpha_\nu = -0.5$ the estimated galaxy ratios are offset by $[+0.009, +0.010]$ (the galaxy ratio is overestimated by $[1.6, 2]$ %). For $\alpha_\nu = 1.0$ the offset is $[+0.028, +0.033]$ (overestimated by $[5.8, 6.6]$ %).

Galaxy ratio = 0.75

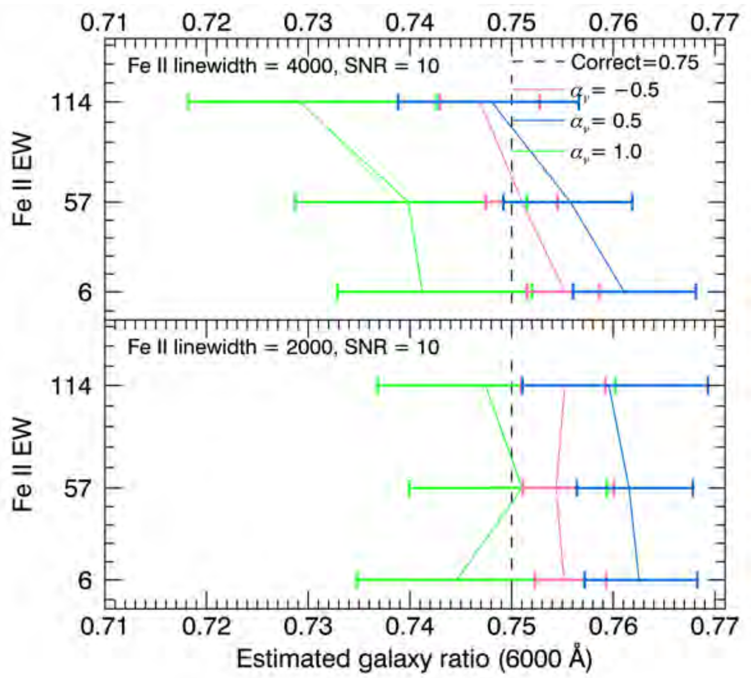


Figure B.4: $\alpha_\nu = 1.0$ is now responsible for the lowest estimated galaxy ratios.
Top: All values of α_ν are underestimating the galaxy ratio when Fe II EW $= 114$ Å. When $\alpha_\nu = -0.5$ the estimated galaxy ratio is offset by $[-0.003, +0.005]$ (the galaxy ratios are estimated to be $[-0.5, 0.8]$ % of the true value). For $\alpha_\nu = 1.0$ the offset is $[-0.021, -0.009]$ (underestimated by $[-2.8, -1.2]$ %).
Bottom: The estimated galaxy ratios are close to constant with increased Fe II EW and generally overestimated. When $\alpha_\nu = -0.5$ the estimated galaxy ratios are offset by $[+0.004, +0.005]$ (estimated to be $[1.2, 1.7]$ % of the true value. For $\alpha_\nu = 0.5$ the offset is $[-0.005, +0.001]$ (estimated to be $[-0.4, 0.8]$ % of the true value).

B.2 Estimated galaxy ratio v. Signal-to-noise ratio

The figures below show how the estimated galaxy ratio changes with S/N ratio for different galaxy ratios as a function of Fe II EW, for different α_ν. Fe II linewidth = 4000 km s^{-1} (top) and Fe II linewidth = 2000 km s^{-1} (bottom). Here I focus on discussing how much the galaxy ratio is overestimated in terms of the *precision* for the different S/N ratios. The accuracy does not change significantly when the S/N ratio is changed, thus the accuracies can still be described by the offsets and percentage errors stated in the previous section. I use the symbol σ to quote the level of precision instead of $IPR/2$ for legibility. Do note: Because the distributions are sometimes highly asymmetric around the median, the true value value can be seen to fall outside the spread shown in the figures below, but still be said to be within 1σ.

In the figure captions below, I do not comment on the intermediate S/N ratios or Fe II EW = 57 Å, because the results for these parameter values tend to fall between the other values being considered.

In the following figures, the x-axis is scaled to cover the same range for the figures associated with each input galaxy ratio, to allow for direct comparison between the figures.

A brief summary of the results: A common trend is that the accuracy of the estimates does not change significantly with increased S/N ratio. The precision increases with increasing S/N ratio is increased. The precision also tends to increase with decreasing Fe II EW.

The true value for the galaxy ratio is almost always within the uncertainties for S/N = 5, due to the large spread in values, or within 2σ at the most. For S/N = 50, the true value is within the uncertainties for some combinations of parameters, mostly when Fe II EW = 114 Å. The precision can be very high for S/N = 50 in some cases (i.e., very small amount of scatter), at the same time as the estimated galaxy ratios can be highly overestimated. This means that the true value can – in the worst case – be as far away as 30σ. In the best case, the true value will also be within the uncertainties for S/N = 50.

The figures begin on the following page.

Galaxy ratio = 0.05

Power-law index, $\alpha_\nu = -0.5$

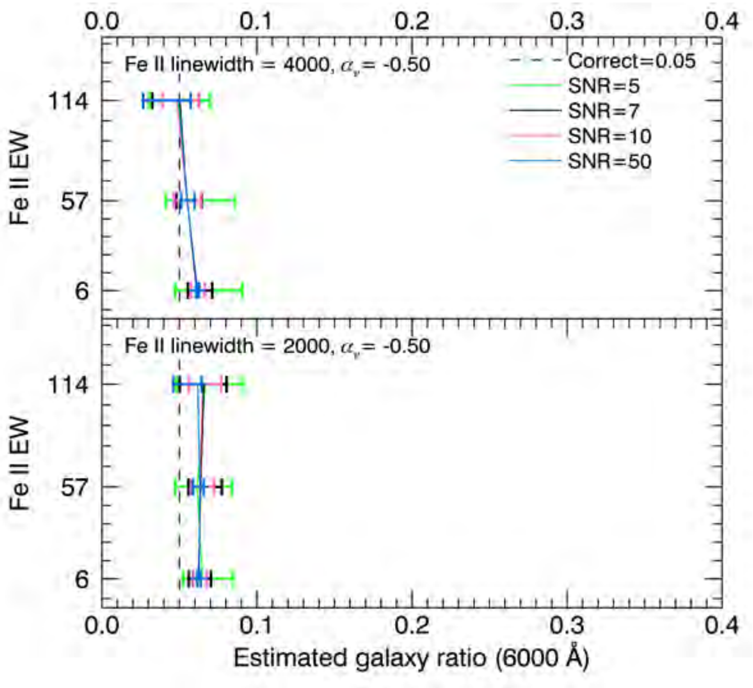


Figure B.5: The galaxy ratio tends to be overestimated. When Fe II EW = 114 Å (top) the distributions are close to centred around the true value. The precision improves with *de*creased Fe II EW for higher S/N ratios. When Fe II EW = 114 Å, the true value of the galaxy ratio is within the uncertainties for both S/N = 5 and S/N = 50 in both panels. Also in both panels, for Fe II EW = 6 Å, the accuracy is approximately within 1σ of the true vale for S/N = 5, but 13σ for S/N = 50.

Power-law index, $\alpha_\nu = 1.0$

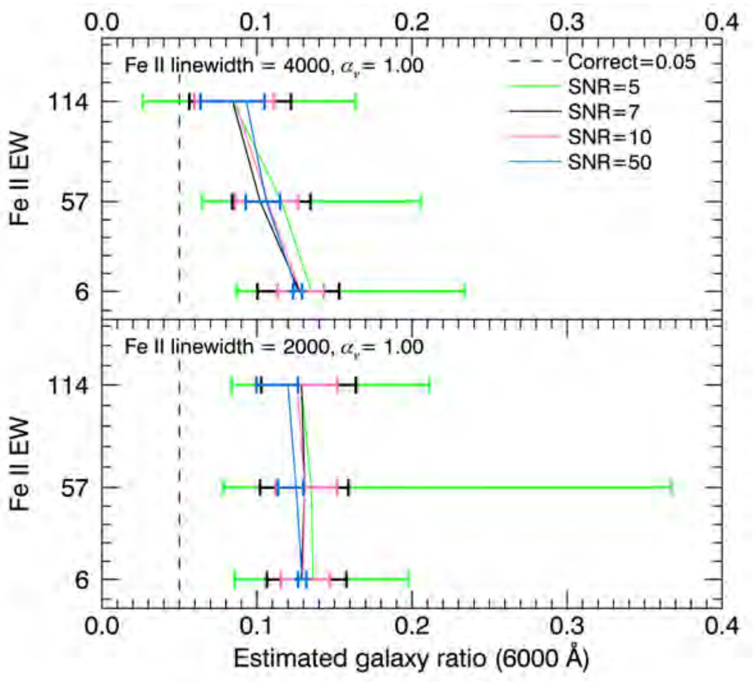


Figure B.6: The estimated galaxy ratio tends to be overestimated relative to the true value. The precision is low for S/N = 5 compared to S/N = 50. The precision improves significantly with *de*creased Fe II EW for higher S/N ratios.
Top: When Fe II EW = 114 Å, the true value of the galaxy ratio is within the uncertainties for S/N = 5, while within 2σ for S/N = 50. For Fe II EW = 6 Å, the true value of the galaxy ratio is within 2σ for S/N = 5, while within 25σ for S/N = 50.
Bottom: When Fe II EW = 114 Å, the true value of the galaxy ratio is within 2σ for S/N=5, and within 5σ for S/N = 50. For Fe II EW = 6 Å, the true value is within 2σ for S/N=5, and within 26σ for S/N = 50.

Galaxy ratio = 0.25

Power-law index, $\alpha_\nu = -0.5$

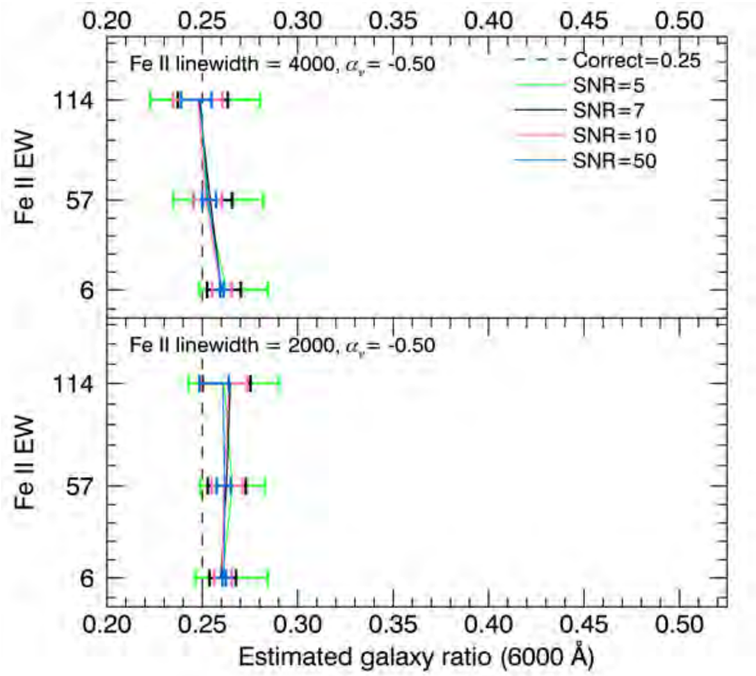


Figure B.7: The galaxy ratio tends to be overestimated. When Fe II EW = 114 Å (top) the distributions are close to centred around the true value. The precision improves with *de*creased Fe II EW for higher S/N ratios. The true value for the galaxy ratio is within the uncertainties for S/N = 5 for all combinations of parameters. For Fe II EW = 114 Å, the true value is approximately within 1σ for S/N = 50 in both panels. For S/N = 50 and Fe II EW = 6 Å, the true value is within 11σ when the iron emission is narrow (bottom) and within 1σ when the iron emission is broad (top).

Power-law index, $\alpha_\nu = 1.0$

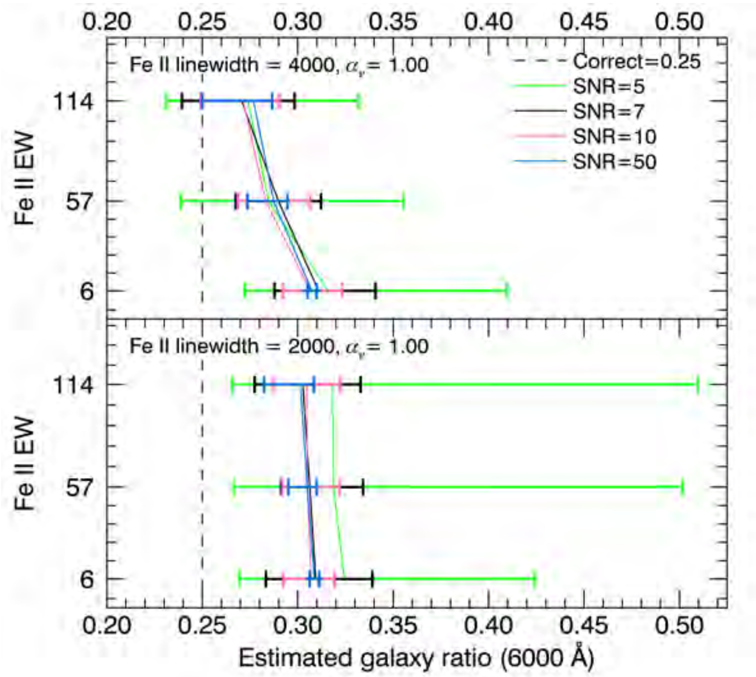


Figure B.8: The galaxy ratio tends to be overestimated. The precision is low for S/N = 5 compared to S/N = 50. The precision improves with *de*creased Fe II EW for higher S/N ratios.
Top: For Fe II EW = 114 Å, the true value is within the uncertainties for all S/N ratios. For Fe II EW = 6 Å, the true value is within 1σ for S/N = 5 and within 29σ for S/N = 50.
Bottom: For S/N = 5 the true value is within 1σ for all values of Fe II EW. For S/N = 50 the true value is within 4σ for Fe II EW = 114 Å, and within 30σ for Fe II EW = 6 Å.

Galaxy ratio = 0.50

Power-law index, $\alpha_\nu = -0.5$

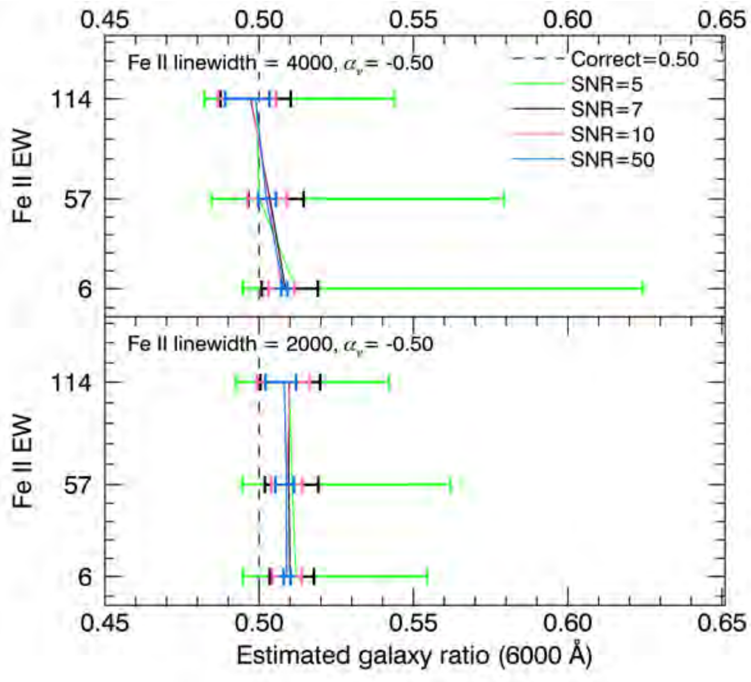


Figure B.9: The galaxy ratio tends to be overestimated. When Fe II EW = 114 Å (top) the distributions are close to centred around the true value. The precision is low for S/N = 5 compared to S/N = 50. The precision improves noticeably with *de*creased Fe II EW for higher S/N ratios.
Top: For Fe II EW = 114 Å, the true value is within the uncertainties for all S/N ratios. For Fe II EW = 6 Å, the true value is within 1σ for S/N = 5 and within 8σ for S/N = 50.
Bottom: For S/N = 5 the true value is within the uncertainties for all values of Fe II EW. For S/N = 50 the true value is within 1σ for Fe II EW = 114 Å, and within 9σ for Fe II EW = 6 Å.

Power-law index, $\alpha_\nu = 1.0$

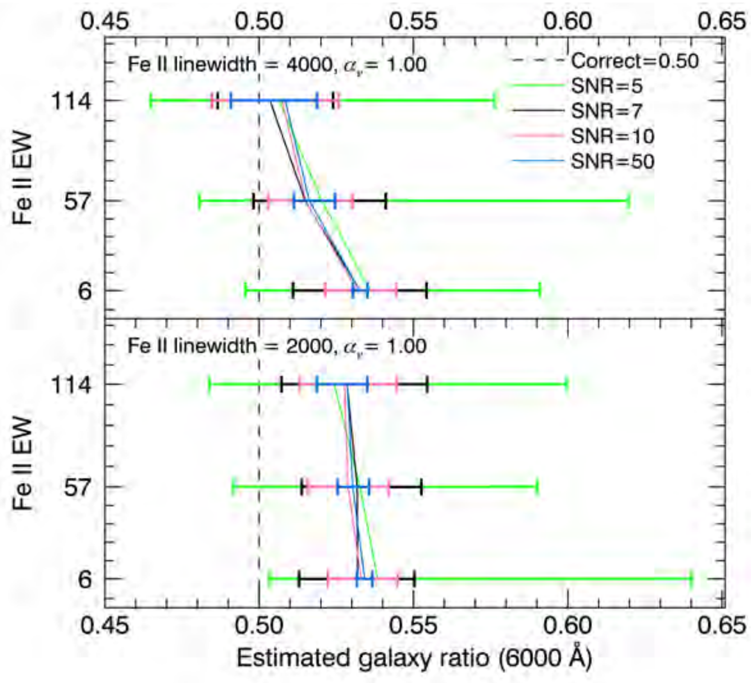


Figure B.10: The galaxy ratio tends to be overestimated. The precision is low for S/N = 5 compared to S/N = 50. The precision improves noticeably with *de*creased Fe II EW for higher S/N ratios.
Top: For Fe II EW = 114 Å, the true value is within the uncertainties for all S/N ratios. For Fe II EW = 6 Å, the true value is within 1σ for S/N = 5 and within 17σ for S/N = 50.
Bottom: For S/N = 5 the true value is within 1σ for all values of Fe II EW. For S/N = 50 the true value is within 1σ for Fe II EW = 114 Å, and within 12σ for Fe II EW = 6 Å.

Galaxy ratio = 0.75

Power-law index, $\alpha_\nu = -0.5$

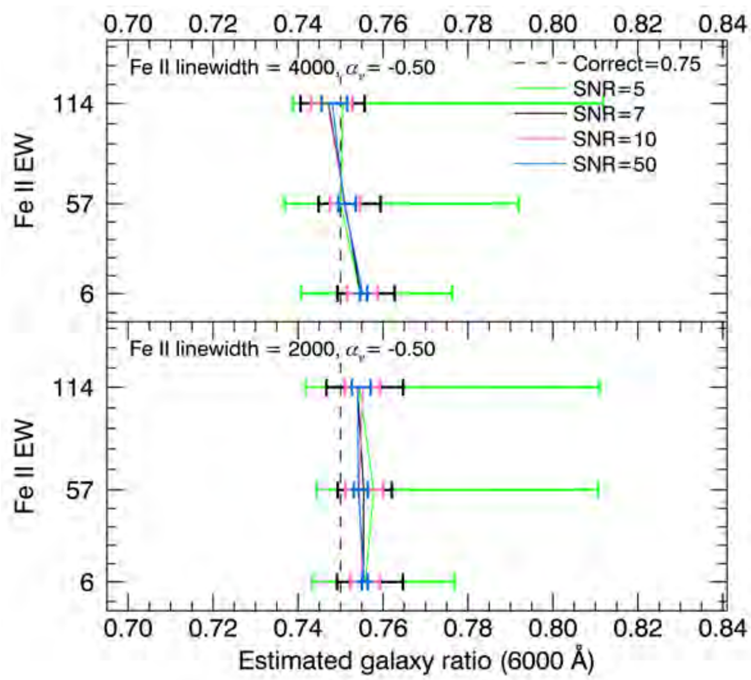


Figure B.11: The precision is low for S/N = 5 compared to S/N = 50. *Top:* The galaxy ratio tends to be overestimated relative to the true value when Fe II EW = 6 Å, and underestimated when Fe II EW = 114 Å. In latter case, the true value for the galaxy ratio is within the uncertainties for all S/N ratios. When Fe II EW = 6 Å, the true value is within the uncertainties for S/N = 5, while it is 5σ away for S/N = 50.
Bottom: The galaxy ratio tends to be overestimated and the accuracy does not change significantly with increasing Fe II EW. The true value for the galaxy ratio is within the uncertainties for S/N = 5 for all values of Fe II EW. For S/N = 50, the accuracy is 2–6σ away from the true value.

Power-law index, $\alpha_\nu = 1.0$

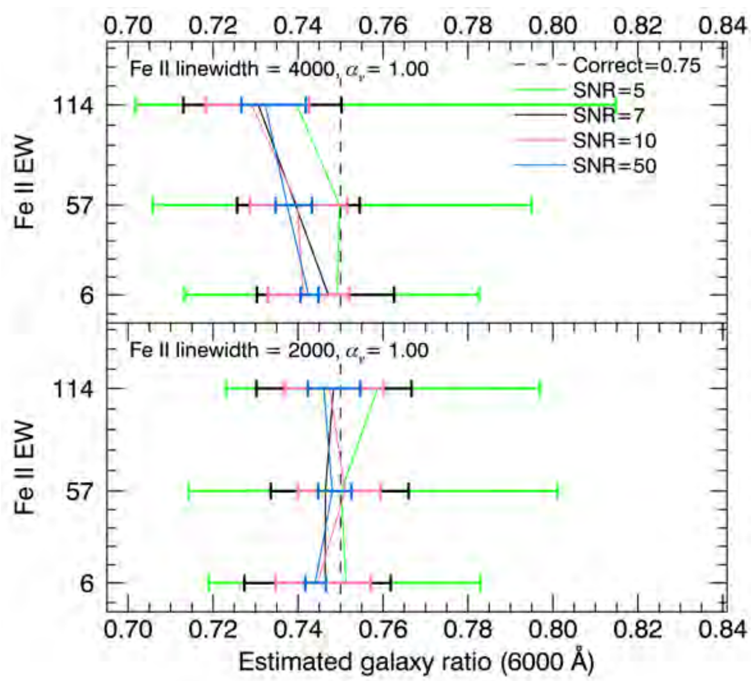


Figure B.12: The precision is low for S/N = 5 compared to S/N = 50. *Top:* The galaxy ratio tends to be underestimated, and more so for increased Fe II EW. The true value for the galaxy ratio is within the uncertainties for S/N = 5 for all values of Fe II EW. For S/N = 50 the accuracy is 4σ away from the true value.
Bottom: The estimated galaxy ratios are close to centred around the true value. The true value of the galaxy ratio is within the uncertainties for all S/N ratios, except for S/N = 50 when Fe II EW = 6 Å. Then the accuracy is 3σ away from the true value.

C Figures: Estimated galaxy ratio (histograms)

The histograms on the following pages show the estimated galaxy ratios for different signal-to-noise ratios (S/N) for all combinations of parameters as described in Table 3.2. Fe II linewidth = 2000 $\mathrm{km\,s^{-1}}$ (left) and 4000 $\mathrm{km\,s^{-1}}$ (right), with Fe II EW = 6 (top), 57 (middle), 114 (bottom).

Outliers have been excluded in order to clearly see the shape of the centre of the distribution. The definition of an outlier is given in Section 5.1. The number of outliers are stated in the legend of each histogram. The range of values that are found to be outliers are given in the subcaptions, with one range for outliers below the median and one range for the outliers above the median: [range below median, range above median].

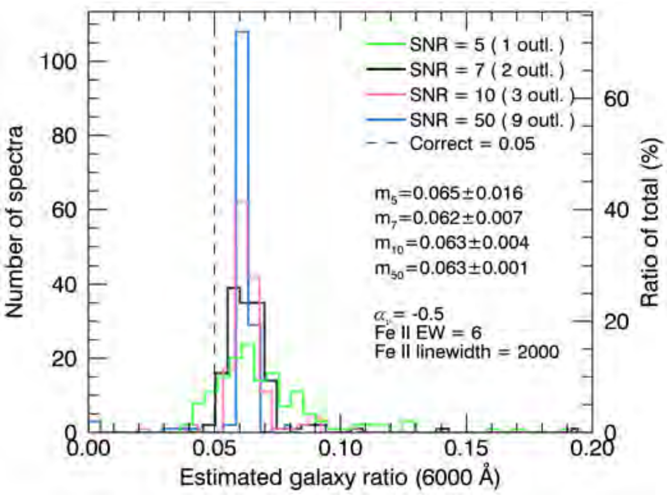
The accuracy and precision of the distributions is noted in the figures as $m_{\mathrm{S/N}} \pm IPR/2$.

For each page of figures, the x-axis is the same in all figures so that the distributions can be compared directly.

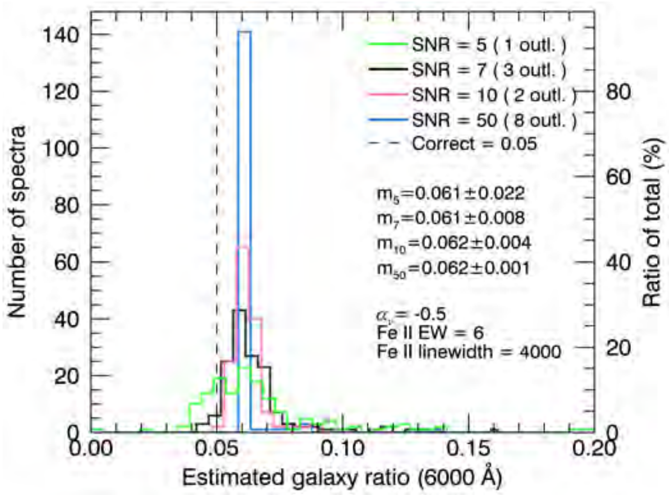
The figures begin on the following page.

Galaxy ratio = 0.05

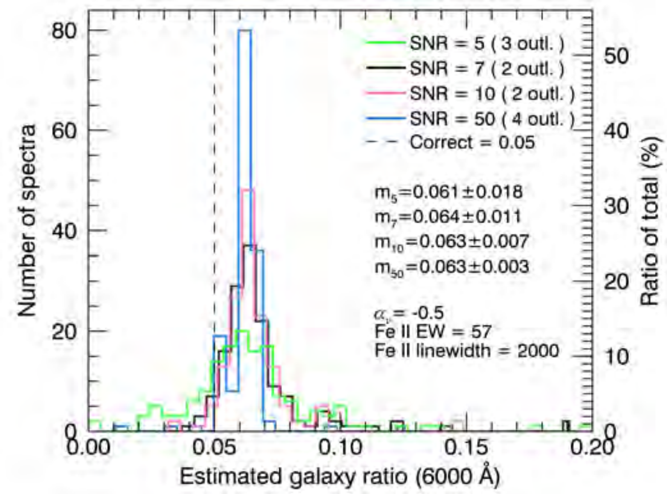
Power-law index = –0.5



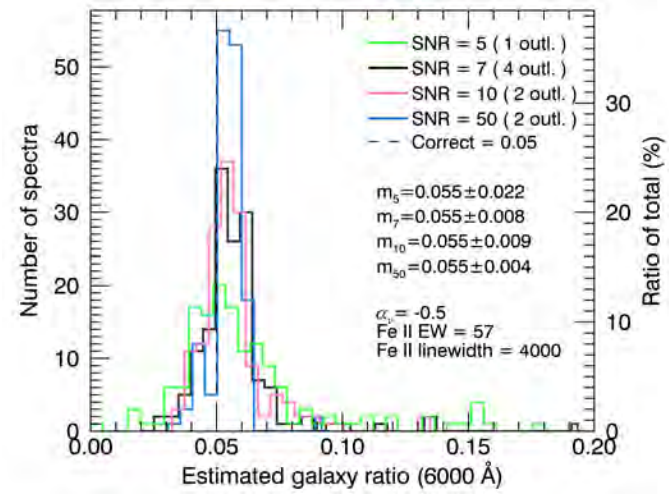
(a) Outliers: [none, 0.191], [none, 0.141–0.193], [0.019, 0.106–0.335], [0.000–0.040, 0.078–0.083]



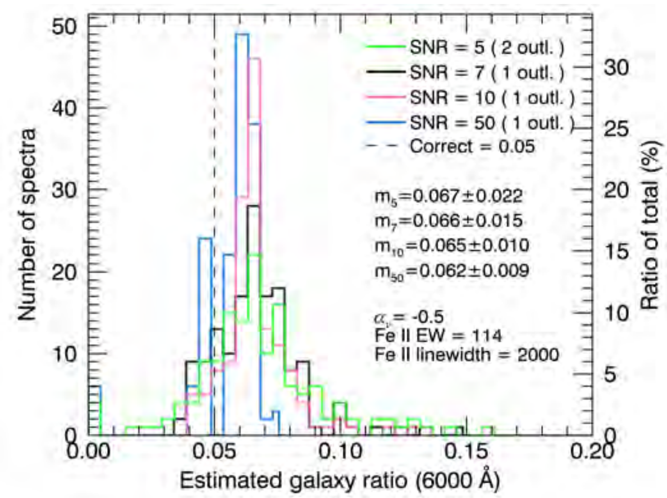
(b) Outliers: [none, 0.062], [none, 0.130–0.161], [none, 0.101–0.112], [0.000, 0.070–0.139]



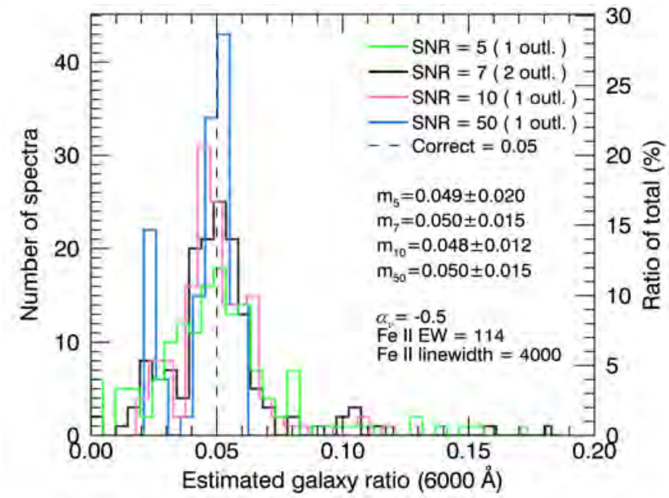
(c) Outliers: [none, 0.197–0.259], [none, 0.190–0.191], [none, 0.145–0.146], [0.011–0.030, 0.097–0.123]



(d) Outliers: [0.029, none], [none, 0.114–0.193], [none, 0.130–0.132], [none, 0.092–0.092]



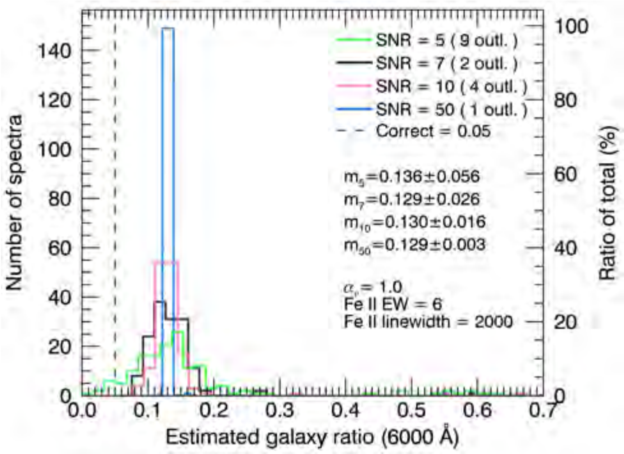
(e) Outliers: [none, 0.229–0.375], [none, 0.052], [none, 0.083], [none, 0.063]



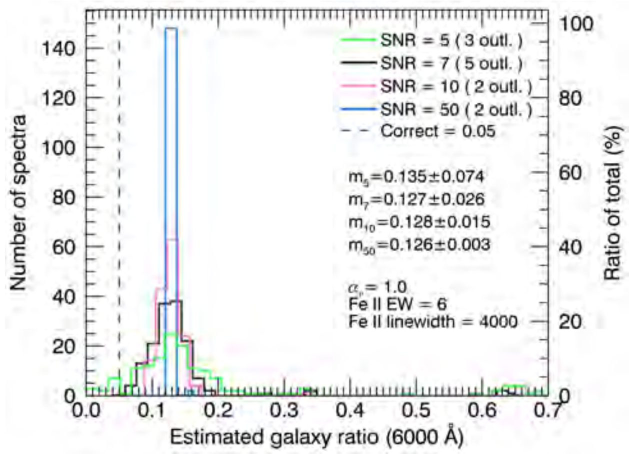
(f) Outliers: [none, 0.062], [none, 0.158–0.182], [none, 0.064], [0.042, none]

Figure C.1

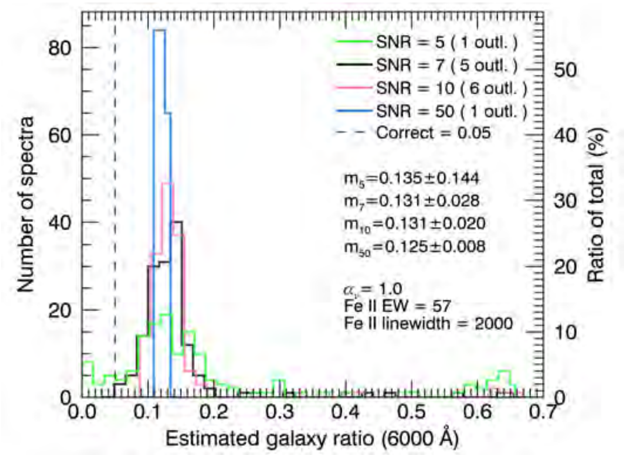
Power-law index = 1.0



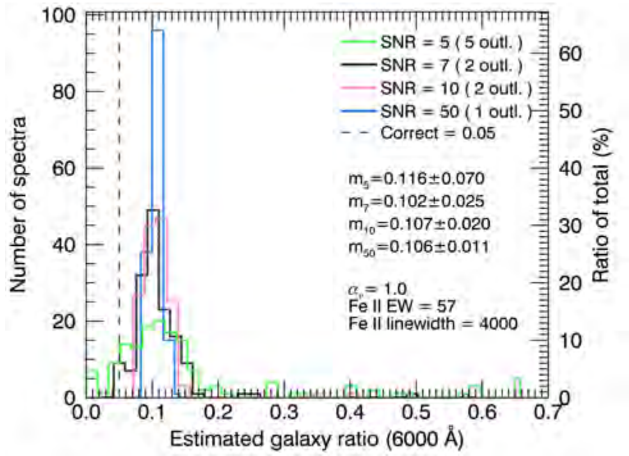
(a) Outliers: [none, 0.543–0.647], [none, 0.603–0.729], [none, 0.272–0.672], [none, 0.164]



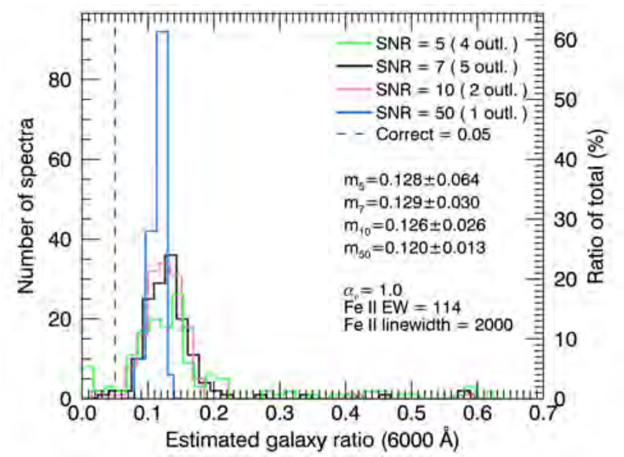
(b) Outliers: [none, 0.677–0.693], [none, 0.344–0.641], [none, 0.334–0.584], [none, 0.155–0.159]



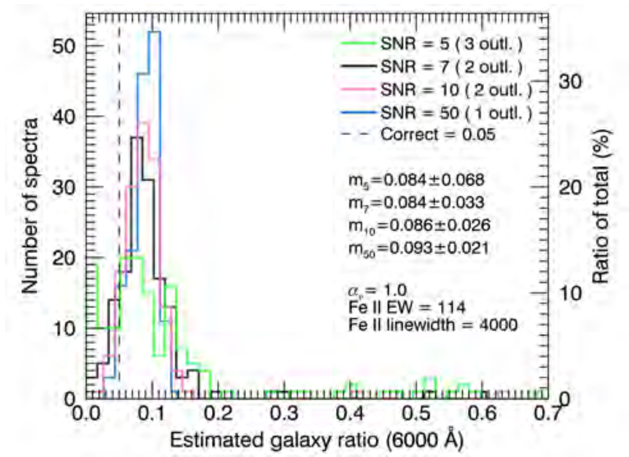
(c) Outliers: [none, 0.638], [none, 0.432–0.652], [none, 0.409–0.654], [none, 0.125]



(d) Outliers: [none, 0.650–0.658], [none, 0.501–0.580], [none, 0.355–0.402], [none, 0.112]



(e) Outliers: [none, 0.594–0.625], [none, 0.407–0.593], [none, 0.409–0.592], [none, 0.126]

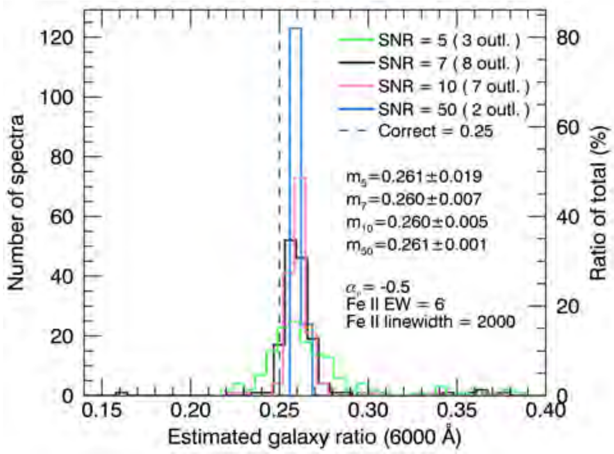


(f) Outliers: [none, 0.575–0.686], [none, 0.513–0.607], [none, 0.402–0.637], [none, 0.090]

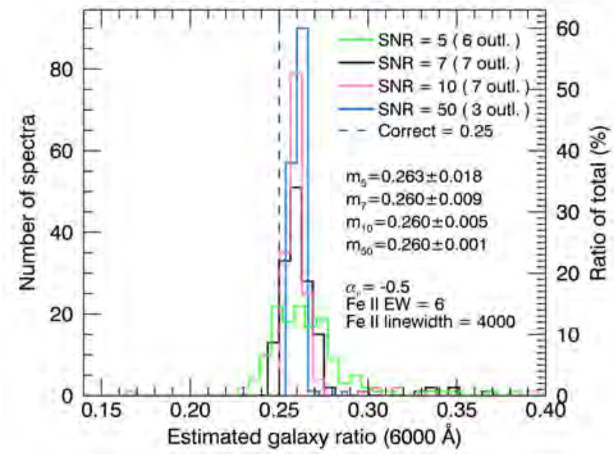
Figure C.2

Galaxy ratio = 0.25

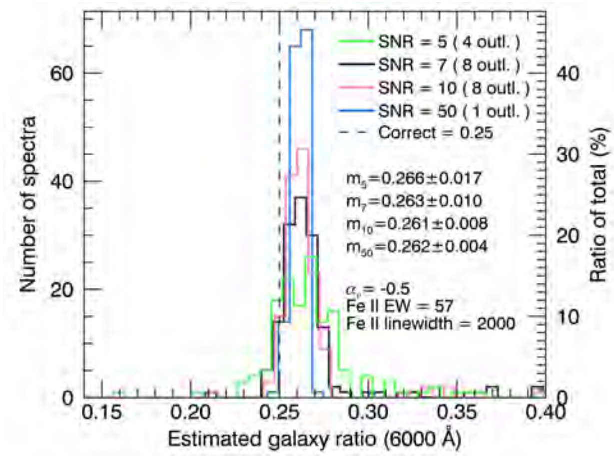
Power-law index = –0.5



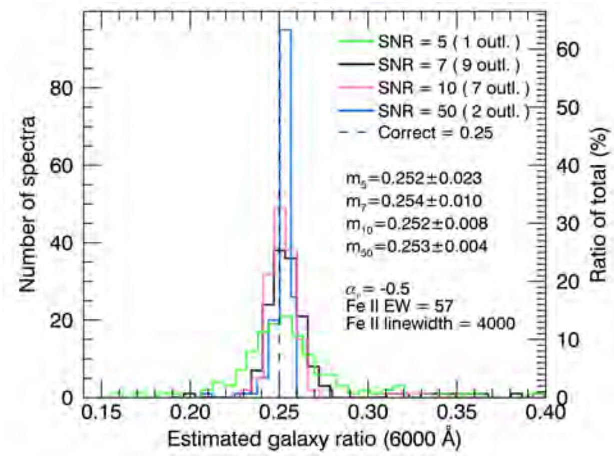
(a) Outliers: [none, 0.424–0.433], [0.126–0.159, 0.347–0.406], [0.220–0.227, 0.301–0.370], [none, 0.299–0.367]



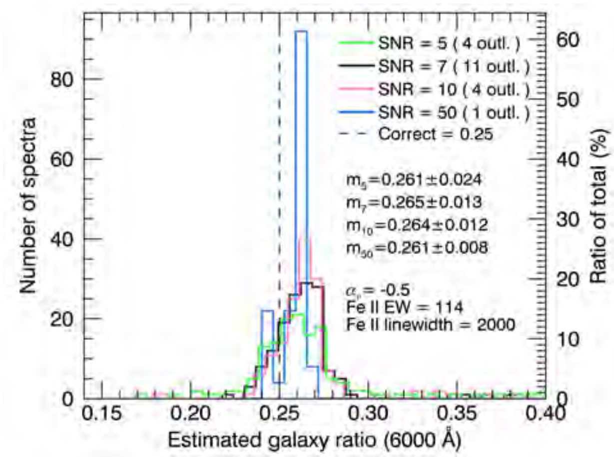
(b) Outliers: [0.0744, 0.412–0.727], [none, 0.329–0.407], [none, 0.296–0.395], [0.000, 0.268–0.287]



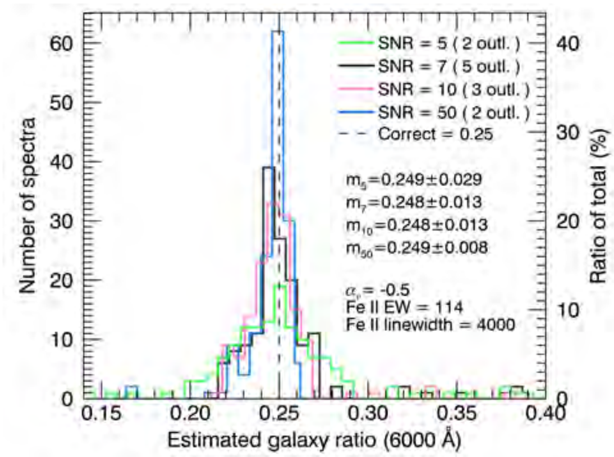
(c) Outliers: [0.049, 0.414–0.455], [none, 0.357–0.416], [none, 0.323–0.425], [none, 0.302]



(d) Outliers: [none, 0.438], [none, 0.330–0.412], [none, 0.306–0.394], [0.206–0.228, none]



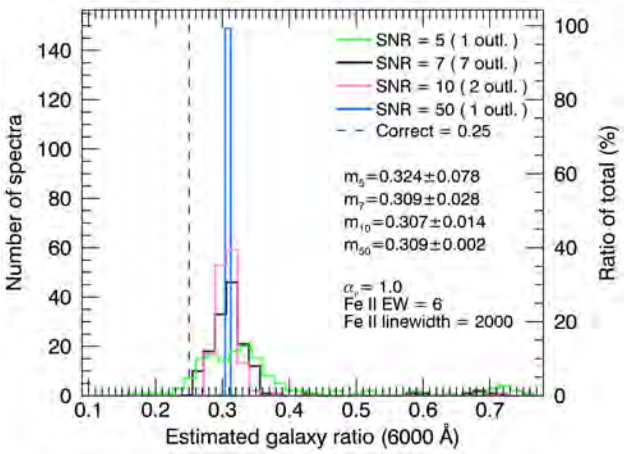
(e) Outliers: [none, 0.436–0.481], [none, 0.351–0.435], [none, 0.366–0.495], [none, 0.366]



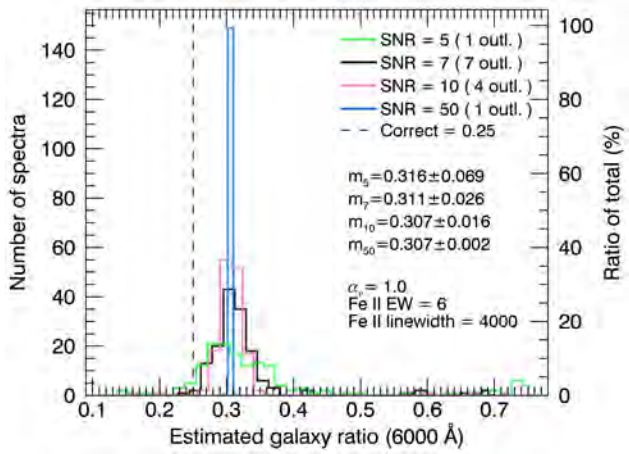
(f) Outliers: [0.000, none], [none, 0.354–0.391], [none, 0.377–0.711], [0.164, none]

Figure C.3

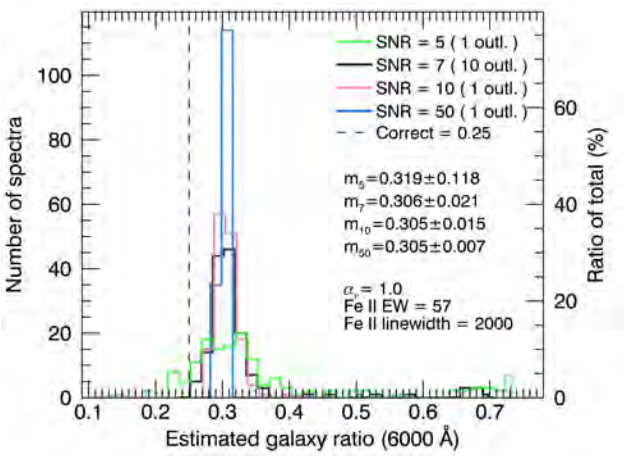
Power-law index = 1.0



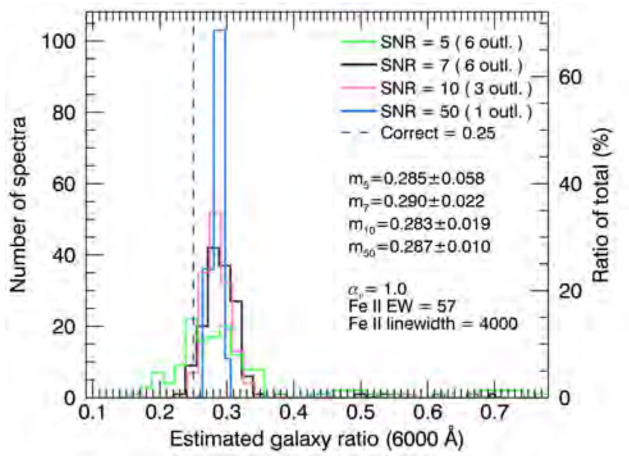
(a) Outliers: [0.241, none], [none, 0.592–0.710], [none, 0.683–0.722], [none, 0.310]



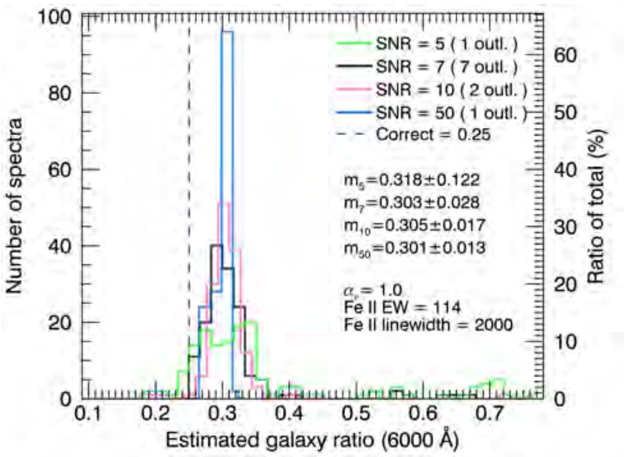
(b) Outliers: [none, 0.756], [none, 0.573–0.696], [0.156–0.196, 0.602–0.661], [none, 0.310]



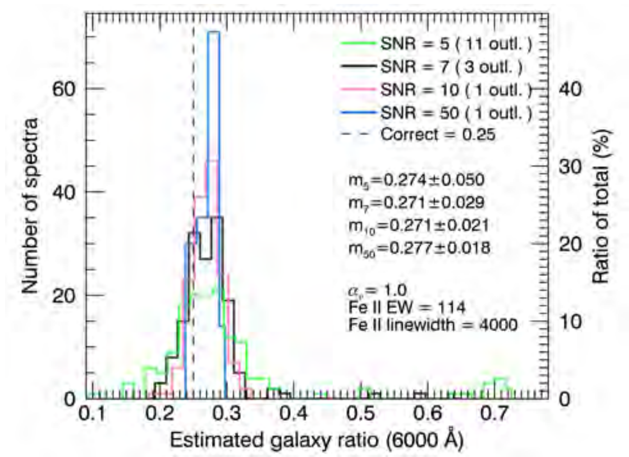
(c) Outliers: [none, 0.330], [none, 0.463–0.690], [none, 0.317], [none, 0.507]



(d) Outliers: [none, 0.717–0.790], [none, 0.501–0.695], [none, 0.515–0.670], [none, 0.294]



(e) Outliers: [0.196, none], [none, 0.558–0.673], [0.058, 0.730], [none, 0.279]

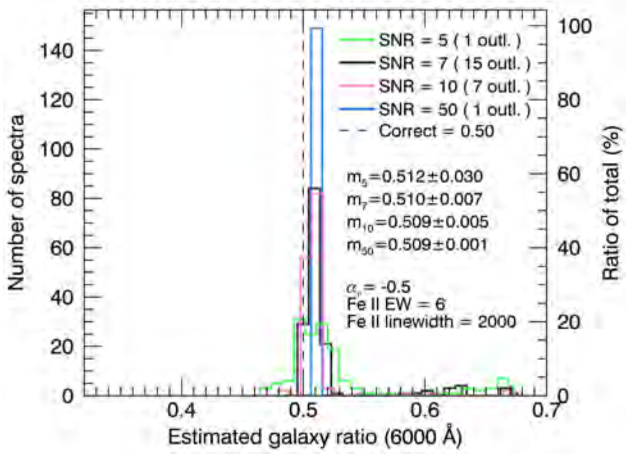


(f) Outliers: [none, 0.641–0.722], [none, 0.526–0.642], [none, 0.254], [none, 0.282]

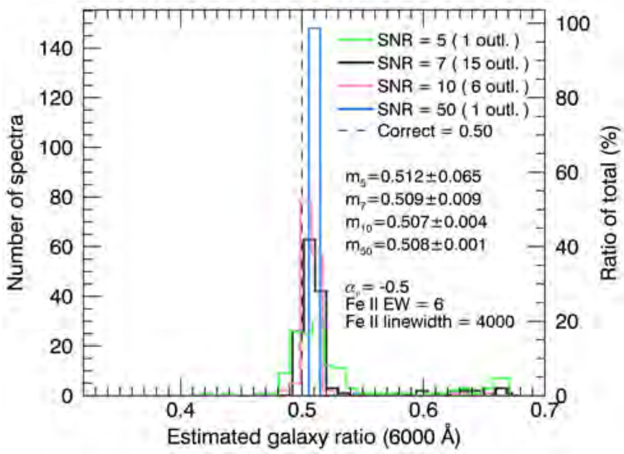
Figure C.4

Galaxy ratio = 0.50

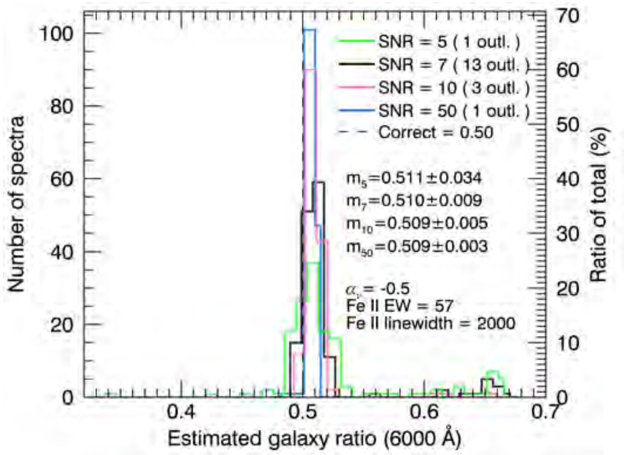
Power-law index = –0.5



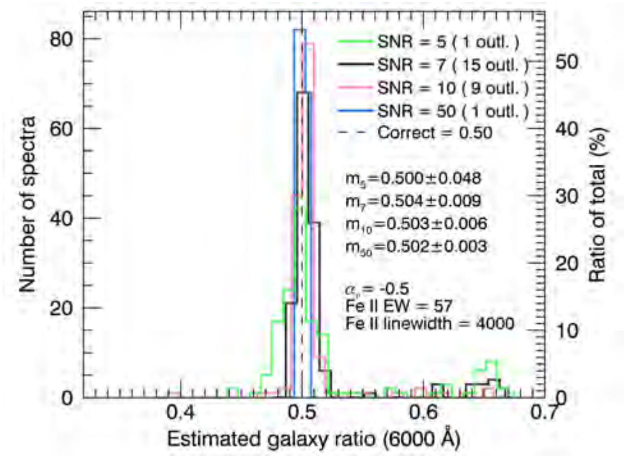
(a) Outliers: [none, 0.645], [none, 0.583–0.672], [none, 0.546–0.669], [none, 0.553]



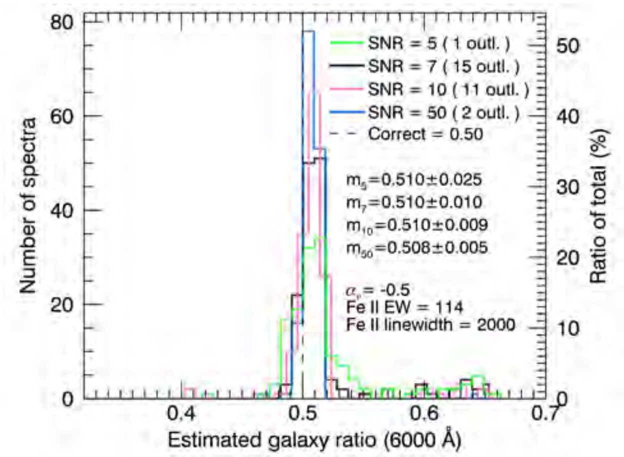
(b) Outliers: [0.499, none], [none, 0.580–0.671], [0.462, 0.590–0.655], [none, 0.546]



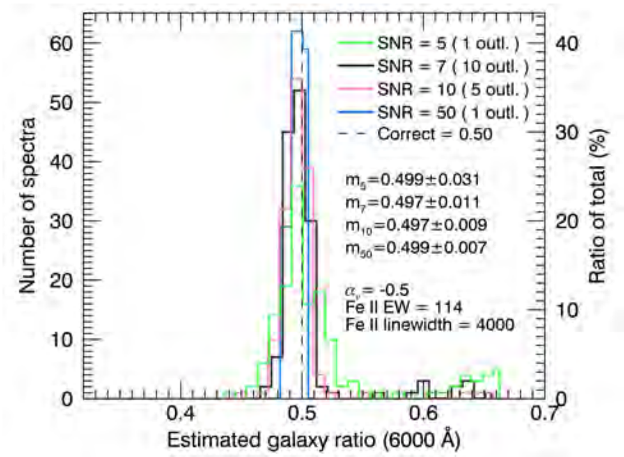
(c) Outliers: [none, 0.503], [none, 0.614–0.667], [none, 0.613–0.655], [0.482, none]



(d) Outliers: [none, 0.669], [none, 0.608–0.668], [0.389, 0.548–0.652], [0.496, none]



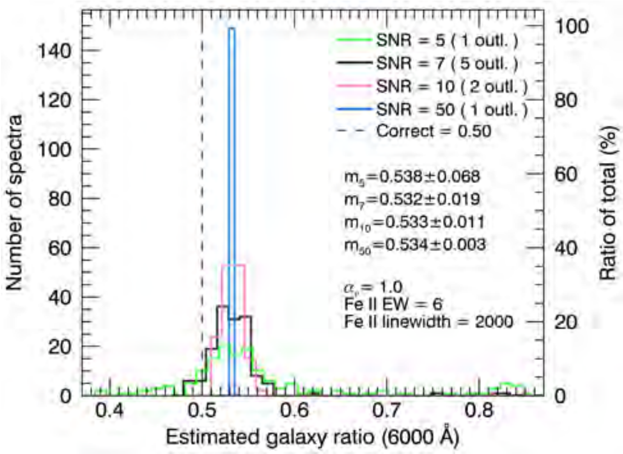
(e) Outliers: [0.000, none], [none, 0.594–0.656], [0.403–0.408, 0.592–0.653], [0.463, 0.644]



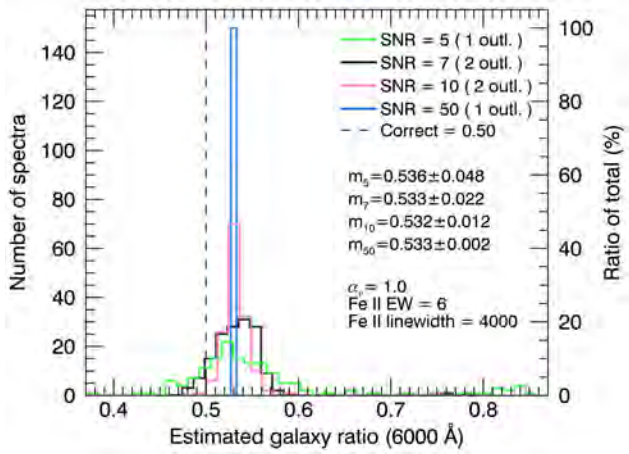
(f) Outliers: [0.000, none], [none, 0.588–0.654], [none, 0.610–0.650], [0.494, none]

Figure C.5

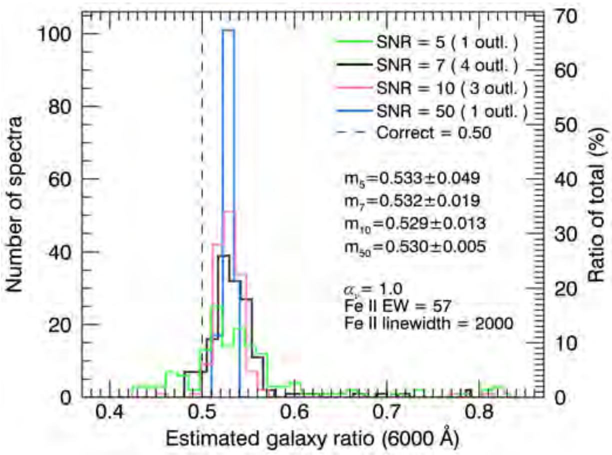
Power-law index = 1.0



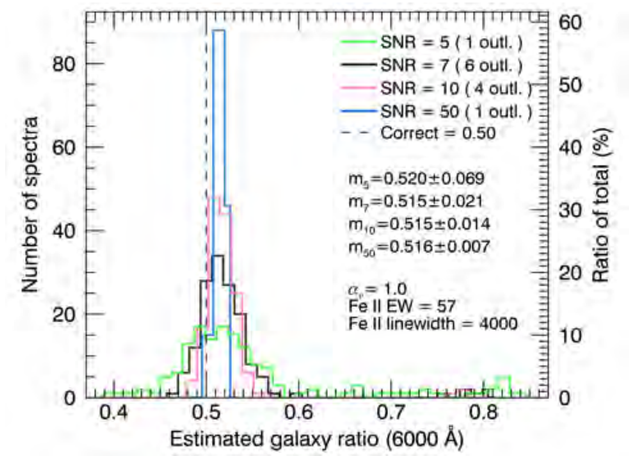
(a) Outliers: [none, 0.552], [none, 0.752–0.915], [none, 0.798–0.809], [none, 0.534]



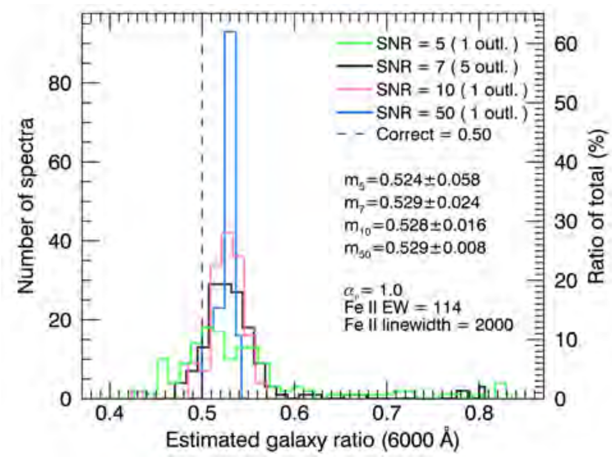
(b) Outliers: [0.497, none], [none, 0.766–0.791], [none, 0.717–0.751], [none, 0.536]



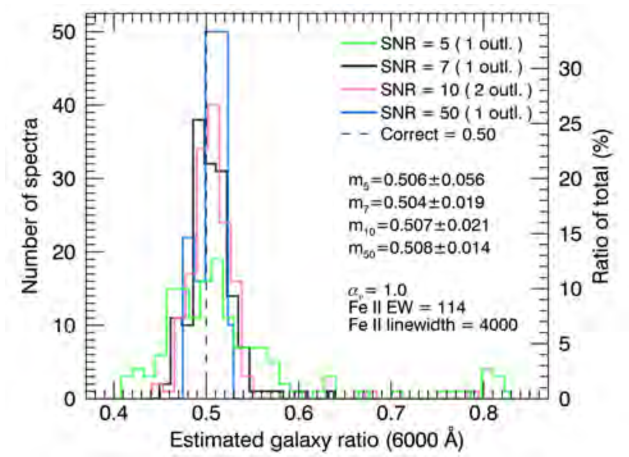
(c) Outliers: [0.119, none], [none, 0.699–0.789], [0.011, 0.776–0.812], [none, 0.526]



(d) Outliers: [none, 0.517], [none, 0.756–0.807], [0.0267, 0.770–0.779], [none, 0.506]



(e) Outliers: [0.000, none], [none, 0.777–0.810], [none, 0.546], [0.450, none]

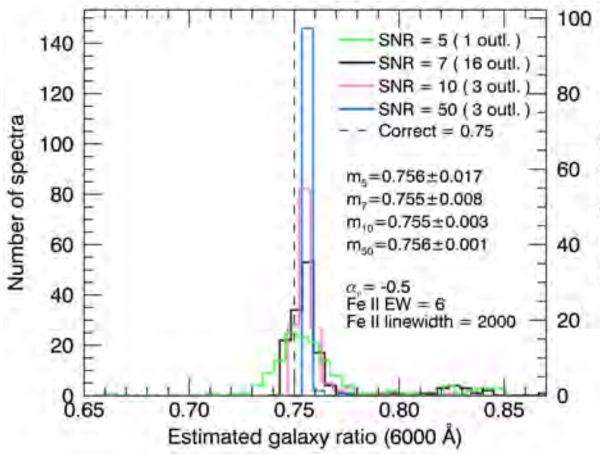


(f) Outliers: [none, 0.521], [none, 0.642], [none, 0.677–0.793], [0.477, none]

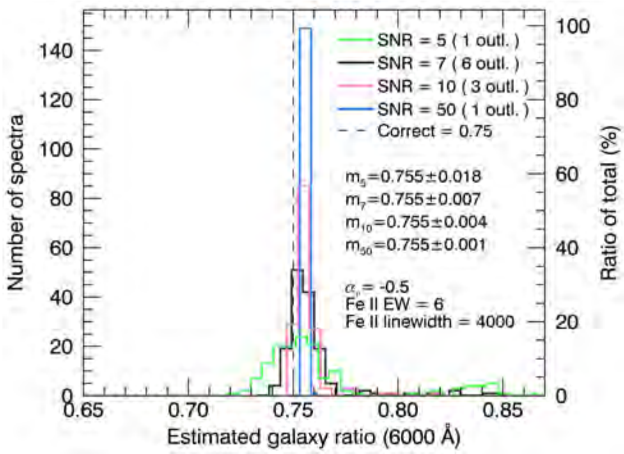
Figure C.6

Galaxy ratio = 0.75

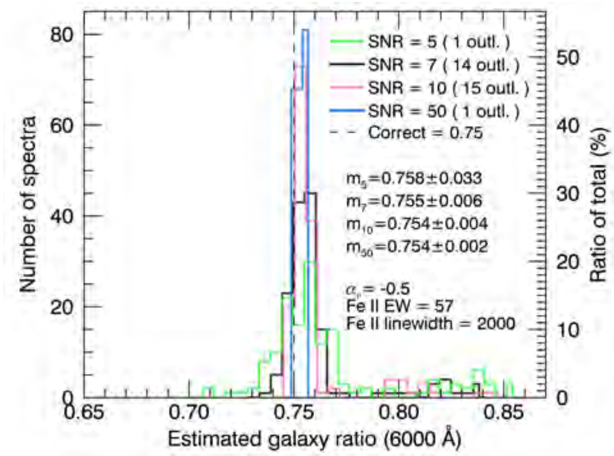
Power-law index = –0.5



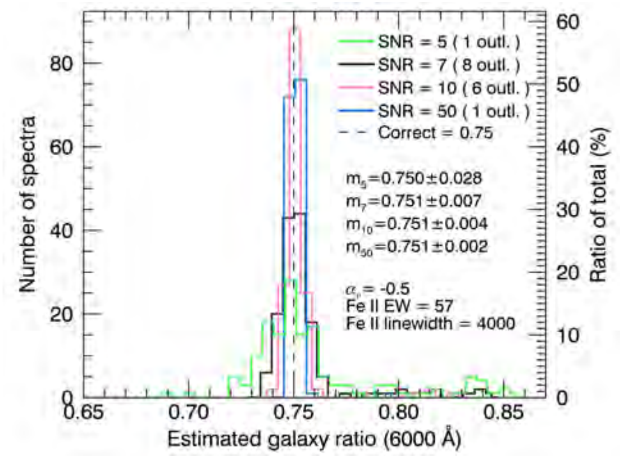
(a) Outliers: [none, 0.752], [none, 0.812–0.869], [none, 0.788–0.823], [none, 0.761–0.775]



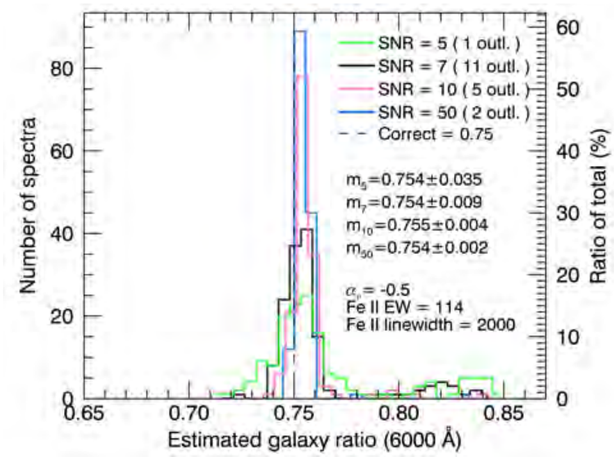
(b) Outliers: [0.000, none], [none, 0.804–0.850], [none, 0.794–0.817], [none, 0.764]



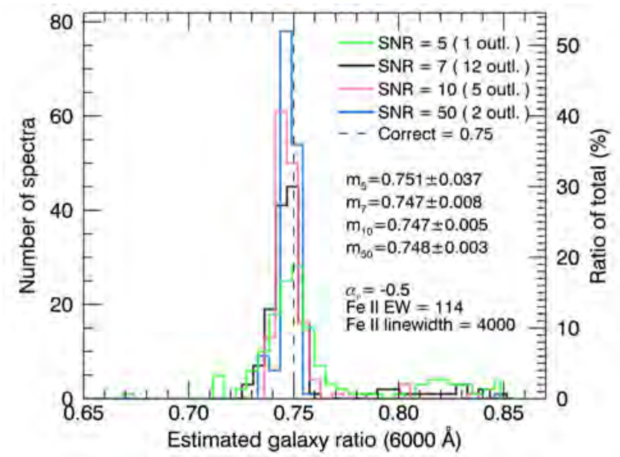
(c) Outliers: [0.748, none], [none, 0.801–0.839], [none, 0.795–0.844], [none, 0.753]



(d) Outliers: [0.000, none], [none, 0.809–0.842], [none, 0.783–0.827], [none, 0.794]



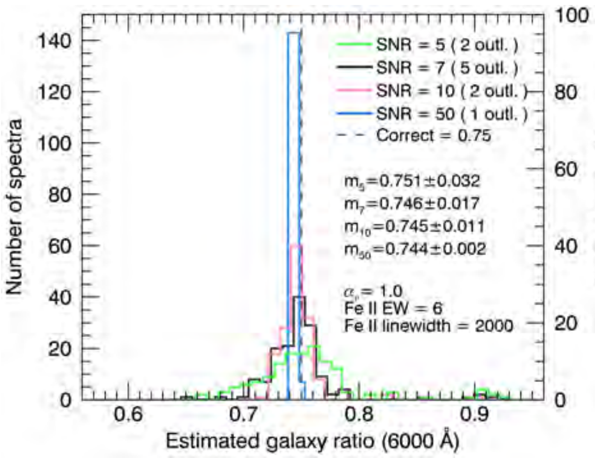
(e) Outliers: [0.745, none], [none, 0.820–0.845], [none, 0.785–0.839], [none, 0.777–0.833]



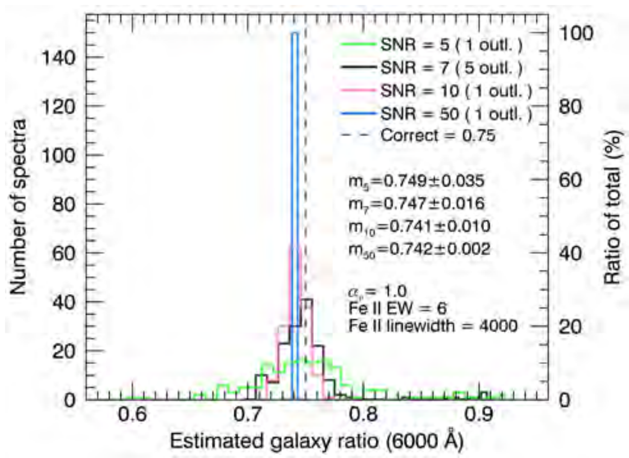
(f) Outliers: [none, 0.759], [none, 0.804–0.849], [none, 0.783–0.833], [0.626, 0.847]

Figure C.7

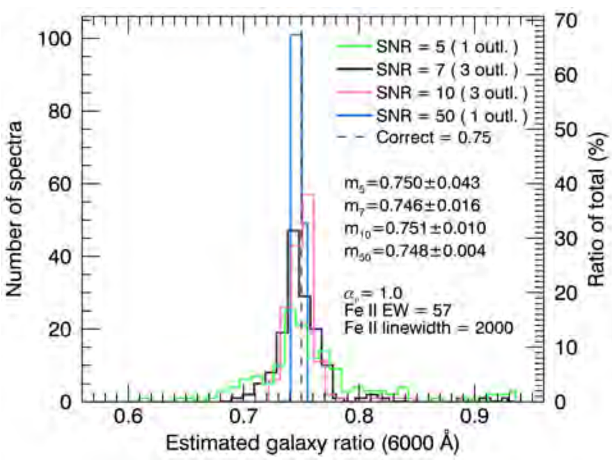
Power-law index = 1.0



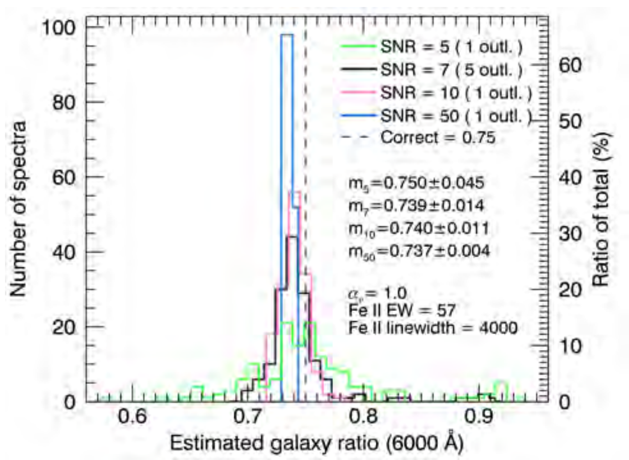
(a) Outliers: [0.346, 0.996], [none, 0.896–0.920], [none, 0.828–0.899], [0.740, none]



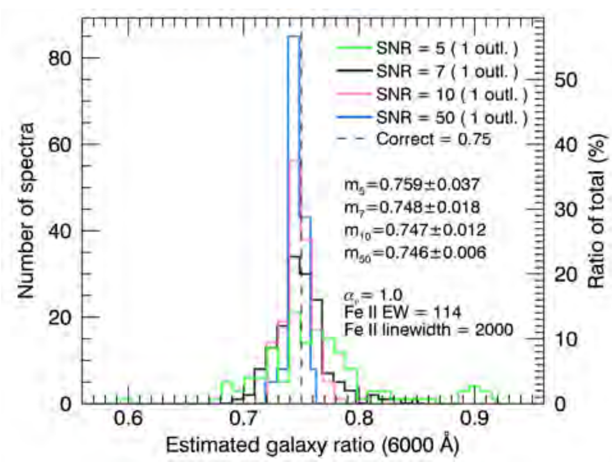
(b) Outliers: [0.000, none], [none, 0.875–0.912], [0.736, none], [0.743, none]



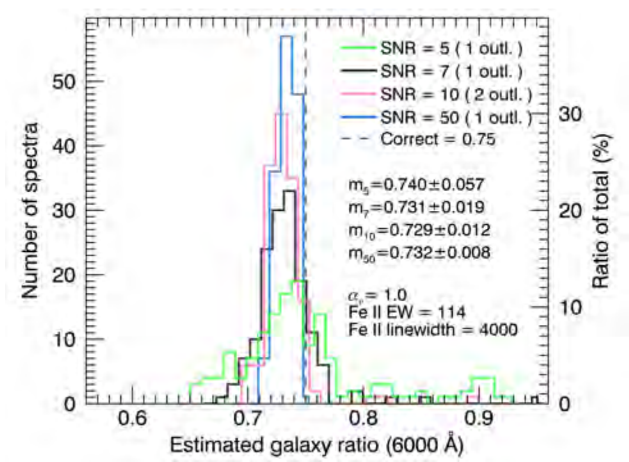
(c) Outliers: [0.443, none], [none, 0.890–0.929], [none, 0.831–0.897], [none, 0.753]



(d) Outliers: [none, 0.916], [none, 0.882–0.911], [0.750, none], [0.738, none]



(e) Outliers: [none, 0.907], [0.728, none], [0.749, none], [none, 0.753]



(f) Outliers: [0.734, none], [none, 0.950], [none, 0.821–0.893], [0.709, none]

Figure C.8

D Figures: Estimated power-law index

The figures on the following pages show how the estimated power-law index α_ν changes with input signal-to-noise ratio (S/N ratio) in the case of $\alpha_\nu = -0.5$ (steep in the UV) and 1.0 (flatter in the UV) as a function of Fe II EW for different galaxy ratios. The results for $\alpha_\nu = 0.5$ has been omitted as to not overwhelm the reader. The results for $\alpha_\nu = 0.5$ fall in-between those of the two other extremal values of α_ν and does not provide new insight.

For each value of α_ν, the figures shown have the same range on the x-axis so that the results can be compared directly. Fe II linewidth = 4000 $\mathrm{km\,s^{-1}}$ (top) and 2000 $\mathrm{km\,s^{-1}}$ (bottom).

A brief summary of the results: The power-law index α_ν is almost always *under*estimated, and more so for steep spectra. Note that the accuracy of the estimated value of α_ν does not changed with increased amount of iron when Fe II linewidth = 2000 $\mathrm{km\,s^{-1}}$, while the accuracy improves with increased amount of iron when Fe II linewidth = 4000 $\mathrm{km\,s^{-1}}$. For more details and comments, see the figure captions on the following pages and the summary in Section 5.1.2.

The figures begin on the following page.

Power-law index $\alpha_\nu = -0.5$

Galaxy ratio = 0.05

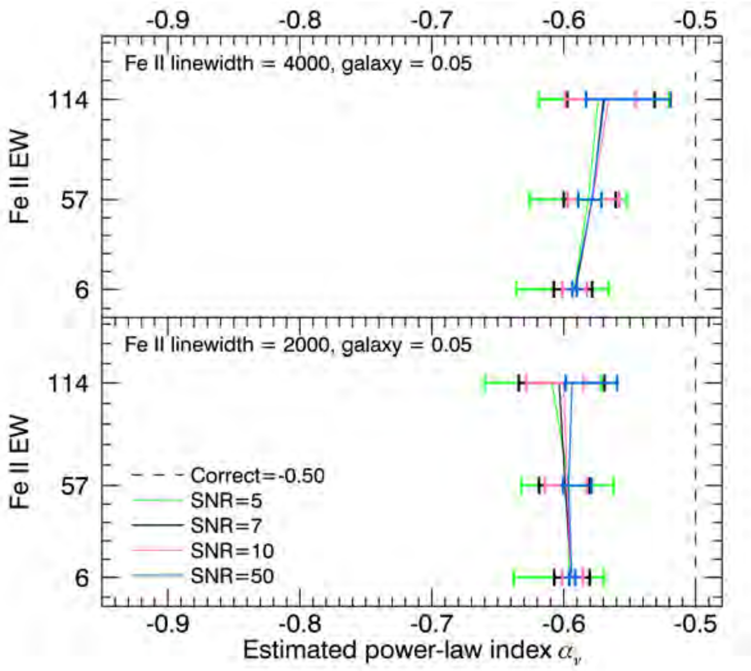


Figure D.1: The value of α_ν is always underestimated. *Top:* The accuracy of the estimated values improves with *in*creasing Fe II EW, while the precision *de*creases. The estimated values of α_ν are offset by [–0.092, –0.066] (underestimated by [–19, –14] %) relative to the true value. *Bottom:* The accuracy of the estimates is about the same for all S/N ratios. The precision *de*creases with *in*creased Fe II EW. The estimated values are offset by [–0.109, –0.094] (underestimated by [–22, –18] %) relative to the true value.

Galaxy ratio = 0.25

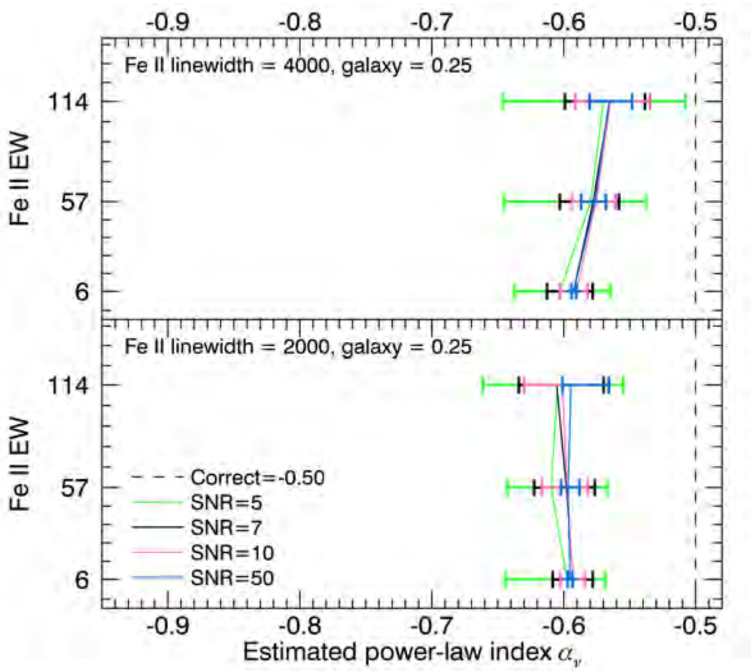


Figure D.2: The value of α_ν is always underestimated. *Top:* The accuracy of the estimated values improves with *in*creasing Fe II EW, while the precision *de*creases. The estimated values of α_ν are offset by [–0.103, –0.065] (underestimated by [–21, –13] %) relative to the true value. *Bottom:* The accuracy of the estimated values are about the same for all S/N ratios. The precision *de*creases with *in*creased Fe II EW. The estimated values are offset by [–0.110, –0.093] (underestimated by [–22, –18] %) relative to the true value.

Galaxy ratio = 0.50

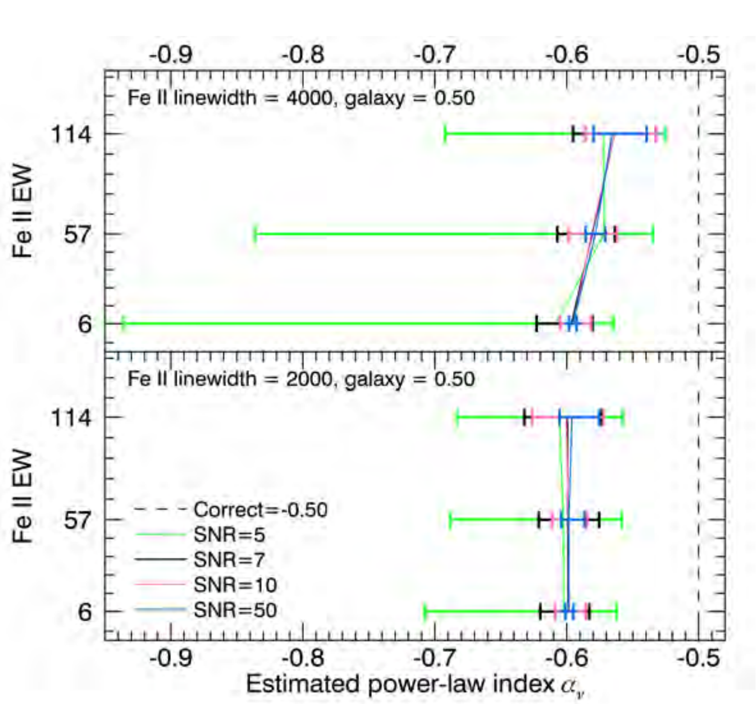


Figure D.3: The value of α_ν is always underestimated. *Top:* The accuracy of the estimated values improves with *in*creasing Fe II EW. The precision *de*creases with *in*creasing Fe II EW for all S/N ratios except S/N = 5, for which the opposite is the case. The estimated values are offset by [–0.109, –0.064] (underestimated by [–22, –12] %) relative to the true value. *Bottom:* The accuracy of the estimates are more or less constant with increased Fe II EW, with the estimates from S/N = 5 being shifted to slightly lower estimated values compared to the other S/N ratios. The estimated values are offset by [–0.105, –0.096] (underestimated by [–21, –19] %) relative to the true value.

Galaxy ratio = 0.75

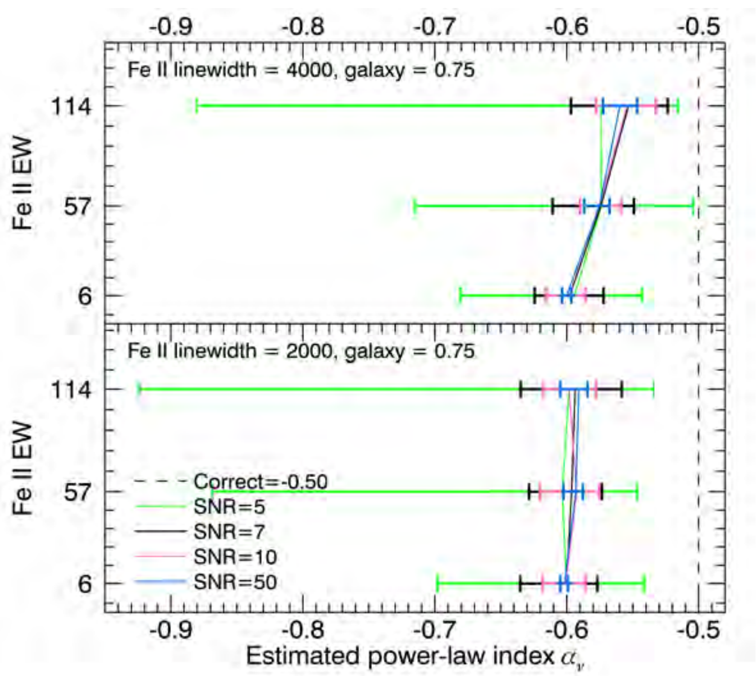


Figure D.4: The value of α_ν is always underestimated. *Top:* The accuracy of the estimates improves with *in*creasing Fe II EW, while the precision *de*creases for S/N = 10 and 50. The estimated values are offset by [–0.100, –0.053] (underestimated by [–20 ,–10] %) relative to the true value. *Bottom:* The accuracy of the estimated values are more or less constant with increased Fe II EW, while the precision *de*creases for S/N = 50. The estimated values are offset by [–0.103, –0.091] (underestimated by [–24 –18] %) relative to the true value.

Power-law index $\alpha_\nu = 1.0$

Galaxy ratio = 0.05

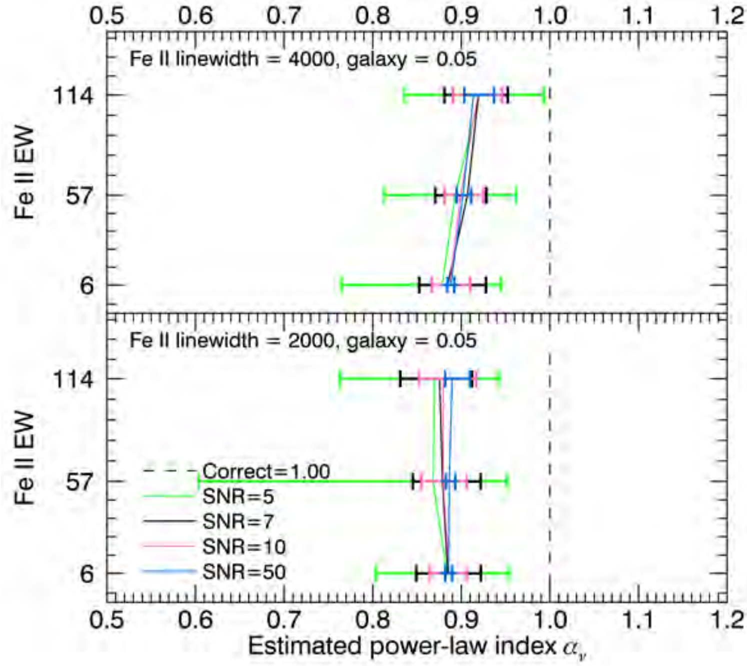


Figure D.5: The value of α_ν is always underestimated. *Top:* The accuracy of the estimates improves with *in*creasing Fe II EW, while the precision *de*creases for S/N = 50. The estimated values are offset by [–0.122, –0.079] (underestimated by [–12, –8] %) relative to the true value. *Bottom:* The accuracy of the estimates are more or less constant with increased Fe II EW, while the precision *de*creases for S/N = 50. The estimated values are offset by [–0.132, –0.110] (underestimated by [–13, –11] %) relative to the true value.

Galaxy ratio = 0.25

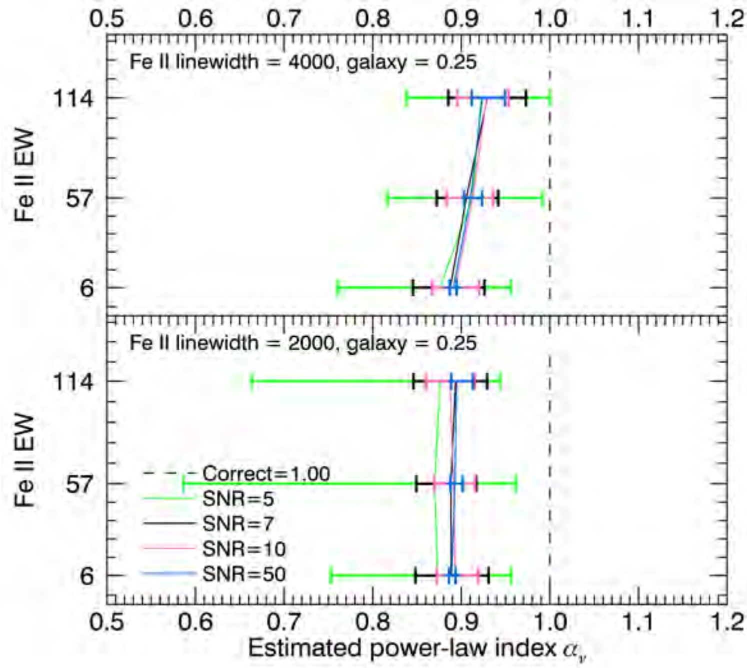


Figure D.6: The value of α_ν is always underestimated. *Top:* The accuracy of the estimates improve with *in*creasing Fe II EW, while the precision *de*creases for S/N = 50. The estimated values are offset by [–0.124, –0.070] (underestimated by [–13,–7] %) relative to the true value. *Bottom:* The accuracy of the estimates are more or less the same with increased Fe II EW, but the precision *de*creases for S/N = 50. The estimated values for S/N = 5 are noticeably shifted to lower values, while the other S/N ratios produce close to identical offsets relative to the true value. The estimated values are offset by [–0.130, –0.105] (underestimated by [–13,–10] %) relative to the true value.

Galaxy ratio = 0.50

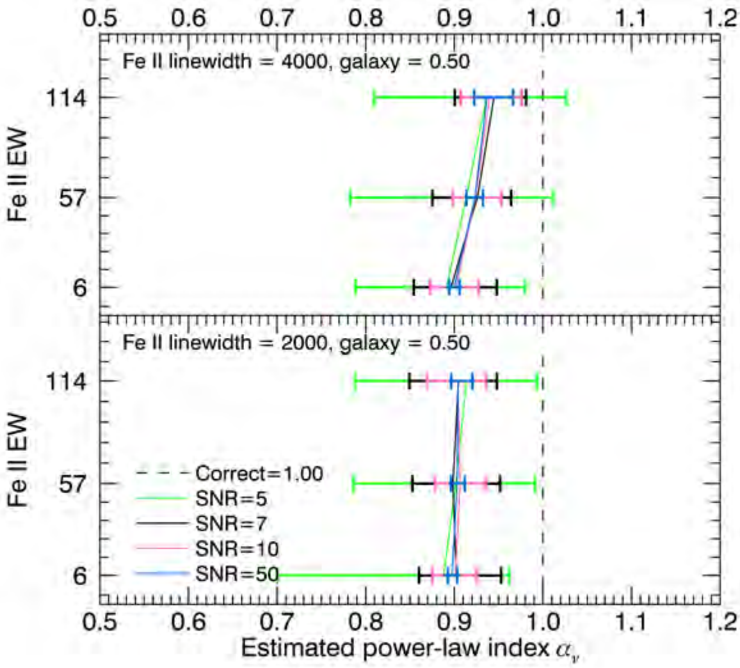


Figure D.7: The value of α_ν is always underestimated. *Top:* The accuracy of the estimates improve with *in*creasing Fe II EW, while the precision noticeably *de*creases for S/N = 50. The estimated values are offset by [−0.109, −0.055] (underestimated by [−11, −5.3] %) relative to the true value. *Bottom:* The accuracy of the estimated values are more or less the same with increased Fe II EW. The precision *de*creases with increased Fe II EW for S/N = 50. The estimated values are offset by [−0.112, −0.086] (underestimated by [−11, −8.5] %) relative to the true value.

Galaxy ratio = 0.75

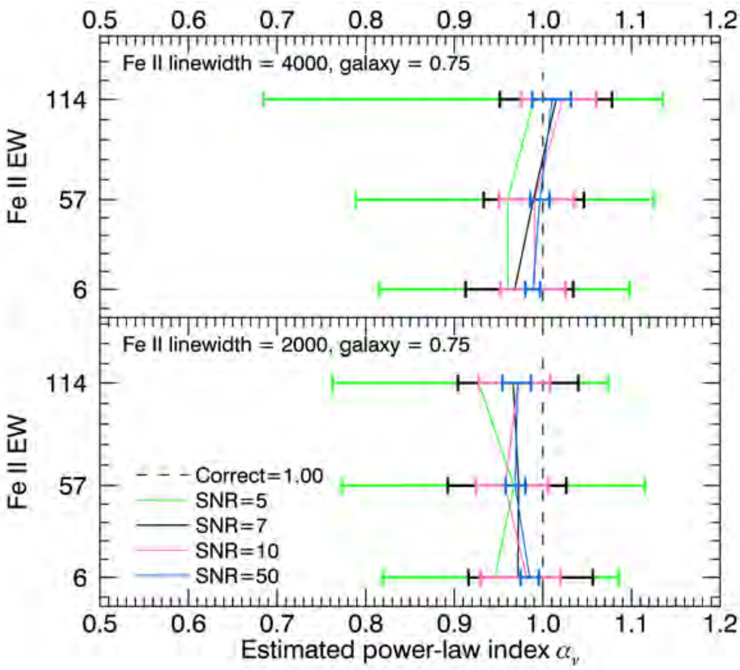


Figure D.8: *Top:* The values of the estimated power-law indices increase with increased Fe II EW. The precision *de*creases with *in*creased Fe II EW when S/N = 50. The power-law index tends to be underestimated for low Fe II EW and overestimated for high Fe II EW. The estimated values are offset by [−0.040, +0.022] (under- and overestimated by up to [−4, 2.5] %, respectively) relative to the true value. The accuracy for S/N = 10 and 50 are similar, while the estimated values for lower S/N ratios are shifted towards lower values. The true value is mostly within the uncertainties. *Bottom:* The offsets relative to the true value does not change much with Fe II EW except for S/N = 5 which is more underestimated for low and high Fe II EW. The true value is within the uncertainties for all S/N ratios except S/N = 50. The estimated values are offset by [−0.015, −0.072] (underestimated by [−5.5, −1.3] %) relative to the true value.

E Figures: Estimated Fe II EW

The figures on the following pages show how the estimated Fe II EW changes with input signal-to-noise ratio (S/N ratio) in the three different cases Fe II EW = 6, 57 and 114 as a function of α_ν for different galaxy ratios.

For each value of Fe II EW all the figures have the same range on the x-axis so that the results can be compared directly. Fe II linewidth = 4000 km s^{-1} (top) and 2000 km s^{-1} (bottom).

A brief summary of the results: The iron content in the spectra is almost always estimated to be zero when Fe II EW = 6 Å. When Fe II EW = 57 Å and 114 Å, the iron content in the spectra tends to be *under*estimated for Fe II linewidth of 4000 km s^{-1} (broad). For Fe II linewidth of 2000 km s^{-1} (narrow) the iron content is *over*estimated when the galaxy ratio is low and *under*estimated when the galaxy ratio is high. The fact that the iron emission goes from being overestimated to underestimated with higher galaxy ratios indicates that the strength of the iron emission relative to the galaxy ratio is important for the results. Whether or not the true value of Fe II EW is within the uncertainties strongly depends on the combination of parameters being considered.

For more details and comments, see the figure captions on the following pages and the summary in Section 5.1.2.

The figures begin on the following page.

Fe II *EW = 6*

Galaxy ratio = 0.05

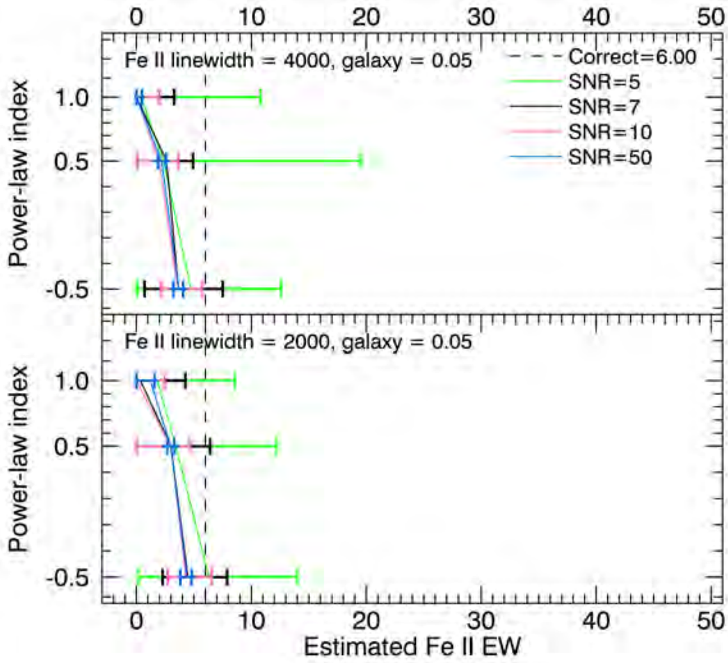


Figure E.1: The accuracy of the estimated Fe II EW generally *de*creases with *in*creasing α_ν (flatter in the UV). The scatter increases with decreasing S/N ratios. The true value is within the uncertainties for S/N = 5, and when S/N = 7 for lower α_ν (steeper). The accuracy is similar for all S/N ratios for a given combination of parameters, but higher in some cases for the estimates when S/N = 5. *Top:* The distributions are offset by [–100, –18] % relative to the true value. *Bottom:* The distributions are offset by [–100, 8.3] % relative to the true value.

Galaxy ratio = 0.25

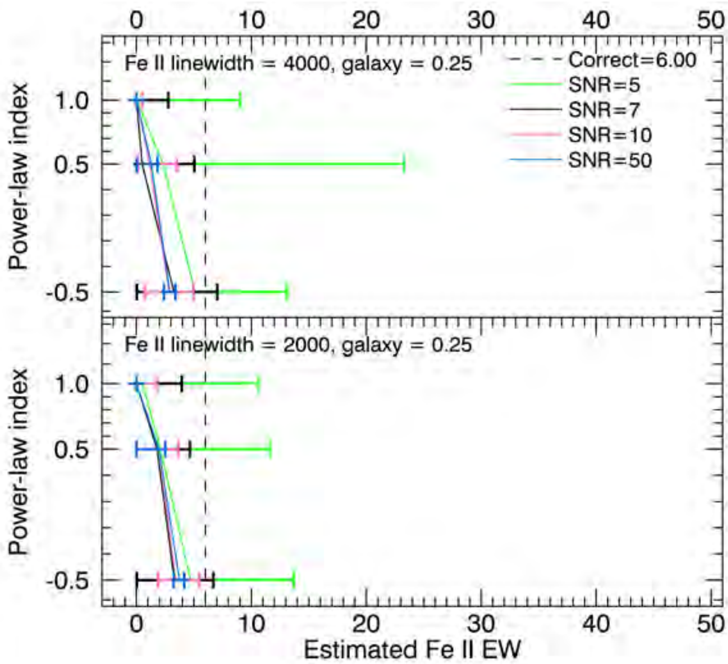


Figure E.2: The accuracy of the estimated Fe II EWs *de*creases with *in*creasing (flatter) α_ν. The scatter increases with decreasing S/N ratios. The true value is within the uncertainties for S/N = 5, and when S/N = 7 for $\alpha_\nu = -0.5$ (steepest). The accuracy is similar for all S/N ratios for a given combination of parameters, but higher in some cases for the estimates when S/N = 5. *Top:* The distributions of estimated Fe II EWs are offset by [–100, –17] % relative to the true value. *Bottom:* The distributions are offset by [–100, –18] % relative to the true value.

Galaxy ratio = 0.50

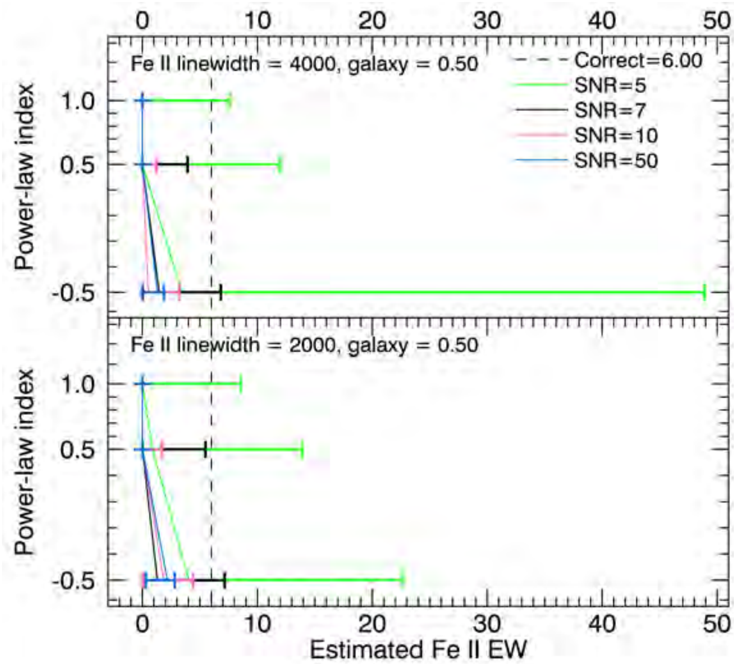


Figure E.3: The accuracy of the estimated Fe II EWs *de*creases with *in*creasing α_ν (flatter in the UV). The scatter increases with decreasing S/N ratios. The true value is within the uncertainties for S/N = 5, and when S/N = 7 for $\alpha_\nu = -0.5$ (steepest). The accuracy is similar for all S/N ratios for a given combination of parameters, but smaller for S/N = 5 for lower α_ν (steeper). In both cases (top & bottom) the distributions of the estimated Fe II EWs are offset by $[-100, -41]$ % relative to the true value.

Galaxy ratio = 0.75

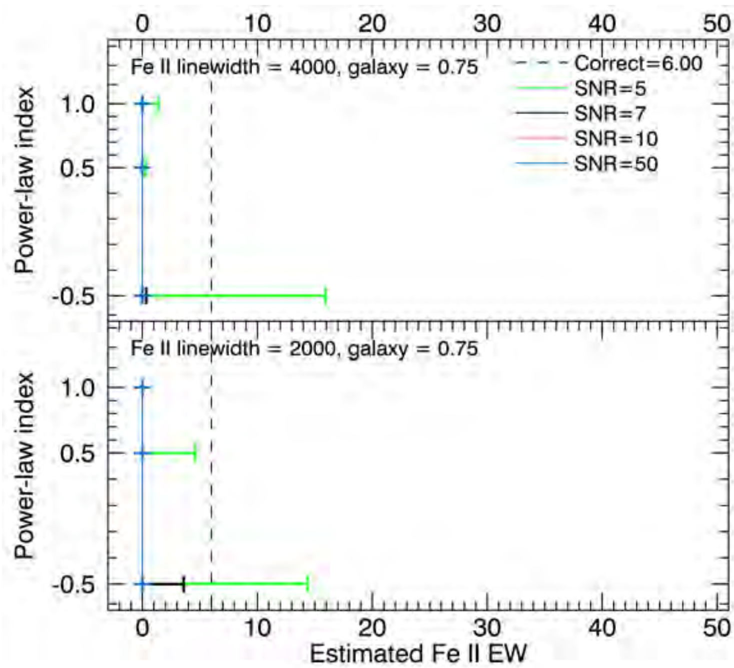


Figure E.4: The offsets of the distributions of estimated Fe II EWs are in all cases -100 % relative to the true value. The scatter is very large when S/N = 5 and $\alpha_\nu = -0.5$. It is possible to recover the true value when S/N = 5 because of the large scatter, but not for the other S/N ratios.

Fe II *EW = 57*

Galaxy ratio = 0.05

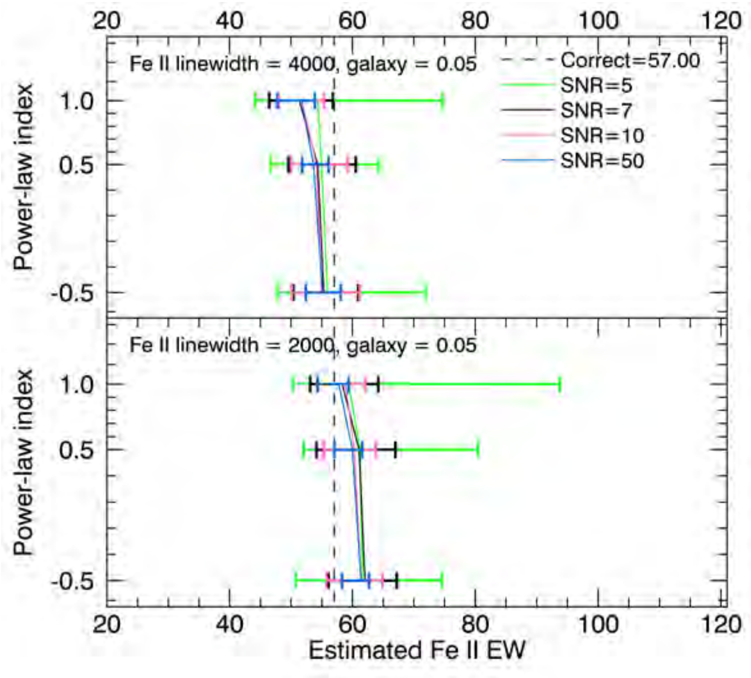


Figure E.5: The estimated values of Fe II EW are generally lower with higher α_ν (flatter in the UV). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by $[-11, -1.7]\,\%$ relative to the true value. *Bottom:* The Fe II EW tends to be *over*estimated, with a noticeably decreased precision for higher α_ν when $\mathrm{S/N} = 5$. The distributions are offset by $[1, 9]\,\%$ relative to the true value.

Galaxy ratio = 0.25

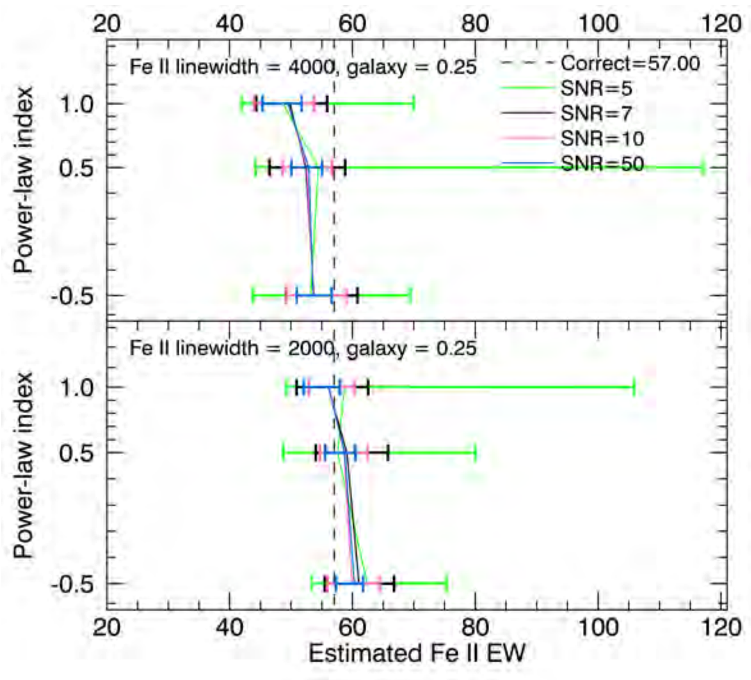


Figure E.6: The estimated values of Fe II EW are generally lower with higher α_ν (flatter in the UV). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The scatter is extremely large for $\alpha_\nu = 0.5$ when $\mathrm{S/N} = 5$. The distributions are offset by $[-14, -3.5]\,\%$ relative to the true value. *Bottom:* The Fe II EW tends to be *over*estimated, with noticeably decreased precision for higher α_ν when $\mathrm{S/N} = 5$. The distributions are offset by $[-1.7, 11]\,\%$ relative to the true value.

Galaxy ratio = 0.50

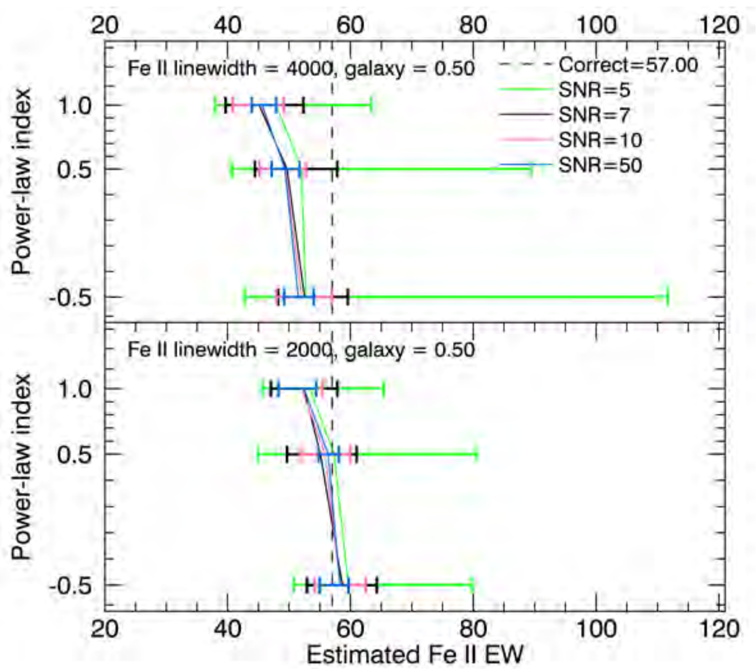


Figure E.7: The estimated values of Fe II EW are generally lower with higher α_ν (flatter). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by [–21, –8.8] % relative to the true value. The precision dramatically decreased with decreasing α_ν (steeper) when S/N = 5. *Bottom:* The Fe II EW tends to be *under*estimated for high (flat) α_ν and *over*estimated for low (steep) α_ν. The distributions are offset by [–8.8, 5.3] % relative to the true value.

Galaxy ratio = 0.75

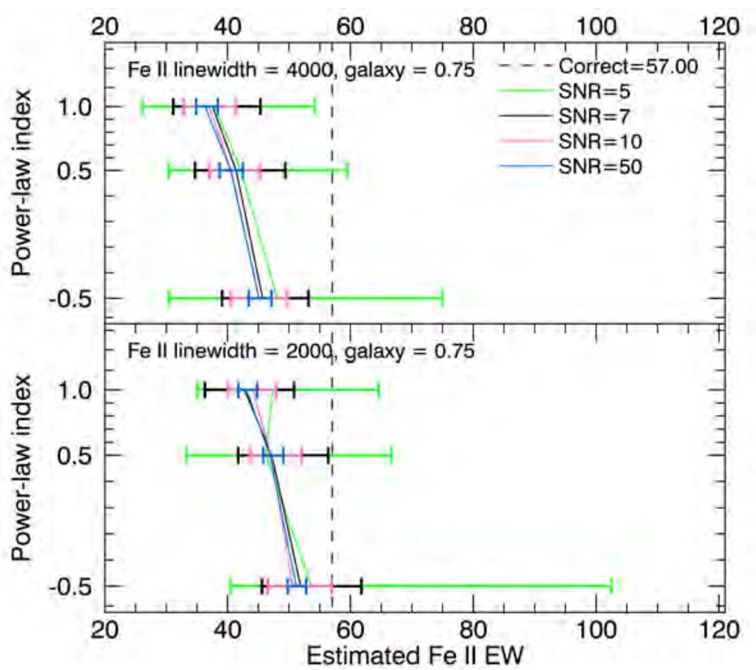


Figure E.8: The estimated values of Fe II EW tends to be underestimated, and are generally more underestimated for higher α_ν (flatter). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by [–37, –15] % relative to the true value. The precision decreased with lower α_ν (steeper) when S/N = 5, with the true value within its uncertainties. *Bottom:* The Fe II EW tends to be *over*estimated. The distributions are offset by [–25, –5.2] % relative to the true value. The true value is within the uncertainties for S/N = 5, and for S/N = 7 when $\alpha_\nu = –0.5$.

Fe II EW = 114

Galaxy ratio = 0.05

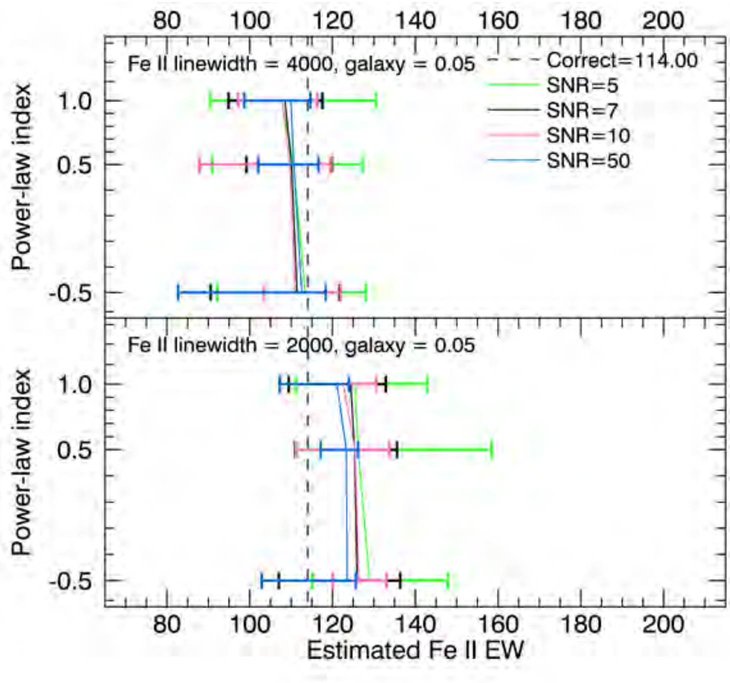


Figure E.9: The estimated values of Fe II EW are generally slightly lower with higher α_ν (flatter). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by [–5.3, –0.8] % relative to the true value. *Bottom:* The Fe II EW is always *over*estimated, with S/N = 50 giving the *most* accurate results and S/N = 5 giving the *least* accurate results. The distributions are offset by [6.1, 13] % relative to the true value.

Galaxy ratio = 0.25

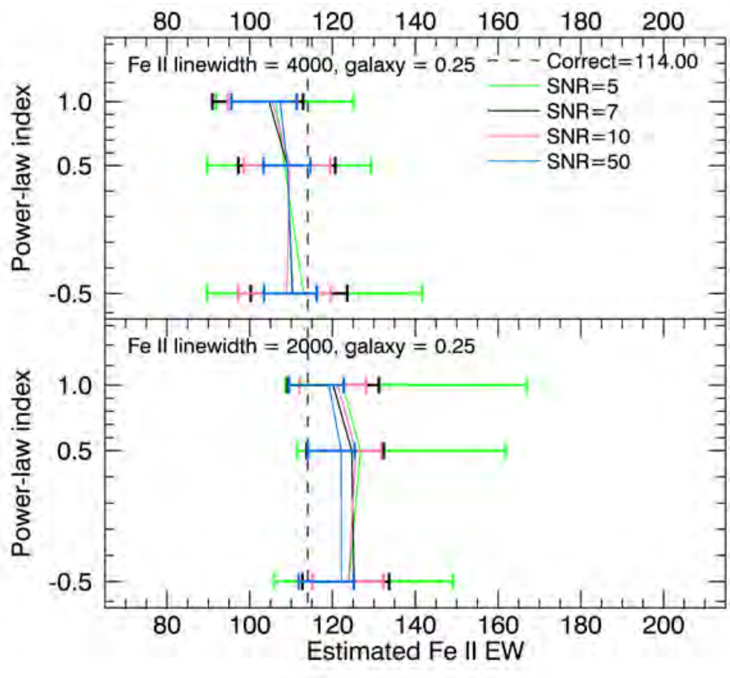


Figure E.10: The estimated values of Fe II EW are generally slightly lower with higher α_ν (flatter). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by [–8, –0.8] % relative to the true value. *Bottom:* The Fe II EW is always *over*estimated, with S/N = 50 giving the most accurate results. The distributions are offset by [4.3, 11] % relative to the true value.

Galaxy ratio = 0.50

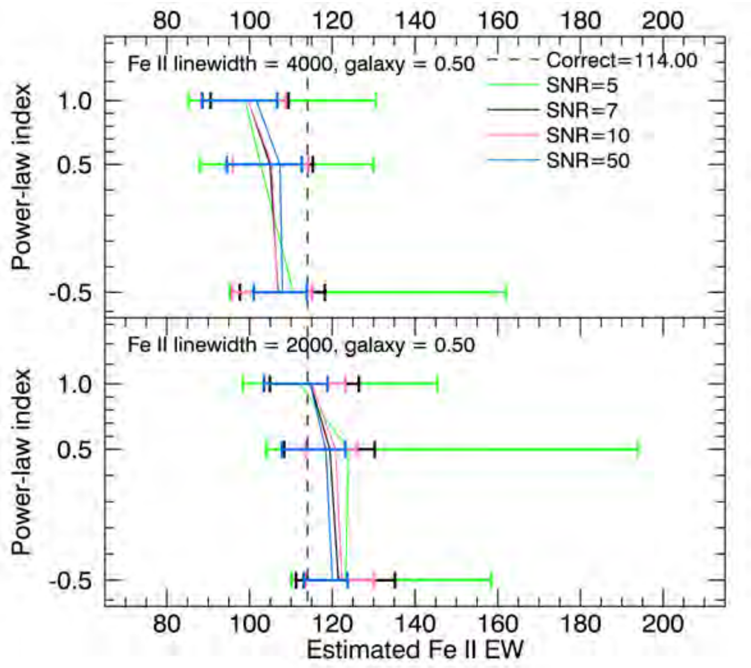


Figure E.11: The estimated values of Fe II EW are generally slightly lower with higher α_ν (flatter). The scatter increases with decreasing S/N ratios. *Top:* The Fe II EW tends to be *under*estimated. The distributions are offset by $[-13, -3.5]\,\%$ relative to the true value. *Bottom:* The Fe II EW tends to be *over*estimated. The distributions are offset by $[-1.8, 10]\,\%$ relative to the true value.

Galaxy ratio = 0.75

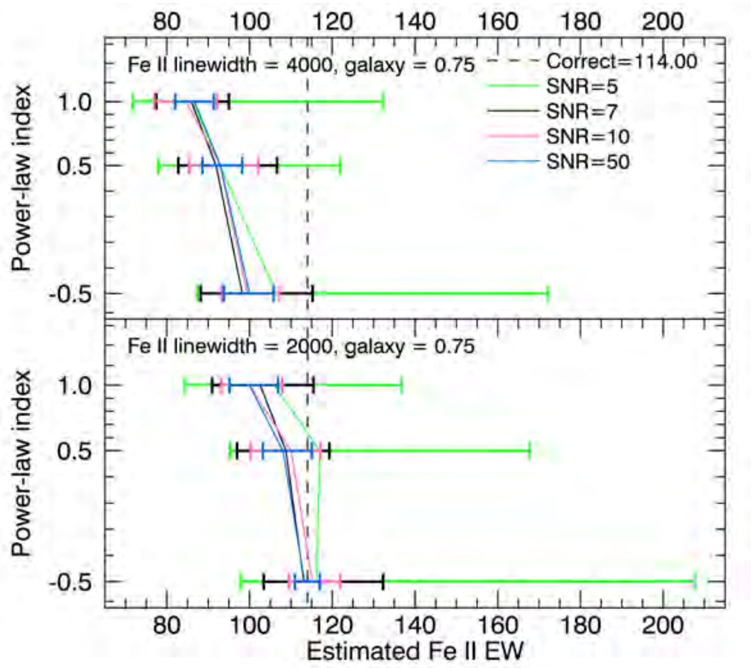


Figure E.12: The estimated values of Fe II EW are generally slightly lower with higher α_ν (flatter). The scatter increases with decreasing S/N ratios. The Fe II EW tends to be *under*estimated. *Top:* The distributions are offset by $[-26, -6.1]\,\%$ relative to the true value. *Bottom:* The distributions are offset by $[-13, 3.5]\,\%$ relative to the true value.

F Figures: Estimated signal-to-noise ratio

The signal-to-noise ratio (S/N ratio) has been measured in four different wavelength ranges, illustrated in Figure F.1, in order to assess whether there is some wavelength range that is preferable for doing such a measurement and whether the preferable range depends on the spectrum at hand.

There is not a big difference from one set of parameters to another, so only plots of the extreme values of α_ν and Fe II EW are included here. In the following figures: $\alpha_\nu = -0.5$ (left) and $\alpha_\nu = 1.0$ (right), with Fe II linewidth = 4000 km s^{-1} (top) and 2000 km s^{-1} (bottom). Each section looks at one S/N ratio, first Fe II EW = 6 Å and then Fe II EW = 114 Å. Fe II EW = 57 has been omitted because it does not provide new insight. The same goes for S/N = 10, which produces results very similar to S/N = 5.

A brief summary of the results: When S/N = 5 all the four ranges are doing approximately equally well. When S/N = 50 the green range (see Figure F.1 for illustration), and sometimes the pink range, stands out as highly underestimating the S/N ratio. The blue and red range does best overall, with the red range being slightly more precise, thus being the superior range of the ones considered here.

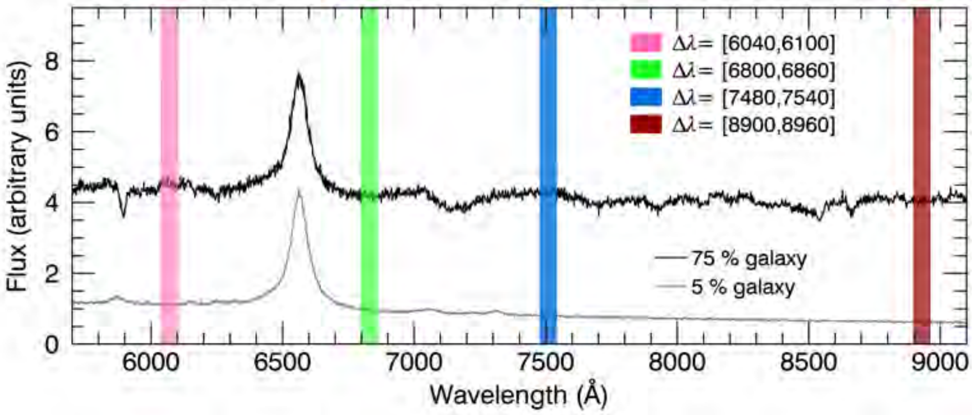


Figure F.1: The S/N ratio was measured for four different wavelength ranges of the mock spectra that are approximately flat. By comparing the black and grey curve representing a mock spectrum with 75 % and 5 % galaxy ratios, respectively, it is evident that the galaxy adds a lot of small-scale features to the spectrum. The pronounced peak is Hα, while the absorption feature leftward of 6000 Å comes from the galaxy template. Here S/N = 50, α_ν = ,0.5, Fe II EW = 57 Å, Fe II linewidth = 2000 km s^{-1}.

F.1 Signal-to-noise ratio = 5

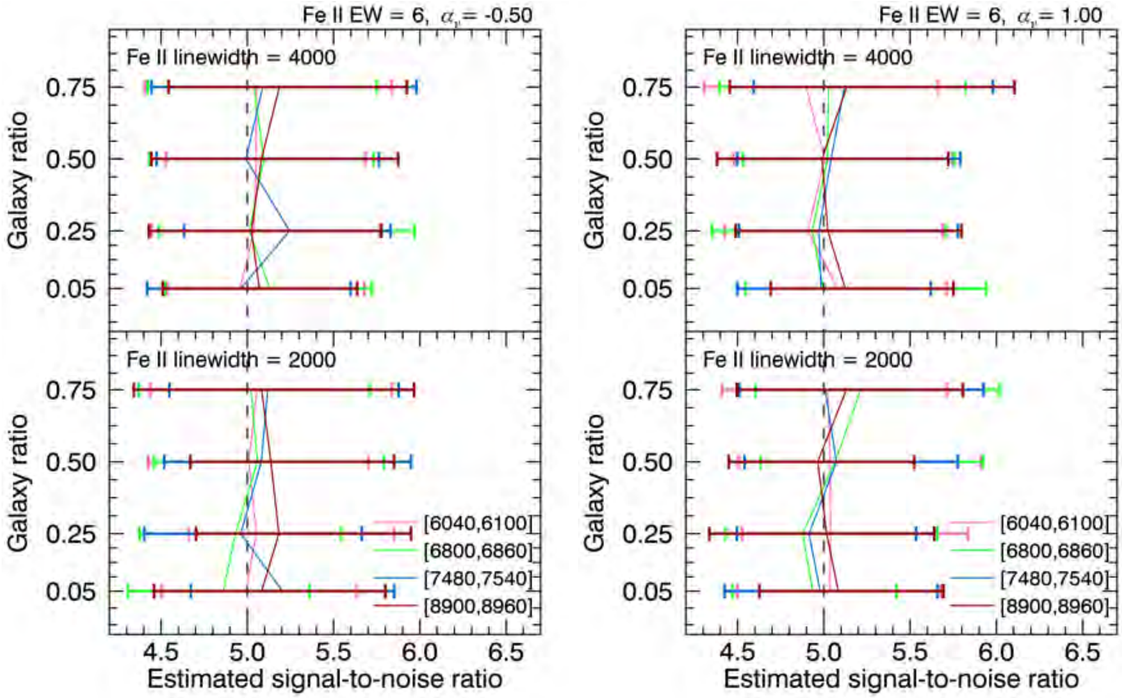


Figure F.2: Here Fe II EW = 6 Å. All the distributions are offset with less than 5 % relative to the true value. The scatter in estimated values tends to be slightly smaller when the galaxy ratio = 0.05, but there is no obvious difference between the four ranges in the four situations considered here.

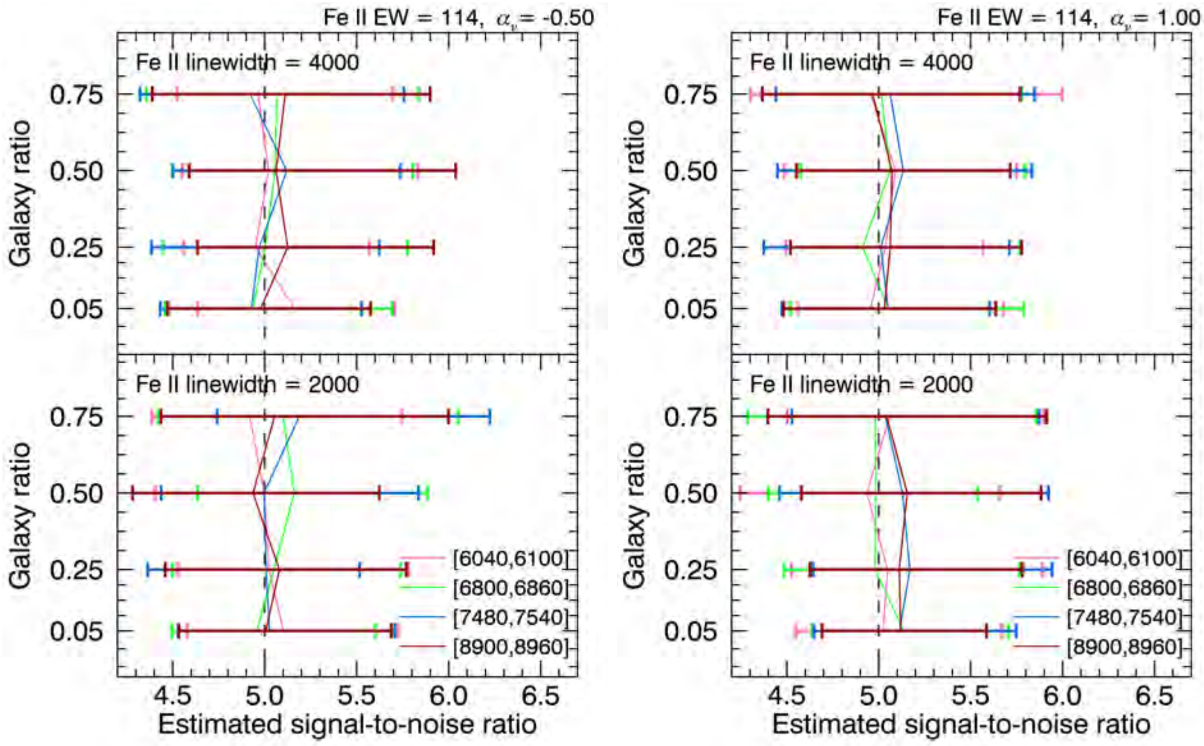


Figure F.3: Here Fe II EW = 114 Å. All the distributions are offset with less than 5 % relative to the true value. There are in a few cases slightly larger scatter than in the above cases.

F.2 Signal-to-noise ratio = 50

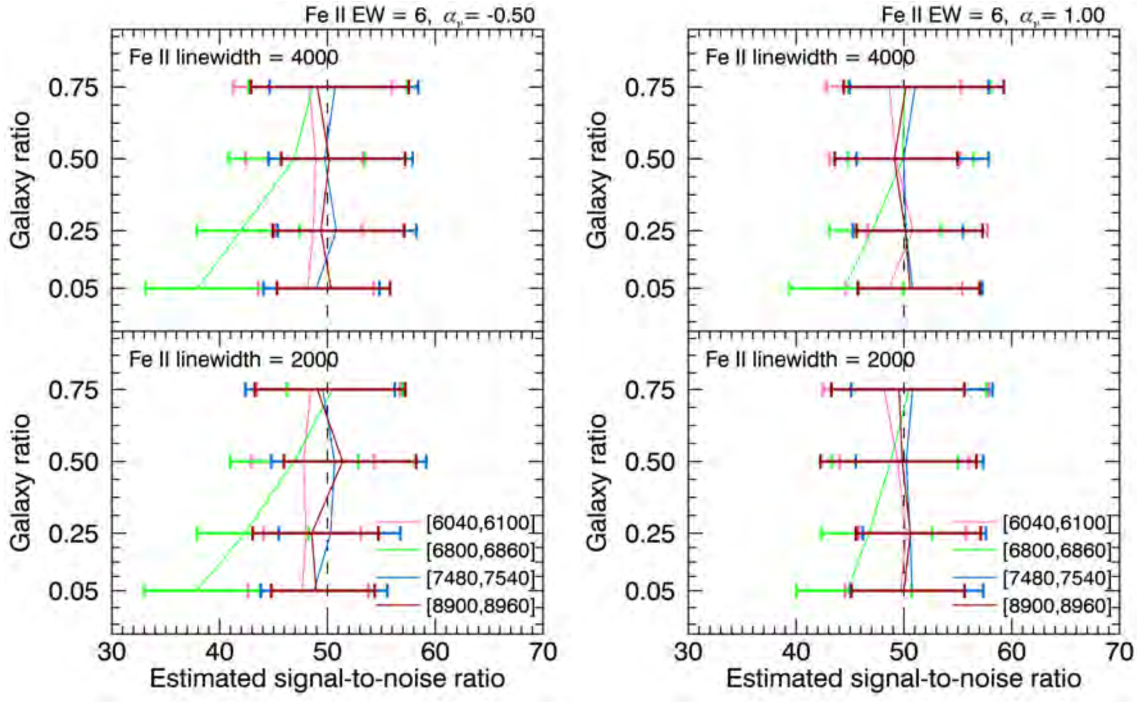


Figure F.4: Fe II EW = 6 Å. The green range is highly underestimating the S/N ratio and more so for lower galaxy ratios and the lower α_ν (left; steeper in the UV). The pink range also tends to slightly underestimate the S/N ratio for $\alpha_\nu = -0.5$. The red and blue range does more or less equally well. The red range tends to have a smaller spread of the two.

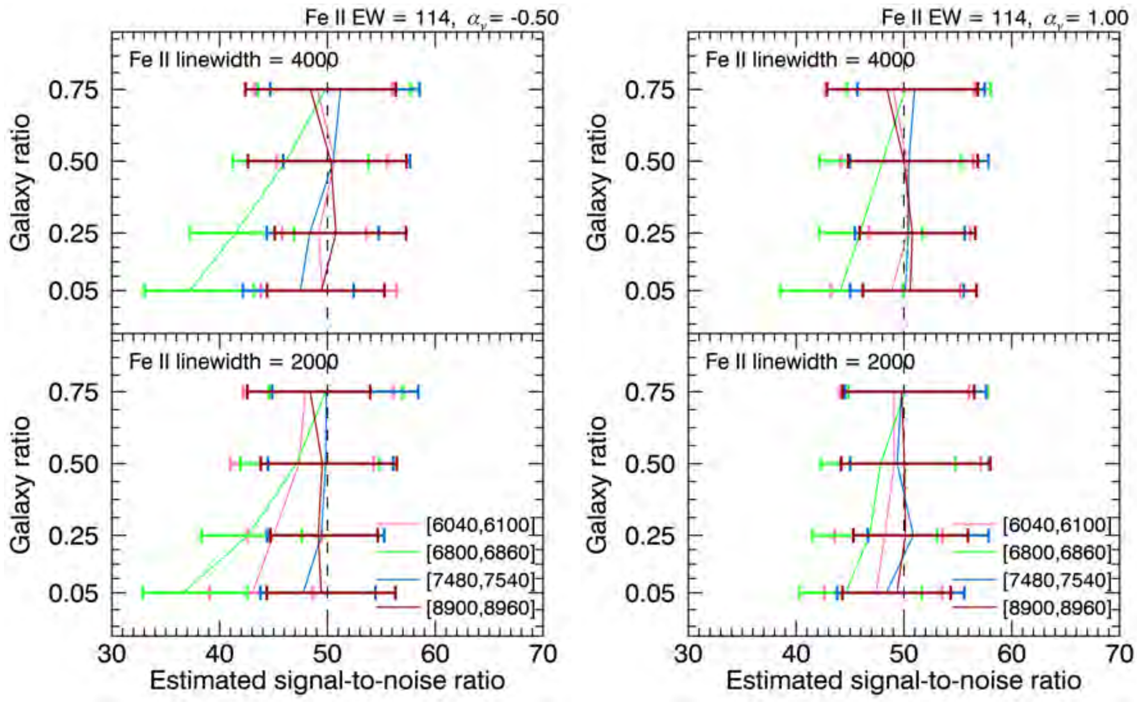


Figure F.5: Fe II EW = 114 Å. The results are similar to the ones in the above figure.

G Figures: Correlations

G.1 Galaxy ratio v. Power-law slope

The correlation coefficient between the estimated galaxy ratio and the estimated power-law index α_ν is shown for different Fe II EW = 6 (top), 57 (middle), 114 (bottom), and for Fe II linewidth = 2000 km s^{-1} (left) and 4000 km s^{-1} (right).

Main result: The estimated galaxy ratio and the estimated power-law index α_ν are highly anti-correlated, mostly in the range $-0.9 \gtrsim r_s > -1.0$.

Signal-to-noise ratio = 5

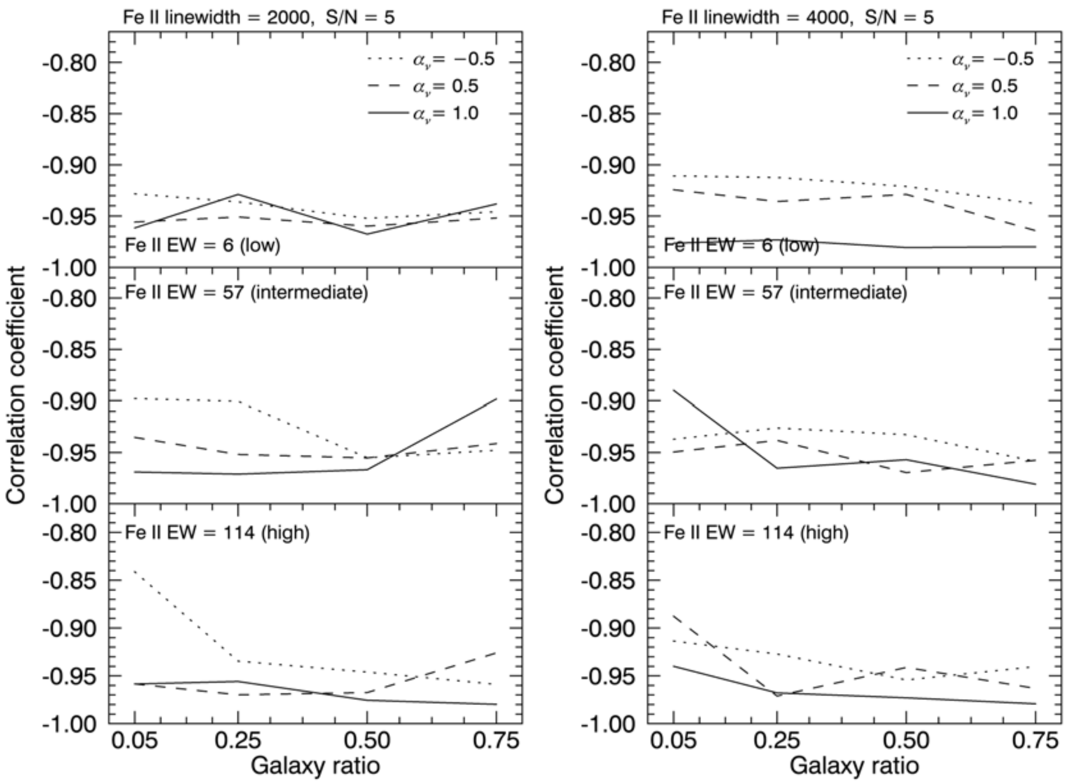


Figure G.1: The estimated galaxy ratio and the estimated power-law index α_ν are highly anti-correlated, mostly in the range $-0.9 \gtrsim r_s > -1.0$. The anti-correlation tends to be stronger for *flatter* spectra (higher α_ν) than for steep spectra, and it also tends to become slightly stronger for higher galaxy ratios, with some exceptions when Fe II linewidth = 2000 km s^{-1} (left).

Signal-to-noise ratio = 10

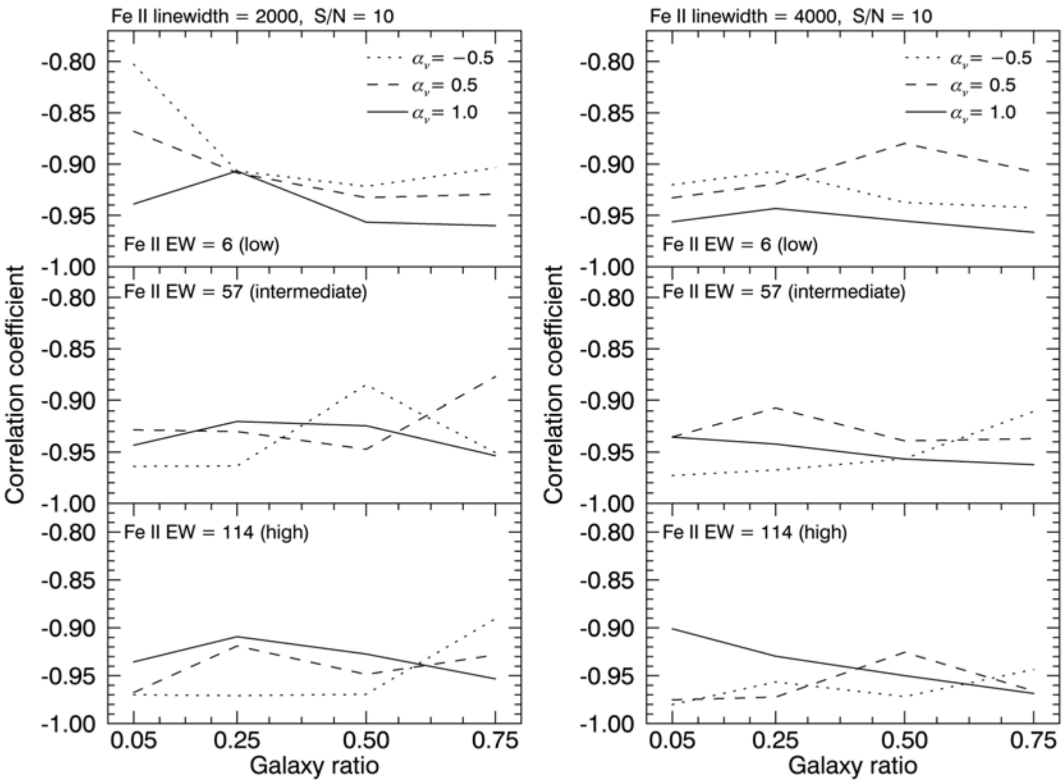


Figure G.2: The estimated galaxy ratio and the estimated power-law index α_ν are highly anti-correlated, mostly in the range $-0.9 \gtrsim r_s > -1.0$. The *flattest* spectra ($\alpha_\nu = 1.0$) show a stronger anti-correlation for higher galaxy ratios for all Fe II EW, with the anti-correlation often being weakest when the galaxy ratio is around 0.25, indicating that the relative strength of the spectral components is important at this galaxy ratio. The *steeper* spectra ($\alpha_\nu = 0.5$ and –0.5) tends to have a stronger anti-correlation with higher galaxy ratios when Fe II EW is low, while the anti-correlation tends to become weaker for higher galaxy ratios when Fe II EW is higher.

Signal-to-noise ratio = 50

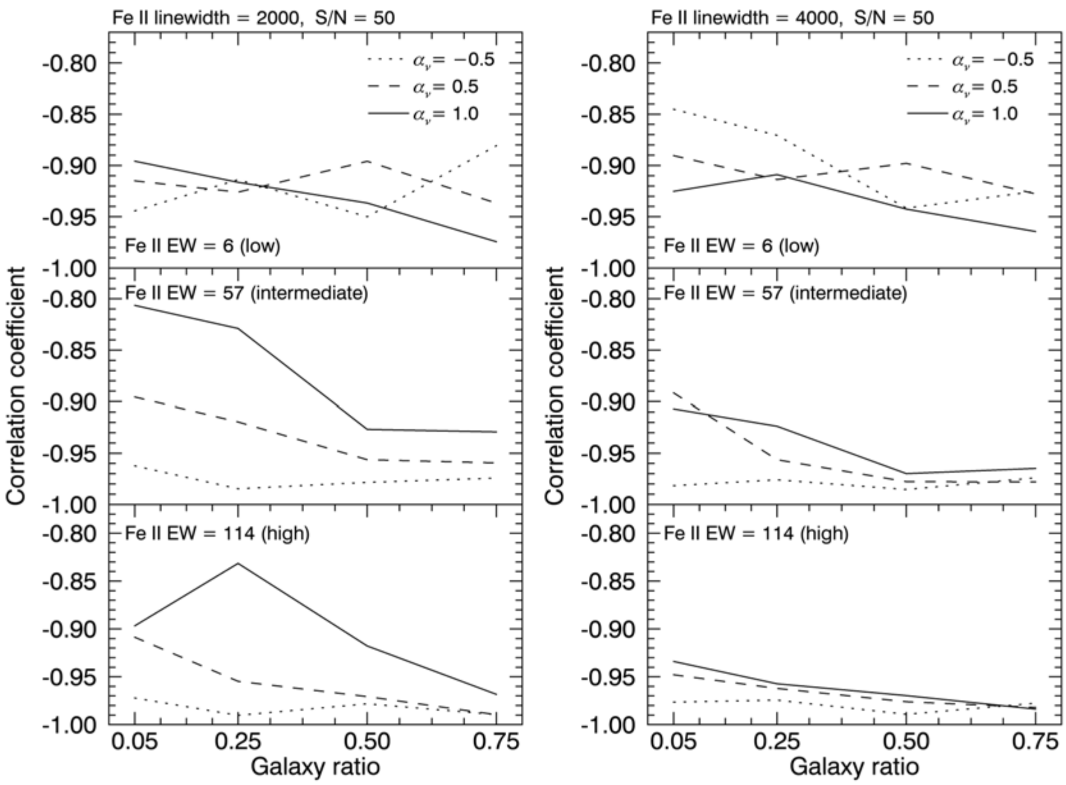


Figure G.3: When there is an *intermediate to high* amount of iron (Fe II EW = 57 Å and 114 Å) in the spectra the anti-correlation between the estimated galaxy ratio and the estimated power-law index α_ν tends to become stronger with higher galaxy ratios. The *steeper* spectra tends to have a stronger anti-correlation than the flatter ones, and are consistently found in the range $-0.9 \gtrsim r_s > -1.0$, while the *flatter* spectra can go up to $r_s \lesssim -0.8$, which is still a strong anti-correlation. The anti-correlation tends to be stronger for Fe II linewidth = 4000 km s^{-1} (right) than Fe II linewidth = 2000 km s^{-1} (left).

When Fe II EW = 6 Å, the strength of the anti-correlation fluctuates between the various galaxy ratios, but the anti-correlation is mostly stronger than $r_s \lesssim -0.9$.

G.2 Galaxy ratio v. Fe II EW

The correlation coefficient between the estimated galaxy ratio and the estimated Fe II EW is shown for power-law index $\alpha_\nu = -0.5$ (top), 0.5 (middle), 1.0 (bottom), and for Fe II linewidth = 2000 km s^{-1} (left) and 4000 km s^{-1} (right).

In some of the figures data is missing for Fe II EW = 6. This is because Fe II EW is estimated to be zero for all the spectra for these particular sets of parameters, and the calculation of the correlation coefficient becomes `NaN` due to division by zero when there is no variance in the Fe II EW estimates.

Main result: The estimated galaxy ratio is correlated with the estimated value of Fe II EW, with the correlation coefficient mostly in the range $0.6 \lesssim r_s \lesssim 0.8$.

Signal-to-noise ratio = 5

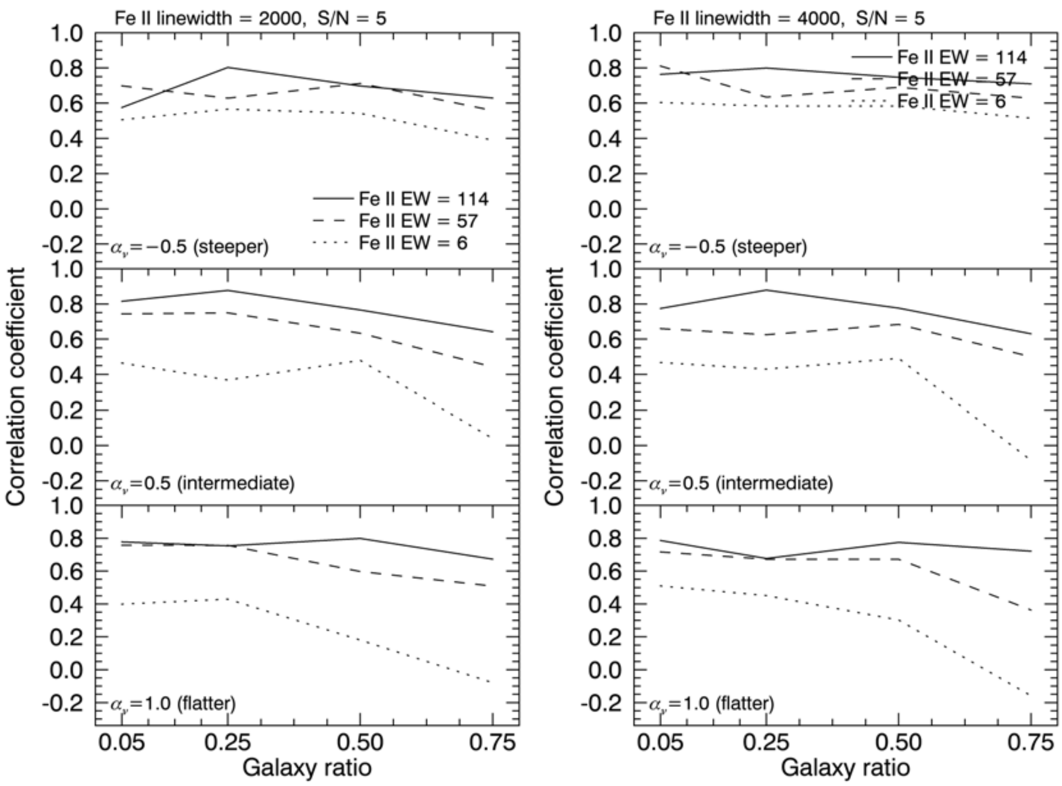


Figure G.4: The estimated galaxy ratio is correlated with the estimated amount of iron in the spectra (described by Fe II EW), with the correlation coefficient mostly in the range $0.6 \lesssim r_s \lesssim 0.8$. The correlation tends to become weaker for higher galaxy ratios, but does not change much with the power-law index α_ν. For Fe II EW = 6 Å the correlation is dramatically weaker for higher galaxy ratios, and it is also weaker for flatter spectra than for steep spectra. There is not an obvious difference between Fe II linewidth = 2000 km s^{-1} (left) and Fe II linewidth = 4000 km s^{-1} (right).

Signal-to-noise ratio = 10

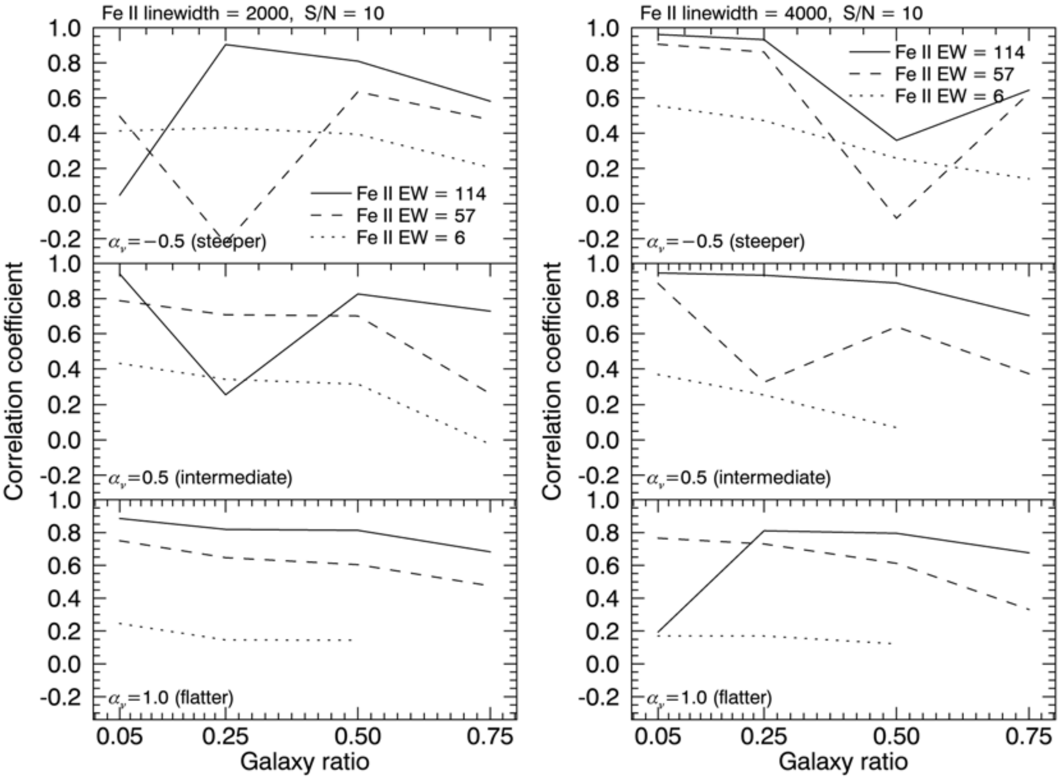


Figure G.5: The estimated galaxy ratio is correlated with the estimated value of Fe II EW. The correlation coefficient tends to be weak when Fe II EW is small and strong when Fe II EW is large. The correlation coefficient greatly fluctuates with the galaxy ratio for the *steeper* spectra ($\alpha_\nu = -0.5$), with noticeable dips or peaks in the correlation coefficient for galaxy ratio = 0.25 and 0.50, depending on the power-law index α_ν and the amount of iron. For *flatter* spectra the correlation coefficient tends to be weaker for higher galaxy ratios, with a noticeable exception when Fe II linewidth=4000 km s^{-1} and galaxy ratio = 0.05 where the correlation coefficient is very low before it peaks for galaxy ratio = 0.25. For Fe II EW = 6 Å the correlation is weaker for flatter spectra than for steep spectra, and the correlation weakens for higher galaxy ratios.

Signal-to-noise ratio = 50

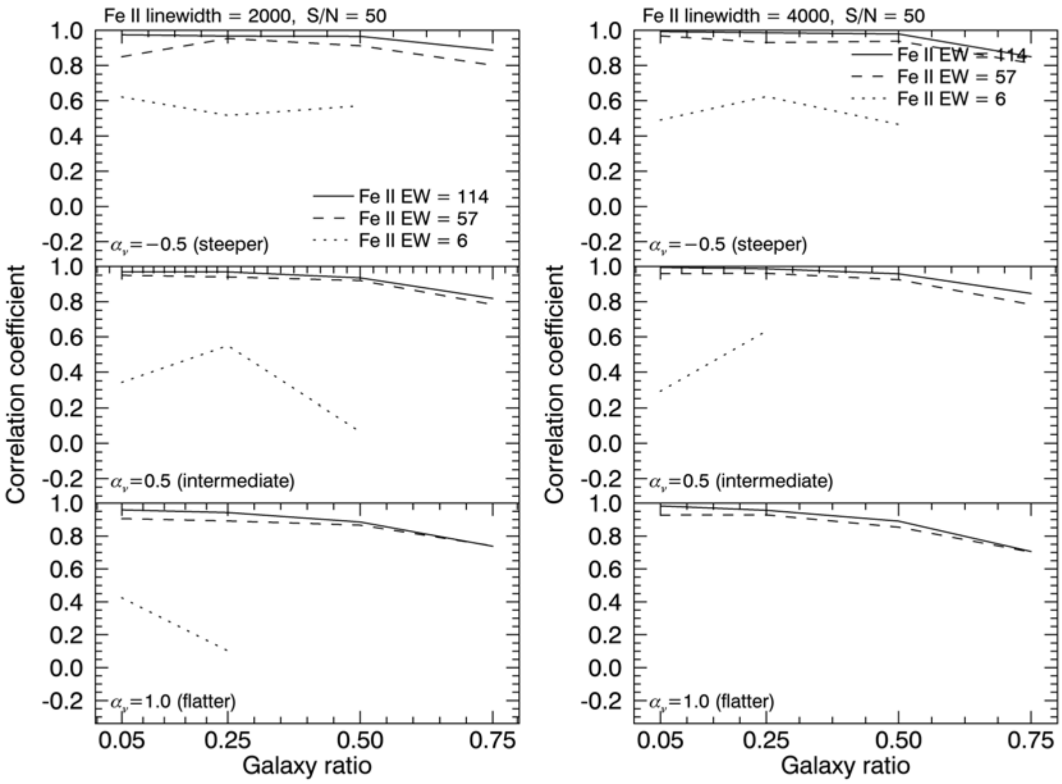


Figure G.6: The estimated galaxy ratio is correlated with the estimated Fe II EW. The correlation is very strong when there is a large amount of iron in the spectra – mostly in the range $0.8 < r_s < 1.0$ – and strongest when Fe II EW = 114 Å. The correlation coefficient hardly changes with the power-law index α_ν, but the correlation tends to be weaker for higher galaxy ratios. When Fe II EW = 6 Å the correlation coefficient is moderate to very low in the cases where the correlation coefficient can be computed. There is not an obvious difference between Fe II linewidth = 2000 km s^{-1} (left) and Fe II linewidth = 4000 km s^{-1} (right).

List of Figures

List of Tables

References

Barth, A.J. et al. "The Lick Monitoring Project 2011: Spectroscopic Campaign and Emission-Line Curves". In: *The Astrophysical Journal Supplement Series* 217 (2015).

Bentz, M.C. et al. "The Radius-Luminosity Relationship for Active Galactic Nuclei: The Effect of Host-Galaxy Starlight on Luminosity Measurements". In: *The Astrophysical Journal* 644 (2006), pp. 133–142.

Bentz, M.C. et al. "The Radius-Luminosity Relationship for Active Galactic Nuclei: The Effect of Host-Galaxy Starlight on Luminosity Measurements. II. The Full Sample of Reverberation-Mapped AGNs". In: *The Astrophysical Journal* 697.1 (2009), pp. 160–181.

Blandford, R.D. and McKee, C.F. "Reverberation Mapping of the Emission Line Regions of Seyfert Galaxies and Quasars". In: *The Astrophysical Journal* 255.1 (1982), pp. 419–439.

Blandford, R.D. and Reese, M.J. "The standard model and some new directions". In: *AIP Conference Proceedings* 254.3 (1992).

Borosen, T.A. and Green, R.F. "The Emission-Line Properties of Low-Redshift Quasi-Stellar Objects". In: *The Astrophysical Journal Supplement Series* 80 (1992), pp. 109–135.

Ciesla, L. et al. "Constraining the properties of AGN host galaxies with Spectral Energy Distribution modelling". In: *Astronomy & Astrophysics* 576 (2015).

Denney, K.D. et al. "Systematic Uncertainties in Black Hole Mass Determined from Single-Epoch Spectra". In: *The Astrophysical Journal* 692 (2009), p. 246.

Dietrich, M. et al. "Continuum and Emission-line Strength Relations for a Large Active Galactic Nuclei Sample". In: *The Astrophysical Journal* 581 (2002), pp. 912–924.

Dietrich, M. et al. "High-Redshift Quasars and Star Formation in the Early Universe". In: *The Astrophysical Journal* 564 (2002), pp. 581–591.

Fan, X. et al. "High-Redshift Quasars Found in Sloan Digital Sky Survey Commissioning Data". In: *The Astronomical Journal* 118.1 (1999), pp. 1–13.

Ferrarese, L. and Ford, H. "Supermassive Black Holes in Galactic Nuclei: Past, Present and Future Research". In: *Space Science Reviews* 116.3 (2005), 523?624.

Ferrarese, L. and Merritt, D. "A Fundamental Relation Between Supermassive Black Holes and Their Host Galaxies". In: *The Astrophysical Journal Letters* 539 (2000), pp. 9–12.

Francis, P. et al. "A High Signal-to-Noise Ratio Composite Quasar Spectrum". In: *The Astrophysical Journal* 373 (1991), pp. 465–470.

Gebhardt, K. et al. "A Relationship Between Nuclear Black Hole Mass and Galaxy Velocity Dispersion". In: *The Astrophysical Journal Letters* 539.1 (2000), pp. 13–16.

Grandi, S.A. "The 3000 Å Bump in Quasars". In: *The Astrophysical Journal* 255 (1982), pp. 25–38.

Hastings, W.K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57 (1970), p. 97.

Kaspi, S. et al. "Reverberation Measurements for 17 Quasars and the Size-Mass-Luminosity-Relations in Active Galactic Nuclei". In: *The Astrophysical Journal* 533 (2000), pp. 631–649.

Keel, William. *Optical Spectra of Various Kinds of Active Galactic Nuclei.* 2012. URL: `http://pages.astronomy.ua.edu/keel/agn/spectra.html`.

Kinney, A.L. et al. "Template Ultraviolet to Near-Infrared Spectra of Star-Forming Galaxies and Their Application to K-Corrections". In: *The Astrophysical Journal* 467 (1996), pp. 38–60.

Korista, K.T. and Goad, M.R. "The Variable Diffuse Continuum Emission of Broad-Line Clouds". In: *The Astrophysical Journal* 553 (2013), pp. 695–708.

Kormendy, J. and Ho, L.C. "Coevolution (or not) of Supermassive Black Holes and Host Galaxies". In: *Annual Review of Astronomy and Astrophysics* 51 (2013), pp. 511–653.

Krolik, J.H. *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment.* Princeton, New Jersey, United States: Princeton University Press, 1999.

Laor, A. et al. "The UV Properties of the Narrow Line Quasar I Zwicky 1". In: *The Astrophysical Journal* 489 (1997), pp. 656–671.

Latif, M.A. et al. "Black hole formation in the early universe". In: *Monthly Notices of the Royal Astronomy Society* 433.2 (2013), pp. 1607–1618.

Markwardt, C.B. "Non-Linear Least Squares Fitting in IDL with MPFIT". In: *Astronomical Data Analysis Software and Systems* XVIII*, Quebec, Canada, ASP Conference Series, Vol. 411.* Ed. by Bohlender, D., Dowler, P., and Durand, D. 2009, pp. 251–254.

Matsuoka, Y. et al. "The Sloane Digital Sky Survey Reverberation Mapping Project: Post-Starburst Signatures in Quasar Host Galaxies at $z > 1$". In: *The Astrophysical Journal* 811 (2015).

Metropolis, N. et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21 (1953), p. 1087.

Mo, H., Bosch, F. van den, and White, S. *Galaxy Formation and Evolution.* Cambridge, United Kingdom: Cambridge University Press, 2010.

Moré, J. "The Levenberg-Marquardt Algorithm: Implementation and Theory". In: *Numerical Analysis* 630 (1978), p. 105.

Mortlock, D.J. et al. "A luminous quasar at a redshift of $z = 7.085$". In: *Nature* 474 (2011), pp. 616–619.

Peng, C.Y. et al. “Detailed Structural Decomposition of Galaxy Images”. In: *The Astrophysical Journal* 124 (2002), pp. 266–293.

Peterson, B.M. *An Introduction to Active Galactic Nuclei.* Cambridge, United Kingdom: Cambridge University Press, 1997.

Peterson, B.M. “Reverberation Mapping of Active Galactic Nuclei”. In: *Astronomical Society of the Pacific* 105.685 (1993), pp. 247–268.

Polletta, M. et al. “Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey”. In: *The Astrophysical Journal* 663 (2007), pp. 81–102.

Riviera, G. Calistro et al. “AGNfitter: A Bayesian MCMC Approach to Fitting Spectral Energy Distributions of AGNs”. In: *The Astrophysical Journal* 833 (2016).

Romero, G.E. et al. “Relativistic Jets in Active Galactic Nuclei and Microquasars”. In: *Space Science Reviews* (2017), pp. 1–57.

Serra, P. et al. “CIGALEMC: Galaxy Parameter Estimation Using A Markov Chain Monte Carlo Approach with CIGALE”. In: *The Astrophysical Journal* 740 (2011).

Shen, Y. et al. “A Catalogue of Quasar Properties From Sloane Digital Sky Survey Data Release 7”. In: *The Astrophysical Journal* 194 (2011).

Sparke, L.S. and Gallagher, J.S. *Galaxies in the Universe: An Introduction.* Cambridge, United Kingdom: Cambridge University Press, 2007.

Urry, C. Megan and Padovani, Paolo. “Unified Schemes for Radio-Loud Active Galactic Nuclei”. In: *Publications of the Astronomical Society of the Pacific* 107 (1995), pp. 803–845.

Véron-Cetty, M.-P., Joly, M., and Véron, P. “The unusual emission line spectrum of I Zw 1”. In: *Astronomy & Astrophysics* 417.2 (2004), pp. 515–525.

Vestergaard, M. and Wilkes, B.J. “An Empirical Ultraviolet Template for Iron Emission in Quasars as Derived from I Zwicky 1”. In: *The Astrophysical Journal Supplement Series* 134 (2001), pp. 1–33.

Wills, B.J., Netzer, H., and Wills, D. “Broad Emission Features in QSOs and Active Galactic Nuclei. II. New Observations and Theory of Fe II and H I emission”. In: *The Astrophysical Journal* 288 (1985), pp. 94–116.

Winter, L.M. “Extragalactic X-ray Surveys of ULXs and AGNs”. In: (2008).

Xiao, T. et al. “Exploring the Low-Mass End of The M_{BH}–σ_* Relation with Active Galaxies”. In: *The Astrophysical Journal* 739 (2011).