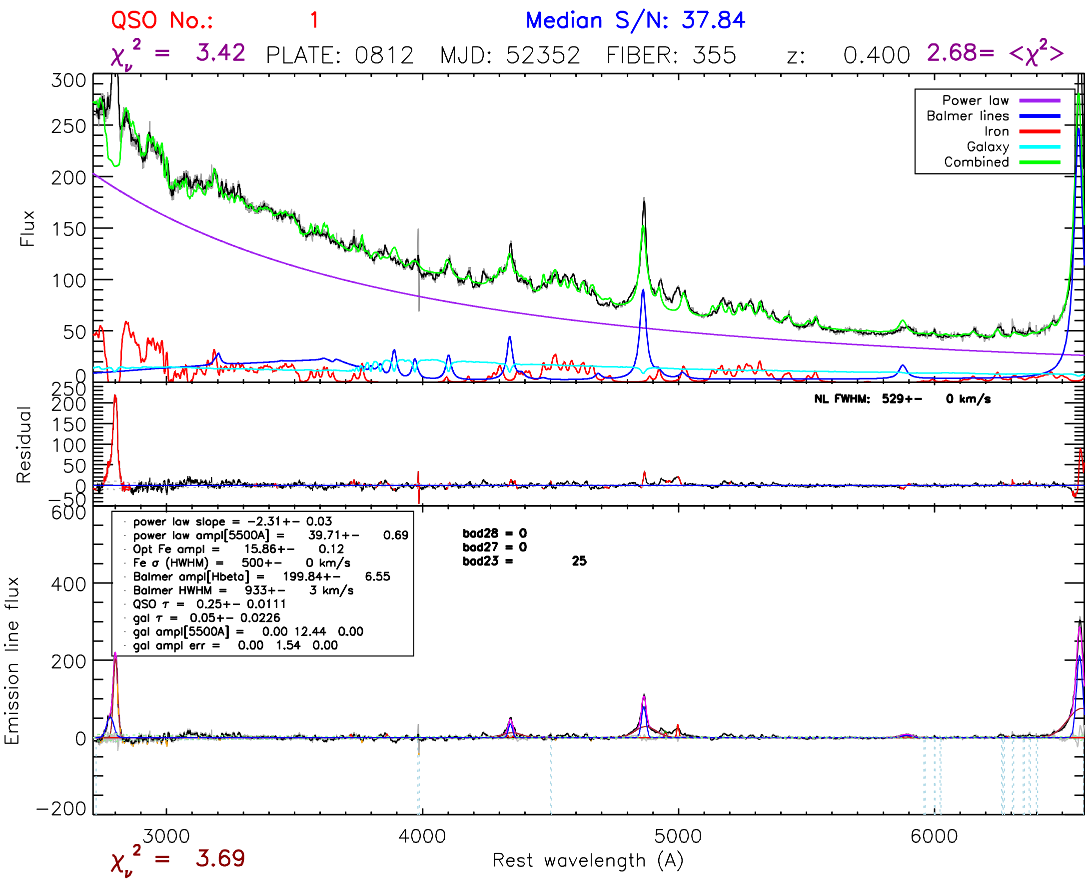
# Investigating the Host Galaxy Contribution in Spectra of Active Galactic Nuclei with Simulations



**Bachelor Project**


**By Lukas S. Engedal**

*Niels Bohr Institute*

*University of Copenhagen*


*June 2013*


**Supervisor:**

**Marianne Vestergaard**

*Dark Cosmology Centre*

*University of Copenhagen*

*Front page illustration:*
*A plot of the results of doing spectral decomposition of an AGN spectrum*
*The flux is in units of* $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ \text{Å}^{-1}$

## Abstract

In this thesis we examine the effect of tree different parameters on how accurately the host galaxy contribution to the spectrum of Active Galactic Nuclei can be determined using spectral decomposition. The three parameters are; the signal-to-noise ratio, the spectral resolution and the ratio between the host galaxy flux and the flux from the core. This is done using sets of simulated data based on two different datasets from the Sloan Digital Sky Survey, and created by adding additional galaxy flux to the Sloan datasets, changing the resolution of them and adding additional random noise to them, for a lot of different values of the three parameters. The value of the host galaxy contribution for each of these sets of simulated data is then determined through spectral decomposition, and the results are compared in different ways. The results are somewhat ambiguous, but generally follow our expectations. They indicate that lowering the signal-to-noise ratio and the spectral resolution decreases the precision of the measurements of the host galaxy contribution, but surprisingly it does not systematically lower the accuracy. The results also indicate that the ratio between the galaxy flux and the flux from the core influences the accuracy of the measurements, particularly when the ratio is low, and finally that the spectral decomposition seems to result in a systematic underestimation of the host galaxy contribution of around 10 %. Based on our results we make a cautious recommendation to other people working with AGNs, and with determining the host galaxy contribution for AGNs, to choose data with a signal-to-noise ratio of no less than 10 to 15, a spectral resolution of no more than about 400 km/s, and to be extra careful when working with datasets with a low host galaxy contribution, and to generally expect large uncertainties.

## Acknowledgements

I would like to thank my supervisor Marianne Vestergaard, for providing me with the very interesting subject of this thesis, the carefully chosen data used to examine the subject, the fascinatingly complex tree of programs for handling the data, and for all the help and guidance in general. I would also like to thank the Dark Cosmology Centre for providing me with access to their computational server, without which the data processing would still be ongoing.

# Contents

# 1. Introduction

Active Galactic Nuclei, shortened AGNs, are some of the most luminous objects in the universe, detectable at distances almost stretching to the edge of the known universe. This makes them prime candidates for studying the parts of the universe that are far away from us, and from where the light we are now receiving was emitted a long time ago. Such studies of AGNs are conducted using techniques such as reverberation mapping, and using high resolution and high signal-to-noise ratio data from some of the best instruments we currently have, such as the Hubble Space Telescope.

But the study of AGNs is naturally also associated with some problems and uncertainties, and there are different things that need to be taken into account when working with AGNs. One such thing is the host galaxy contribution to the AGN spectrum, which is particularly important for low redshift and low luminosity AGNs, where the host galaxy contribution might be accountable for most of the measured light. But how accurately we can actually determine, and subsequently subtract, the host galaxy contribution, and whether it depends on such technical details as the amount of noise in the spectrum or the spectral resolution, is not very well known. And with the Hubble telescope going out of service any day now, and no current plans for a successor, one cannot help wonder whether perhaps data with a lower signal-to-noise ratio and a lower resolution might be used to study AGNs as well.

With this thesis we hope to be able to shed some light on the importance of three different parameters on our ability to accurately and precisely determine the host galaxy contribution to the AGN spectrum, and hopefully to be able to give a cautious recommendation of what values of the three parameters other people should go for when choosing data for studying AGNs.

Below, the specific purpose of this thesis is described briefly. Following that is a section called 'The Big Picture', where the overall field of research we are dealing with is introduced. This is followed by a section describing the process of creating and analyzing our data, and a section describing the results of this process. And finally, there is a section with a discussion of these results, one with the overall conclusion and one with a list of references.

## 1.1. Purpose

The purpose with this thesis is to examine the effect of three different parameters on how accurately we can determine the host galaxy contribution to the AGN spectrum.

The first parameter is *the Galaxy/Power-law ratio*, which is the ratio between the flux from the host galaxy and the flux from the AGN, which is modeled by a power law, at a given wavelength. The second parameter is *the signal-to-noise ratio*; a way of expressing the ratio between the amount of noise in the spectrum and the strength of the spectrum itself. Finally, the third parameter is *the spectral resolution*, which is a measurement of how large an interval, i.e. how many Ångström or kilometers per second, is represented by each pixel in the spectrum.

Using AGN data from the Sloan catalog, we create simulated data with different values of the three parameters by adding galaxy flux to the spectrum, changing the resolution and adding additional random noise. For sets of simulated data with different combinations of these three parameters we then determine the host galaxy contribution using spectral decomposition. The different values of the host galaxy contribution are then plotted in different ways in order to examine the effect varying the three parameters has on how accurately and how precisely the host galaxy contribution can be determined.

Based on the results of this examination we then finally try and predict the result of doing spectral decomposition of an AGN known as NGC5548, for which we know the value of the three parameters beforehand.

# 2. The Big Picture

In this thesis we are dealing with a kind of objects known as Active Galactic Nuclei, or AGNs. AGNs and the components they are believed to be composed of will be described briefly below, followed by a description of a particular technique used to study AGNs called reverberation mapping. Finally, there is a discussion of some of the important uses of the results of doing reverberation mapping, such as the luminosity - $r_{BLR}$ relationship. This section is based in part on Bradley M. Peterson's '*Active Galactic Nuclei'*.

## 2.1. Active Galactic Nuclei

*Active Galactic Nuclei* is a term covering a variety of different phenomena observed in galaxies which at first glance has only one thing in common; an unusually luminous core. Among these phenomena are *Seyfert galaxies*, *quasars*, *QSOs* and a range of other objects, and one of the big questions is whether they are actually different phenomena or just different names for the same kind of phenomenon, possibly seen from different angles, under different conditions and/or at different stages in their evolution. One thing these phenomena have in common and that sets them apart from other objects such as stars, is that their emission is strong across the entire electromagnetic spectrum, even in the high energy X-ray and γ-ray regions and in the low energy radio region, where for instance stars are normally quite weak. But this is also where some of them differ, some being what we call radio loud and some ratio quiet. In this thesis we will assume they are indeed just different names for the same kind of phenomenon, and simply refer to them as AGNs.

Another thing that AGNs have in common is that they are extremely bright; the luminosity of an AGN is equivalent to billions or even trillions of stars, and some of them easily outshine their host galaxies. And what is perhaps even more puzzling is that the light seems to come from an area in the center of the galaxy no bigger than a few cubic parsecs, which is just a tiny fraction of the size of a normal galaxy. The small volume and the large luminosity means it is currently not possible to resolve the actual core of these galaxies, and therefore no one really knows what is going on, although there are some commonly accepted ideas, which will briefly be presented below.

### 2.1.1. Components of an AGN

At the center, literally, of an AGN is believed to be a *super massive black hole*. Just to remind ourselves; a *black hole* is an object so compact and with such a strong gravitational pull that, within a certain radius, not even light can escape it. By using the emission lines in the AGN spectrum and simple virial arguments, one can calculate a mass of these black holes in the order of $10^8$ – $10^{10}$ $M_\odot$. This means we are talking about objects with masses of up to a 10 billion times the mass of the sun, but with a radius of only around one third the size of the solar system, which is a tiny object seen from a galactic scale, hence the name *super massive* black holes.

Surrounding the black hole is believed to be what is called an *accretion disk*, which is a disk of gas rotating around the black hole and slowly but steadily falling onto it. If we assume all the light from the AGN is being supplied by turning mass into energy, one can do some simple calculations on how much fuel is needed to sustain the AGN. Doing so shows that surprisingly only mass equivalent to a few solar masses per year is needed to keep the AGN going. What actually supplies the AGN with the required energy though, is not very well understood and probably involves more than just turning matter into energy. A lot of different components are also believed to be involved in creating the AGN spectrum, such as thermal radiation from the accretion disk, synchrotron radiation from the jets and emission lines originating in the narrow- and broad line regions.

Also originating somewhere near the black hole in the core is the so-called *jets*. They are believed to be caused by very strong and intricate magnetic fields, whose origin is not very well understood. When a particle is caught in these magnetic fields, it is then forced away from the black hole and out of the AGN in a narrow cone, perpendicular to the accretion disk, at high velocity. If the AGN has the right inclination, the jet could potentially be pointed right at us, and that could possibly explain some of the differences between the different subclasses of AGNs.

Further out from the black hole, in the same plane as the accretion disk, is believed to be a giant, doughnut shaped cloud of dust and gas known as a *torus.* The material in this torus is also orbiting the black hole, and some of it is possibly also falling inward and unto the accretion disk. The main difference between the accretion disk and the torus is that while the accretion disk has almost zero height, the torus is quite voluminous and, seen from the black hole, obscures a large part of the sky. Depending on whether the AGN is positioned so that the torus is between the black hole and the observer or not is probably another of the main differences between some of the AGN subclasses.

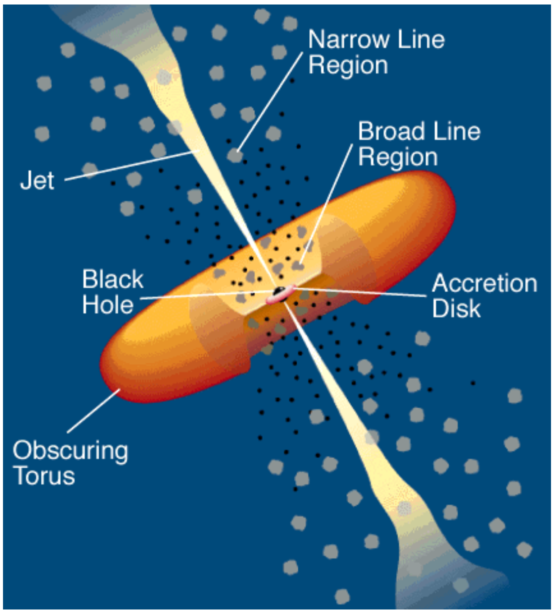



**Figure 2.1 – An illustration of what an AGN might look like up close. Diagram from Urry & Padovani 1995**

Further out from the black hole, perpendicular to the accretion disk, is the *Broad Line Region* or BLR, named so because of the width of the emission lines believed to originate there. The broad line region is traditionally believed to consist of a large number of clouds consisting mainly of hydrogen gas, which absorbs some of the light coming from the core, and in return sends out emission lines. What then makes theses emission lines broad is that the individual clouds are moving at high velocity in different directions with regard to each other, causing the emission lines to be Doppler shifted. However, more recent studies (W. Kollatschny, M. Zetzl 2013) seem to indicate that the broad line region actually consists of a more even distribution of gas being blown out from the accretion disk by a sort of wind. This wind blows the gas out perpendicular to the accretion disk, and at the same time it rotates with the disk, causing the gas to move out from the accretion disk in a sort of spiral, which is then the cause of the velocity dispersion. A typical value for the velocity dispersion in the broad line region is in the order of 10.000 km/s.

Even further out than the broad line region we got the *Narrow Line Region*, or NLR. Like the BLR, the NLR consists of a large number of clouds of gas, but, unlike the BLR, the NLR is also sufficiently far away from the black hole that dust grains can form and be part of the clouds. The clouds in the NLR are bigger and more massive but also less numerous than the clouds in the BLR. And again, as the name implies, the narrow line region is named so because of the narrow emission lines believed to originate there. A typical value for the velocity dispersion in the narrow line region is in the order of 500 km/s.

## 2.2. Reverberation Mapping

One way of studying AGNs is by doing what is called *Reverberation Mapping.* It is a technique that takes advantage of the fact that the broad emission lines in the AGN spectrum originates in the broad line region, and that they are in turn 'fueled', so to speak, by the light produced in the center of the AGN. Any variations in the strength of this light should then also be expected to result in variations of the emission lines.

The continuum is produced in the center of the AGN in the neighborhood of the black hole, and from there the light is then radiated away, eventually reaching the broad line region where some of it is absorbed by the gas clouds, leading to reemission observed as the broad emission lines of the AGN spectrum. The strength of these emission lines depends on the amount of light absorbed, which then again depends on the amount of light emitted from the core. Therefore, an increase or decrease in the amount of light emitted in the core should lead to a corresponding increase or decrease in the strength of the emission lines.

What's interesting then is that whenever a variation in the amount of light emitted in the core happens, the light then has to travel from the core out to the BLR before it can cause a change in the strength of the emission lines. This leads to a time delay between the variation of the core continuum and the variation in the emission line strength, denoted $t_d$, which is what one might observe and use. Assuming that the broad line region has a low filling factor, i.e. consists mainly of empty space with only a limited amount of gas, we can then simply divide the time delay $t_d$ by the speed of light $c$, in order to get an estimation of the distance from the core to the broad line region, referred to as $r_{BLR}$.

The use of reverberation mapping relies on a few crucial assumptions, mainly concerning the different timescales involved. First of all, the time it takes for the light to be absorbed by the BLR clouds and reprocessed into emission line photons must be significantly shorter than the time delay $t_d$ in order not to cause even further delay. And on the other hand, the time delay $t_d$ must be shorter than the timescale on which the geometry of the gas changes significantly, otherwise variations in the geometry will lead to variations in the emission line strength, potentially causing us to over- or underestimate the time delay.

One of the disadvantages of reverberation mapping is actually the timescales involved, because it requires that you spend a lot of observation time looking at the same object, hoping for a variation in the continuum sufficiently strong to be able to accurately measure the time delay. And in some cases, the time delay might be several years, so you need a lot of patience when doing reverberation mapping. That is also why a lot of effort has been put into trying to find a connection between the characteristics of AGNs that are simpler and quicker to measure, and the radius of the broad line region. One such is the $r_{BLR} - L$ relationship.

### 2.2.1. The $r_{BLR} - L$ relationship

One of the very interesting uses of the $r_{BLR}$ we get from reverberation mapping is the so-called $r_{BLR} - L$ *relationship*, where $L$ is the luminosity of the AGN, typically measured in the optical range. By using measurements of the luminosity and $r_{BLR}$ of different AGNs to determine this relationship, one can then determine the $r_{BLR}$ of other AGNs by simply measuring their luminosity, and then the $r_{BLR}$ can be used to estimate a range of other interesting things. Before discussing this further, let's do a quick theoretical estimation of what the $r_{BLR} - L$ *relationship* should be.

To derive such a relationship we start by looking at the main constituent of the BLR clouds; hydrogen. The energy required to ionize hydrogen is $h\nu_H = 13.6\ eV$, and this energy is supplied by the radiation coming from the core of the AGN. The amount of energy emitted per second at a certain frequency $\nu$ is $L_\nu$, so the amount of photons with frequency $\nu$ emitted per second is $L_\nu / h\nu$. We can then calculate the total amount of photons emitted per second with sufficient energy to ionize hydrogen, denoted $Q(H)$, by integrating over all frequencies from $\nu_H$ and upwards.

$$Q(H) = \int_{\nu_H}^{\infty} \frac{L_\nu}{h * \nu} d\nu \tag{2.1}$$

An interesting thing then, is to compare this with the number of hydrogen atoms in the broad line region, in order to get an idea of how many hydrogen atoms are actually ionized, the so-called *ionization parameter* $U$. The distance from the core to the broad line region is $r_{BLR}$ , and so the surface area of the BLR being hit by light from the core is $4\pi {r_{BLR}}^2$. Assuming that the light moves with velocity $c$ through the cloud, the total volume of the BLR that the light can reach during one second is then $c * 4\pi {r_{BLR}}^2$. To get the total number of hydrogen atoms in this volume we simply multiply it with the hydrogen density $n_H$. Knowing the number of photons capable of ionizing hydrogen and the total number of hydrogen atoms, we then have our ionization parameter

$$U = \frac{Q(H)}{4\pi {r_{BLR}}^2 c\, n_H} \tag{2.2}$$

Rearranging equations (**2.1**) and (**2.2**) we have

$$r_{BLR} = \sqrt{\frac{Q(H)}{4\pi c\, n_H * U}} \propto Q(H)^{0.5} \tag{2.3}$$

If we then assume that the strength of the ionizing spectrum is a function of the luminosity, but the shape of the ionizing spectrum is independent of luminosity, i.e. the amount of photons capable of ionizing hydrogen at a given frequency is a function of luminosity, but the ratio between the amount of photons capable of ionizing hydrogen at different frequencies is not, then $L \propto Q(H)$ and we get

$$r \propto L^{0.5} \tag{2.4}$$

The slope of the $r_{BLR} - L$ relationship is typically referred to as $\alpha$, and from a theoretical point of view our expectation is $\alpha = 0.5$. Surprisingly this simply derived value of $\alpha$ is actually very well in accordance with what have been found through extensive research using reverberation mapping results from around 40 AGNs (Bentz et al. 2009) yielding a value of $\alpha = 0.519^{+0.063}_{-0.066}$.

Naturally there are also some problems with the $r_{BLR} - L$ relationship, particularly concerning the determination of $L$. First of all, what we actually get from measuring the continuum of the AGN is naturally not the luminosity but the flux, and in order to translate the flux into luminosity we need to know the distance to the AGN. Measuring the distance to the AGNs would require some form of distance indicator or standard candle, such as a Cepheid, and such objects are very rare. What is done instead is to use a distance estimated from the measured redshift, which is a somewhat inaccurate method, especially for low-redshift AGNs. Another uncertainty is in the assumption that all the light is coming from the core, thereby ignoring the contribution from the host galaxy. For high luminosity AGNs this is probably not a bad assumption, but for low luminosity AGNs the contribution from the host galaxy is not negligible, and failing to correct for this can cause the slope of the relationship to be overestimated to as much as 0.7 (Kaspi et al. 2005). Therefore, ideally the contribution from the host galaxy should be identified and subtracted from the spectrum before using the $r_{BLR} - L$ relationship.

By using the $r_{BLR} - L$ relationship people are also trying to estimate other important characteristics of AGNs, such as the mass of the central black hole (Peterson et al. 2004) and the distance to the AGNs (Watson et al. 2011). By approximating the luminosity using the measured flux and the redshift as mentioned before, one can then use the $r_{BLR} - L$ relationship to find the radius of the broad line region. Using velocities based on the velocity dispersion of the emission lines, the mass of the center black hole can then be calculated using simple virial arguments. Or by instead measuring the $r_{BLR}$ of an AGN by doing reverberation mapping, one can then use the $r_{BLR} - L$ relationship to get a value of the luminosity, and with this luminosity and with the measured flux of the AGN one can then calculate the distance to the AGN. And particularly methods allowing people to measure the distance to objects are very popular and sought after, and by using AGNs it could potentially be possible to measure distances to objects as far away as $z \approx 4$, which is quite a lot better than current methods.

Both of these techniques ultimately depends on the accuracy of the $r_{BLR} - L$ relationship though, and the slope of this relationship in turn depends to some degree on how well we can determine the host galaxy contribution to the AGN spectrum and correct for this, particularly when dealing with low luminosity low redshift AGNs. And how accurately the host galaxy contribution can be determined under different conditions is exactly what we aim to examine below.

# 3. Data Processing

The overall method used in this thesis is actually quite simple. We choose a dataset, add additional galaxy flux to it, change the spectral resolution, add additional noise and then measure the host galaxy contribution using spectral decomposition. We do this for a range of different values of the galaxy flux, the resolution and the noise, and then we plot the results hoping it will give us an idea of the effect of varying these three parameters. Each of these steps is explained in a bit more detail below.

## 3.1. Data

In this thesis we are using a total of three datasets, from two different sources.

Two of the datasets are from the Sloan Digital Sky Survey, the first one, referred to as *sdss1*, has the rather dull name $\mathrm{J}120924.07+103612.0$ and the second one, *sdss2*, is called $\mathrm{J}154530.23+484608.9$. In the Sloan database, the two dataset are also identified by the three parameters $plateid$, $mjd$ (mean Julian date) and $fiberid$. The first dataset, sdss1, has plateid = 1229, mjd = 52723 and fiberid = 489, and the second dataset, sdss2, has plateid = 0812, mjd = 52352 and fiberid = 355. These two particular datasets have been chosen firstly because their host galaxy contribution is small, and secondly because sdss2 has strong Fe emission and sdss1 has week Fe emission, which could possibly also play a role in how effectively the host galaxy contribution can be determined.

The third object we are working with is the AGN known as NGC5548. We have chosen NGC5548 because it is one of the most studied AGN's, and therefore we have a very precise value of what the host galaxy contribution is for the spectra (Bentz et al. 2013), which allows us to test how well we can predict the result of doing spectral decomposition of NGC5548, based on the results from looking at the two Sloan datasets. We originally intended to use a whole range of different NGC5548 datasets to test our predictions, but we later found out that of all the datasets only three were actually suitable for doing spectral decomposition. And of these three datasets, two of them turned out to cause the fitting routine to fail, for reasons unknown, leaving us with just one.

## 3.2. Simulating data

For simulating a whole range of datasets with different values of the host galaxy contribution, the spectral resolution and the signal-to-noise ratio we use the Sloan datasets, sdss1 and sdss2. For each desired combination of the three parameters we use one of the two Sloan datasets to create a new dataset, by firstly adding additional galaxy flux, secondly degrading the spectral resolution and finally adding additional noise to the data. Each of these three steps is carried out by an IDL program written specifically for this task, and the three programs are described further below.

### 3.2.1. Add galaxy flux

The first of these three programs is called $addgal$, and is used to add additional galaxy flux to the AGN data, using a standard galaxy template, made by averaging a lot of galaxy spectra. The galaxy template can be found in the appendix, Figure 8.1.

First, the galaxy template is normalized so that the mean galaxy flux in the interval 5050Å to 5150Å is equal to the mean flux of the data in the same interval. Second, the galaxy template is then multiplied with the desired galaxy/power-law ratio, and added to the AGN data. Ideally the galaxy flux should be normalized to the power-law flux and not the total flux of the data, but by assuming that the flux contribution from emission lines in the 5100 Å interval is negligible, we can set the power-law flux equal to the total flux. This way we simulate an AGN with a specific, and well known, host galaxy contribution, which is a lot easier than actually having to find different AGNs with different values of the host galaxy contribution, which would otherwise be a more correct way of doing it.

$Example$: If we for instance want a galaxy/power-law ratio of 0.5, the normalized galaxy template is multiplied by a half and added to the data. The host galaxy flux in the 5100 Å interval is then half of the power-law flux in this interval, and a third of the total flux of the simulated data.

In many of the plots below, the variable $rgalpow$ is used to represent the Galaxy/Power-law ratio, so an $rgalpow$ value of 0.5 is the same as a Galaxy/Power-law ratio of 0.5.

### 3.2.2. Degrade the spectral resolution

The second program is called $degrade$, and is used to change the spectral resolution of the AGN data. This is done in one of a few different ways, depending on the circumstances.

If the wavelength resolution is linear, and we want to change it by some integer number, the program simply adds together the flux of the appropriate number of pixels, and assigns the new pixel a wavelength in the middle of the previous ones.

$Example$: If for instance we want to change the resolution from 2 Å per pixel to 4 Å per pixel, the program simply adds the flux of the pixels together in pairs, and assigns the new pixel a wavelength halfway between the previous two.

It is a bit more complicated if we want to change the resolution by some non-integer number. In that case, the program adds the flux of as many whole pixels as possible, and then an appropriate fraction of the flux of the next pixel, and calculates the wavelength in the same way.

$Example$: If for instance we want to change the resolution from 2 Å per pixel to 5 Å per pixel, this is done by adding together the flux of the first pixel and the second pixel and then half the flux of the third pixel, giving us the first of the new pixels. Next, it then adds the remaining half of the flux of the third pixel to the flux of the fourth and the fifth pixels, giving us the second of the new pixels, and so on.

Another problem is if the total number of pixels is not divisible by the total number of pixels in the new wavelength resolution, leaving us with insufficient pixels at the end to create the last pixel of our new resolution. In this case a simple linear approximation is used to create the final pixel.

$Example$: If for instance we are adding together two and a half pixels each time and are left with one and a half pixels at the end, the flux of the one and a half pixels is then added together, and then divided by 1.5 and multiplied by 2.5.

If the wavelength resolution is logarithmic, the procedure is more or less the same. First, we start by taking the base 10 logarithm to the wavelengths, and the result is then put through the same kind of treatment as described above for the linear case. Following this, the wavelengths are then changed back to being logarithmic by calculating 10 to the power of the result of the linear procedure.

$Example$: If for instance the first pixel has a wavelength of $10^{2.01}$ Å, the second one $10^{2.02}$ Å and so forth, the wavelengths are logarithmic, each time changing by $10^{0.01}$ Å. Calculating the logarithm of these pixels, the first pixel is then at 2.01, the second at 2.02 and so on, now changing by 0.01 each time. If we then want a resolution that is a third of the original, the pixels are then added together in pairs of three as described above. The first pixel of the new resolution is then at 2.02, the second at 2.05 and so on, changing by 0.03 each time as intended. Calculating 10 to the power of these numbers gives us the final result, with the first pixel having a wavelength of $10^{2.02}$ Å, the second has a wavelength of $10^{2.05}$ Å and so forth.

In many of the plots below, the variable $ddx$ is used to describe the resolution. A $ddx$ value of 1 means the resolution is the same as in the original data, where as a $ddx$ value of 2 means that the resolution is half of what it was in the original data, a $ddx$ value of 4 means it is a quarter of the original resolution and so on. The Sloan data has a resolution of 69 km/s per pixel, so with a $ddx$ value of 2 the resolution is then 138 km/s per pixel and so forth.

## 3.2.3. Add noise

The third and final program is called $addnoise$, and is used to add additional noise to the AGN data, in order to lower the signal-to-noise ratio, shortened the S/N-ratio. Given the current S/N-ratio and the target S/N-ratio, it calculates how much extra noise is needed, creates it using a Gaussian random number generator in IDL called $randomn$, and adds it to the AGN data.

$Example$: If for instance we have an S/N-ratio of 40, and want to change it to 20, we start by calculating the current noise, $N$

$$S/N = 40 \;\; \rightarrow \;\; N_{initial} = S/40 \tag{3.1}$$

Where $S$ is the signal in the form of the flux. In the same way we can calculate the target noise, and then we can calculate the difference needed

$$N_{target}^2 = N_{initial}^2 + \Delta N^2 \;\; \rightarrow \;\; \Delta N = 0.0433 * S \tag{3.2}$$

The noise would then be calculated by using the IDL routine $randomn$ to create the necessary number of $sigma = 1$ random numbers and then multiplying them by $\Delta N$.

Using a random number generator naturally is a bit tricky due to the obvious randomness of such a generator. In order to account for this, for each combination of our three parameters, we create not just one dataset, but 200 datasets, differing only in the random noise added. By then doing spectral decomposition on each of these 200 datasets and averaging the results, it should remove most of the uncertainties from using a random number generator.

## 3.2.4. Example of simulated dataset

To provide an example of the effect of these three programs, Figure 3.1 shows the original sdss2 dataset and a simulated dataset based on the sdss2 dataset. The simulated dataset has an $rgalpow$ value of 1.00, which means that the average galaxy flux is equal to the average power-law flux in the 5100 Å interval, and a $ddx$ value of 2, which means that the resolution is half that of the original data. Furthermore it has an S/N-ratio of 20, which means that the noise is about 5 % of the flux, compared to an S/N-ratio of around 40 for the original data. The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ \text{Å}^{-1}$.

Looking at the plot of the simulated data, the first thing you notice is that there seems to be a lot more noise than on the plot of the original data, and the extra noise seems to make the features of the spectrum harder to distinguish. Another thing one notices is that the downward slope of the original data when going towards higher wavelengths seems to have more or less disappeared. This is due to the added galaxy flux, because the galaxy template has about 40 times as much flux at high wavelengths as it does at low wavelengths. The effect of the change in resolution is a bit more difficult to see, particularly because the resolution has only been halved. One obvious effect can be seen when looking at the y-axis, because the flux of the simulated data is generally about twice as high as that of the original data, which makes perfect sense considering that when changing the resolution to half that of the original, the flux of the pixels is added together in pairs.

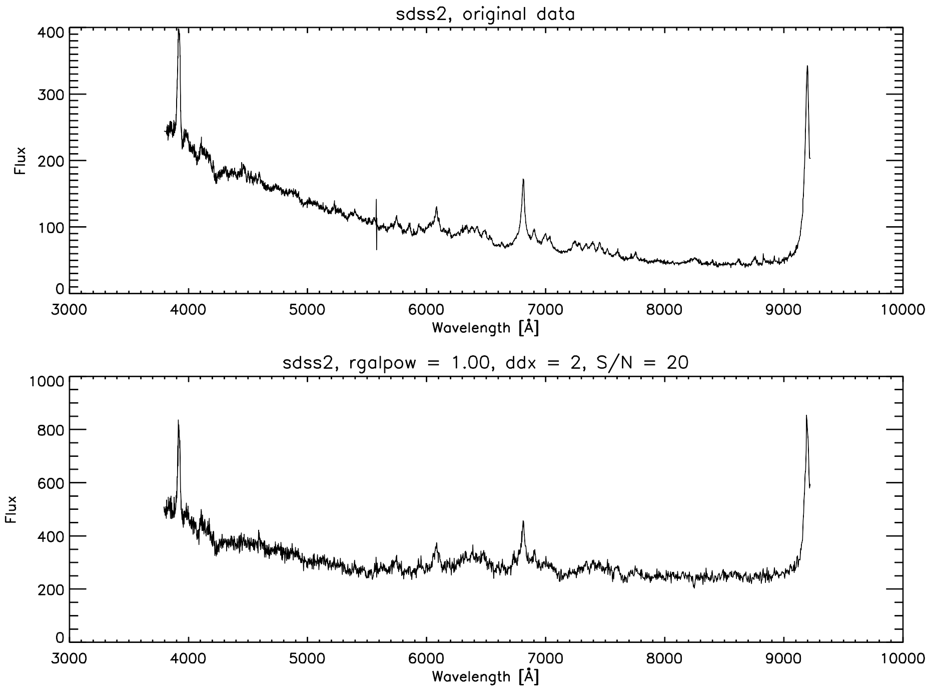



**Figure 3.1**
**Top: The original sdss2 data. Bottom: A simulated dataset based on sdss2**
**The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}$ Å$^{-1}$**

## 3.3. Spectral Decomposition

The next step in the process is the spectral decomposition. This is done using a tree of IDL programs supplied by my supervisor, with the IDL program MPFITFUN at its core. The programs themselves are quite complex, but the process in its basic form is actually not. Put simply, the process involves combining different components into a spectrum, comparing this spectrum with the AGN spectrum, and then adjusting the components in order to achieve the best possible fit to the AGN data. There are three basic components involved; the AGN continuum, the host galaxy spectrum and the emission lines, and each of these components are adjusted through a range of variables.

The AGN continuum is represented by a power law, where the two variables control the amplitude and the slope of the power law. The host galaxy is simulated using the standard galaxy template mentioned earlier, with a single variable controlling the amplitude of the galaxy spectrum. The emission lines are simulated using mathematical functions, typically either Gaussian or a Lorentzian functions. In our case, we are working with two different kinds of emission lines; Balmer lines and Fe, or iron, emission lines. Each of these two kinds of emission lines has two variables; one controlling the width and one controlling the amplitude of the lines.

### 3.3.1. MPFITFUN

In total, this means we got at least 7 different variables to adjust in order to achieve the best possible fit. Doing that manually would be very time consuming, but luckily there is a very clever IDL routine called *MPFITFUN* that can do the work for us. MPFITFUN requires 5 different inputs. The first of these is the wavelengths, the second is the AGN spectrum, and the third input is an array containing an estimation of the uncertainties of the AGN spectrum. The fourth input parameter is an estimate of the different variables, in order for the program to have an idea of what values to use as a starting point, and the fifth and final input parameter is the name of an IDL program that, given the wavelengths and a value for each of the variables, can return a spectrum for the program to compare with the AGN spectrum.

Based on this comparison, MPFITFUN then makes some adjustments to the values of the variables, uses the new values to draw a new spectrum, and compares this spectrum to the AGN spectrum once again. MPFITFUN then continues to change the variables and compare the resultant spectra with the data until it has achieved a fit as good as possible, or at least as good as it believes possible. The changes in the values of the variables and the determination of when the fit is done, is based on a least-chi-square test, where MPFITFUN tries to minimize the chi-square value of the fit, and calculating this fit can take anywhere from a few to several hundred iterations. One problem with the way the best fit is determined is that sometimes the program is what you might call a bit lazy. It might decide it is done when it finds itself in a local minimum in parameter space, although there are other, smaller minima in other parts of the parameter space that it has never explored. Despite this, it is a very powerful and important tool when fitting our data.

Once MPFITFUN has calculated a fit, it then returns an array containing the values of the variables found to give the best possible fit, and the uncertainties on these values. In the program tree we then actually use these values as the start parameters in yet another run of the MPFITFUN program, which surprisingly often runs for a few iterations before returning a new set of best fit values. These values are then returned to the program tree, and eventually saved in a *fits* file together with a lot of other information. This information comes from the fact that the program tree actually also does a variety of other things, and has a whole range of parameters that can be changed in order to deal with things such as extinction and redshift. In particular there is a lot of work put into trying to deal with emission lines, not just Balmer and Fe lines but also Paschen lines and other metal lines such as Mg and O, but it is not really relevant in our case, because we are actually only interested in the galaxy contribution, and therefore most of these features have been turned off in order to save time.

### 3.3.2. The result of spectral decomposition

After having completed the spectral decomposition, the result is a *fits* file containing the hundreds of different parameters and variables and their assigned values from the program tree, and using these values a plot of the fit can be produced, such as Figure 3.2 illustrating the result of doing spectral decomposition of the sdss1 dataset. In the top plot, the black spectrum is the original data, and the green one is our best fit. The four components, the power-law, the galaxy, the Fe-emission and the Balmer emission, are also plotted in respectively purple, cyan, red and blue. As before, the flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ \text{Å}^{-1}$. The middle plot is the residuals, i.e. the difference between the original data and our fit, and we want this to be as small as possible. The bottom plot is for plotting the fits to the remaining emission lines, and in our case it is empty because, as mentioned earlier, we have chosen to turn this feature off. The box in the lower plot contains the values of the most important variables used to create the fit, and we are particularly interested in the galaxy amplitude called $gal\ ampl$ and measured at 5100 Å. A similar plot for the sdss2 data can be found in Figure 8.2 in the appendix.

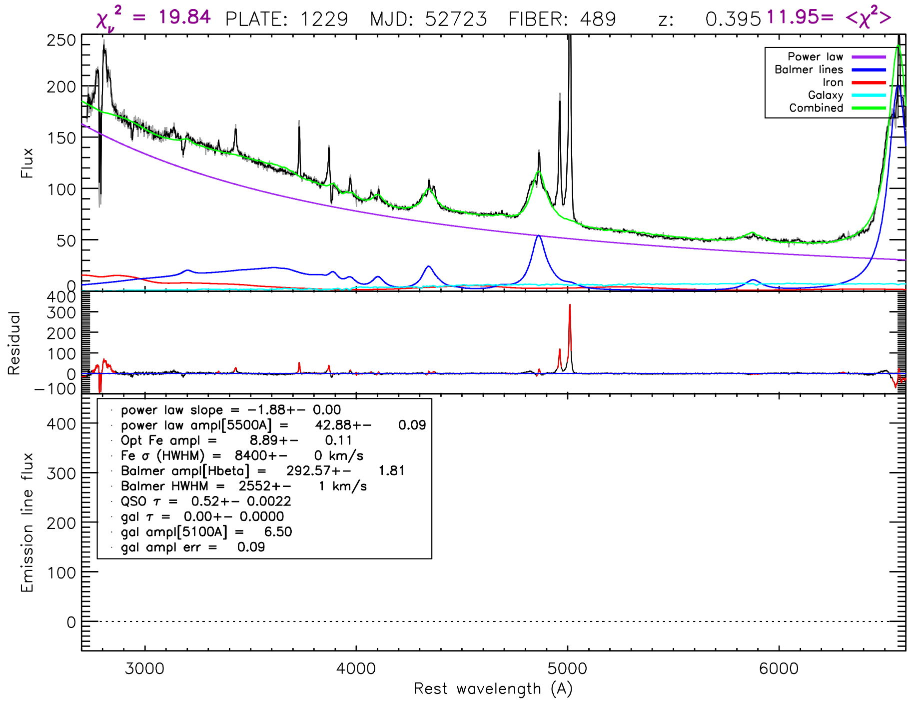



**Figure 3.2**
**The result of spectral decomposition of sdss2**
**The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ \text{Å}^{-1}$**

## 3.4. Expectations

For low values of the Galaxy/Power-law ratio, our expectation is that determining the galaxy contribution should be difficult, because the different features of the galaxy spectrum used to identify it might get lost in the noise, causing the fitting routine to underestimate the galaxy contribution. For high values of the Galaxy/Power-law ratio, the expectation is that generally the determination of the galaxy contribution should be pretty accurate, since the galaxy features should be clearly distinguishable.

When lowering the signal-to-noise ratio by adding more random noise, we would expect the fitting routine to have a harder time determining the correct galaxy contribution, since the added noise might obscure the features of the spectrum used in determining the contribution of the different components. Therefore, lowering the S/N-ratio is also expected to cause the fitting routine to underestimate the galaxy contribution.

And finally, our expectation of the effect of lowering the spectral resolution is also an underestimation of the galaxy contribution by the fitting routine. This is because when lowering the resolution by adding together pixels, the features of the galaxy spectrum might be distorted such that the fitting routine might not recognize them, or might even disappear entirely if the change in resolution is sufficiently large. And as before, we expect this to cause the fitting routine to use less galaxy flux to create its spectrum, thereby underestimating the host galaxy contribution.

# 4. Results

Based on the sdss1 and sdss2 datasets, 527 sets of simulated data with different values of our three parameters, the Galaxy/Power-law ratio, the spectral resolution and the signal-to-noise ratio, have been created and put through data processing; 164 sets of simulated data based on the sdss1 dataset and 363 based on the sdss2 dataset. The main reason for the difference between the number of sets of data for sdss1 and sdss2 is that, for reasons not quite understood, the process of fitting the data took three to four times as long for the sdss1 data as it did for the sdss2 data.

Another big difference between the two datasets became evident when examining the results of the fitting routine. We originally assumed the galaxy contribution of our two datasets to be zero, but this turned out to cause the fitting routine to overestimate the galaxy amplitude by as much as 400% for the sets of data based on sdss1. In order to deal with this, we made a correction for the actually non-zero galaxy amplitude of the sdss1 dataset evident when looking at the fit of the dataset in Figure 3.2 and equivalent to a Galaxy/Power-law ratio of 0.1304. Due to this correction, the Galaxy/Power-law ratios of the sets of simulated data based on sdss1 might seem a bit weird, because when we for instance originally added galaxy flux in order to achieve a Galaxy/Power-law ratio of 0.5, the actual Galaxy/Power-law ratio is now 0.6956 due to the correction. Surprisingly, no correction was needed for the sets of data based on sdss2, which might have something to do with the fitting routine actually returning a galaxy amplitude of zero for the sdss2 dataset, as evident in Figure 8.2 in the appendix.

As mentioned earlier, each of these 527 sets of simulated data contains 200 datasets differing only in the random noise added. For each of the 527 sets of data, the 200 datasets have then been put through spectral decomposition, resulting in 200 different values of each of the fit parameters, such as the galaxy amplitude that we are interested in. The 200 values have then been averaged; giving us a value for the entire set of simulated data, and the errors on the averaged result has been estimated as the standard deviation of the 200 values. The result of all this is then 527 values of the host galaxy contribution that can be plotted in different ways in order to investigate the effect of the variation of the three parameters. Due to the difficulty of producing and interpreting three dimensional plots, the three parameters have been plotted together in pairs, resulting in three different kinds of plots. These three kinds of plots are shown and discussed below.

## 4.1. Signal-to-Noise ratio and Spectral Resolution

The first combination of parameters examined is the combination of varying the signal-to-noise ratio and the spectral resolution. Figure 4.1 shows a plot representing a range of different sets of simulated data with different values of the S/N-ratio and the resolution, all based on the sdss2 dataset. The $rgalpow$ value is fixed at 0.70, and the S/N-ratio is then varied from 5 to 7, 10, 15, 20 and 25. This is done for $ddx$ values of 1, 6, 12, 18 and 24, each one with its own color. The asterisks on the plot represent the average value of the 200 datasets, and the colored lines represent the average value ± the standard deviation of the 200 datasets. On the left y-axis is the ratio between the galaxy amplitude returned by the fitting routine and the actual galaxy amplitude of the data. On the right y-axis is the same ratio, multiplied by 100 to show the ratio as a percentage.

The first, and probably most important, thing to notice about this plot, is the fact that the ratio between the value of the galaxy amplitude we get from our fit and the actual galaxy amplitude of the data seems pretty much fixed at around 0.9, which means that the fitting routine estimates a value of the galaxy amplitude which is about 90 % of the actual value. Changing the S/N-ratio does not really seem to affect this, and neither does changing the resolution. Only at very low S/N-ratios does the data points drift apart, and not in any consistent way. If, for instance, lowering the resolution made the fitting routine underestimate the galaxy contribution, we would expect the data points to be sorted with the blue data point having the highest value, and the cyan colored data point having the lowest value, and vice versa if lowering the resolution made the fitting routine overestimate the galaxy contribution, and this is not the case for any of the S/N-ratios. The only thing sorted in this way is the average ± standard deviation lines, which seems to indicate that lowering the resolution increases the uncertainties of the galaxy contribution values found by the fitting routine.

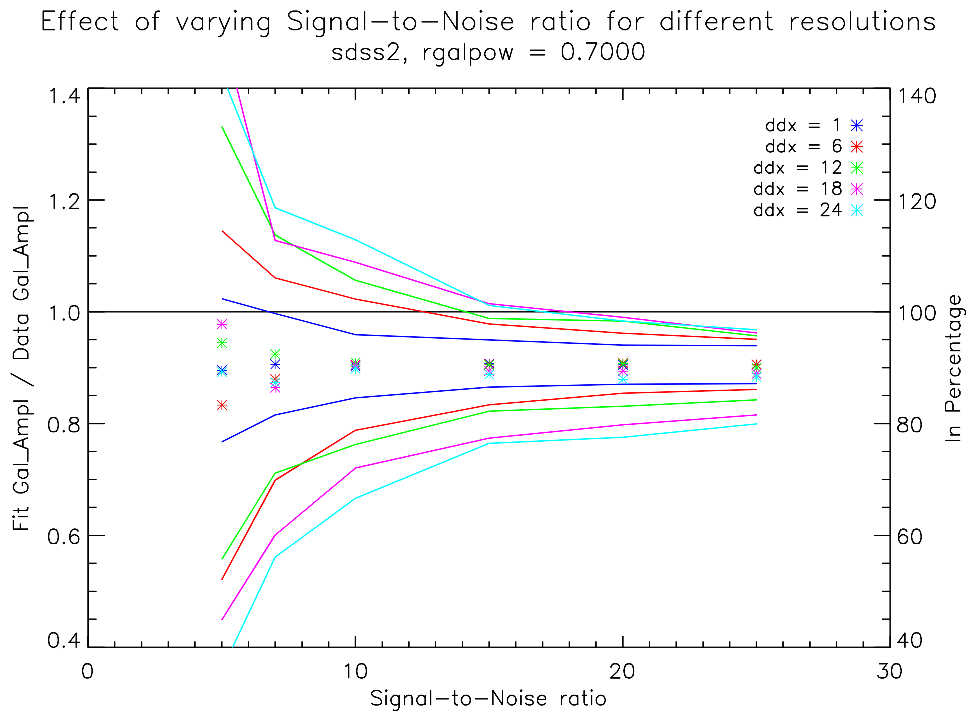



**Figure 4.1**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

Based on this plot, the results seems to indicate that lowering the spectral resolution and/or the signal-to-noise ratio decreases the precision of the values of the host galaxy contribution we measure, but does not consistently increase or decrease the accuracy of the values found by the fitting routine. This is actually a bit surprising, considering that we expected both lowering the resolution and lowering the signal-to-noise ratio to make fitting the data more difficult and cause the fitting routine to underestimate the galaxy contribution.

A similar plot for the sets of data based on sdss1 can be seen in Figure 4.2. This plot is a bit different from the sdss2 plot, in that the data points for the different resolutions no longer have roughly the same values, but actually seem to differ by a noticeable margin. The ranking of the data points for each S/N-value is not consistent though, only in that the blue data point has the highest values and the purple one has the lowest. As mentioned above, in order to follow our expectation that lowering the resolution should cause the fitting routine to underestimate the galaxy contribution, the cyan colored data point should have had the lowest value, but it actually seems to be placed roughly in the middle of the blue and purple data points. One thing that the plot does have in common with the sdss2 plot though is that it seems that lowering the S/N-ratio does not have any effect on how accurately we can determine the host galaxy contribution, it just seems to decrease the precision of our measurements.

All in all though, the huge uncertainties on the measurements make it difficult to say anything conclusive about the plots, other than the general tendencies pointed out. A cautious recommendation based on these plots would be to choose data with as high a resolution as possible, and a signal-to-noise ratio of no less than 10 and preferably 15.

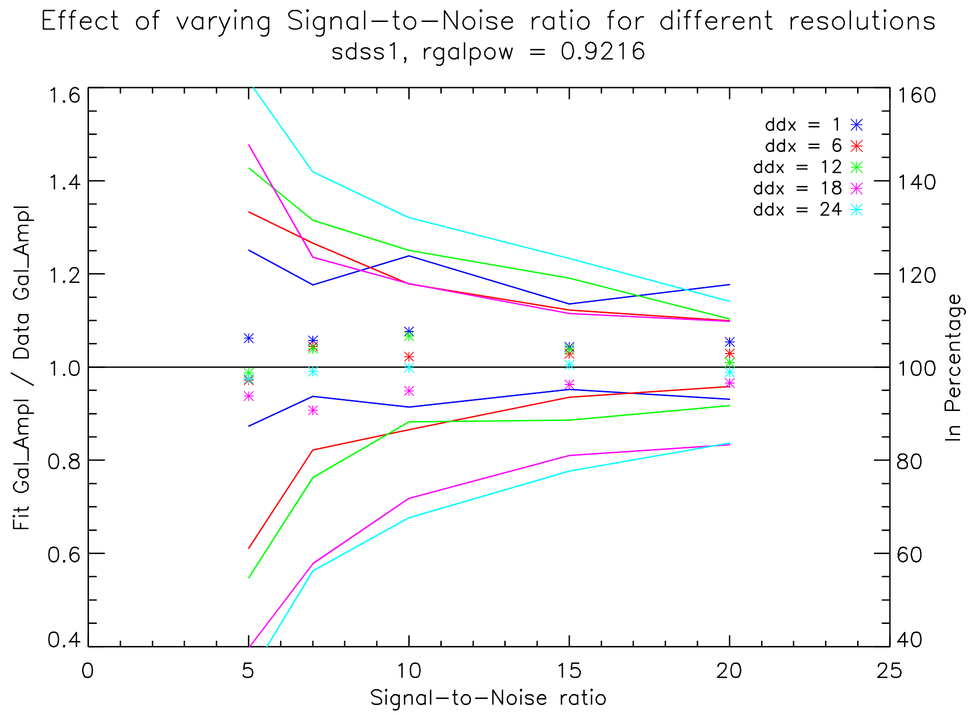



**Figure 4.2**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

## 4.2. Galaxy/Power-law ratio and Signal-to-Noise ratio

The second combination of parameters examined is the combination of varying the Galaxy/Power-law ratio and the signal-to-noise ratio. This is done in Figure 4.3 for sets of simulated data based on sdss1 and in Figure 4.4 for sets of data based on sdss2. Both plots are of sets of data with the original resolution and have Galaxy/Power-law ratios ranging roughly between 0 and 3. The signal-to-noise ratio is varied from 5 to 7, 10, 15 and 20, each with its own color. Once again each data point represents the average value of 200 datasets, and the lines represent the average ± the standard deviation of the 200 datasets. The y-axes are the same as on the previous plot, and the second x-axis added at the top of the plot shows the ratio between the galaxy flux and the total flux of the spectrum, and is non-linear.

Starting with Figure 4.3, it once again looks like the S/N-ratio does not really have a big effect on the accuracy of the values of the galaxy amplitude we find for the different sets of data, since for each value of the Galaxy/Power-law ratio, the five colored dots seem to be placed roughly on top of each other, and the few deviations from this seem random and does not really repeat themselves consistently. Another resemblance to what we found before is that it seems that lowering the S/N-ratio increases the uncertainties on our measurements. All of this is in accordance with what we found above and therefore not highly surprising. The interesting thing then, is to look at what happens when we vary the Galaxy/Power-law ratio. For low $rgalpow$ values the fitting routine seems to overestimate the galaxy contribution by around 5-10%, but as we increase the Galaxy/Power-law ratio, this changes to an underestimation, by as much as 20% for the highest ratios. This is a bit surprising, as we expected low values of $rgalpow$ to cause the galaxy features to be somewhat lost in the noise and the other emission, and would therefore have expected the fitting routine to underestimate the galaxy contribution. This overestimation might be due to the fact that our correction of the data was not quite enough, and that the galaxy amplitude of the sdss1 data might actually be larger than what we found, which would have a particularly strong effect at low Galaxy/Power-law ratios. The behavior at the high end of the scale is also different from what we expected, since with a high galaxy contribution we would expect the features of the galaxy spectrum to be strong and easily detected by the fitting routine.

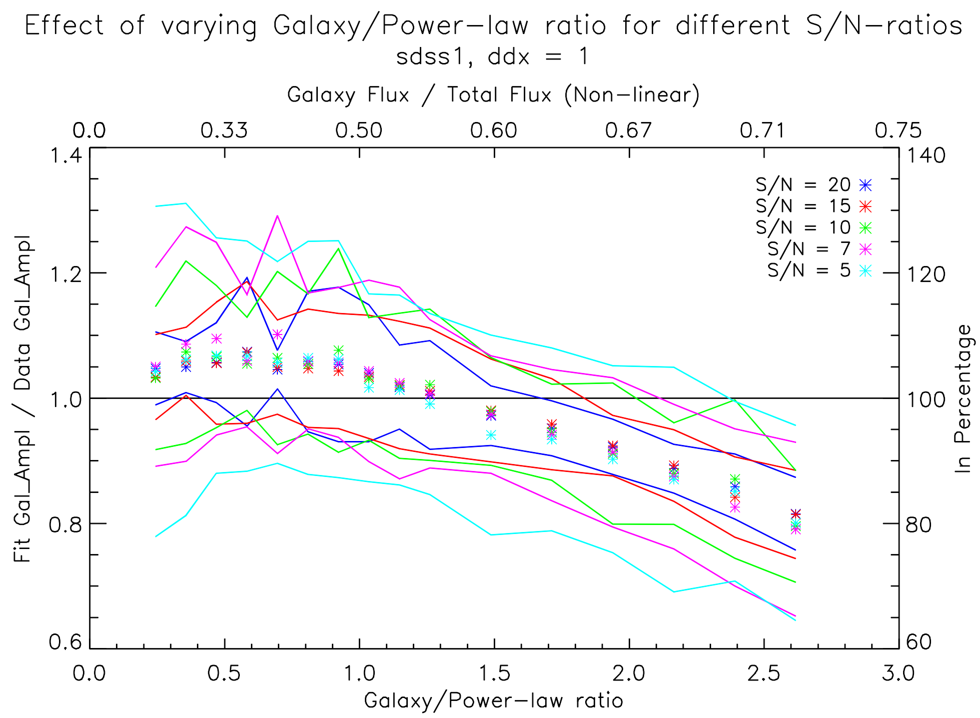



**Figure 4.3**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

The behavior of the sdss2 plot, shown in Figure 4.4, is a bit different from the sdss1 data, at least at the lower end of the plot. Here we can see that our fitting routine significantly underestimates the galaxy contribution, by as much as 80% for the lowest data point. In fact, looking at the actual distribution of the galaxy amplitudes found by the fitting routine for the sets of data with an $rgalpow$ value of 0.1, such as the histogram in Figure 8.3 in the appendix with $ddx = 1$, S/N = 20 and $rgalpow = 0.1$, we can see that for as many as 20% of the 200 datasets the fitting routine has found a galaxy amplitude of zero or very close to zero. This is in accordance with what we expected, as with such a small contribution from the host galaxy, the features of the galaxy spectrum are lost in the noise and in the other components making up the spectrum. At higher Galaxy/Power-law ratios, the data points seem to stabilize at around 0.9, with only a slight drop at the highest ratios. This overall behavior is more or less in accordance with our expectations, particularly for the low Galaxy/Power-law ratios.

The big differences in behavior, both at low and high Galaxy/Power-law ratios, of the two datasets, and of course the big uncertainties of the measurements, once again makes it a bit difficult to say something conclusive based on these two plots. Assuming that the behavior of the sdss1 data at low Galaxy/Power-law ratios is due to the previously mentioned correction not being sufficiently strong, a cautious recommendation might be to avoid data with a Galaxy/Power-law ratio of less than about 40% of the total flux, and to expect the fitting routine to underestimate the host galaxy contribution by around 10% for higher ratios. And once again it might also be advisable to choose data with an S/N-ratio of at least 10-15, in order to reduce the uncertainties on the results.

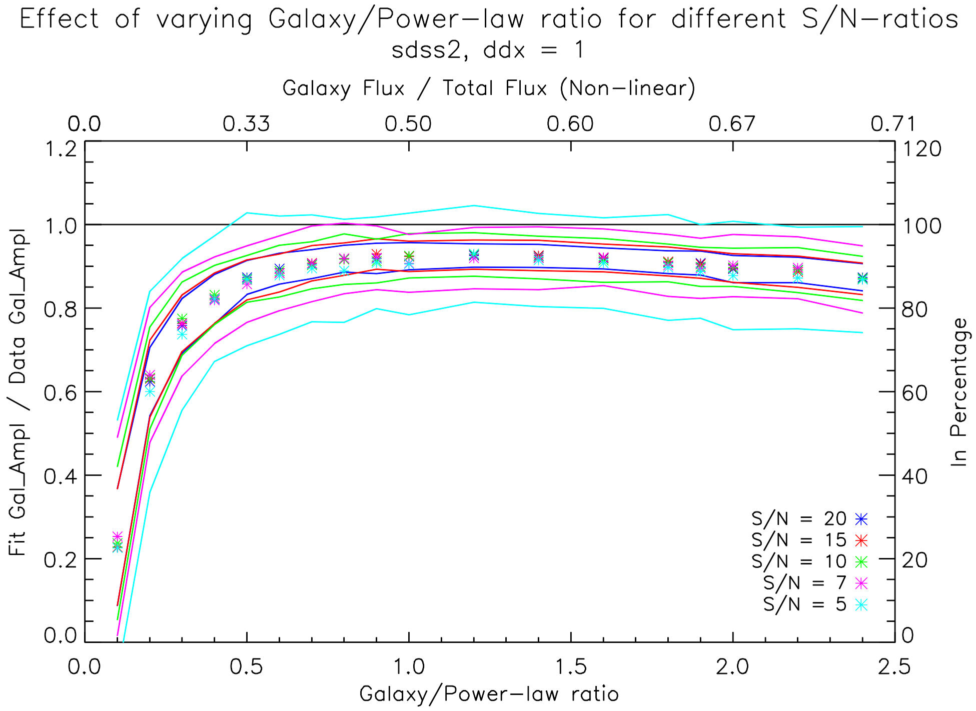



**Figure 4.4**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

## 4.3. Galaxy/Power-law ratio and Spectral Resolution

The third and final way of combining our parameters is the combination of varying the Galaxy/Power-law ratio and the spectral resolution. This is done in Figure 4.6 for sets of data based on sdss1 and in Figure 4.5 for sets of data based on sdss2. The signal-to-noise ratio is locked at 15, and the $rgalpow$ values are once again roughly in the range of 0 to 3. The $ddx$ values are 1, 6, 12, 18 and 24, each with its own color as usual, and the asterisks and lines on the plot are also representing the same things as previously. The different axes are also the same as in the previous plots.

At first glance, Figure 4.5 looks a lot like Figure 4.4 where it was the S/N-ratio we varied instead of the resolution. For low Galaxy/Power-law ratios the galaxy amplitude is highly underestimated by the fitting routine and for Galaxy/Power-law ratios from 0.5 and above, the data points are all roughly at 90% with a slight drop towards the end as was also the case before. A noticeable difference though, is that at the high end of the scale the different colored data points are actually starting to separate and drift apart from each other by as much as 10%. This is in accordance with what we found earlier, in that changing the resolution seems to have an effect on the galaxy amplitude found by the fitting routine, and this time, if you look closely, it actually even seems like the data points are somewhat ordered according to the resolution, with the blue data point with the highest resolution at the top and the cyan data point with the lowest resolution at the bottom.

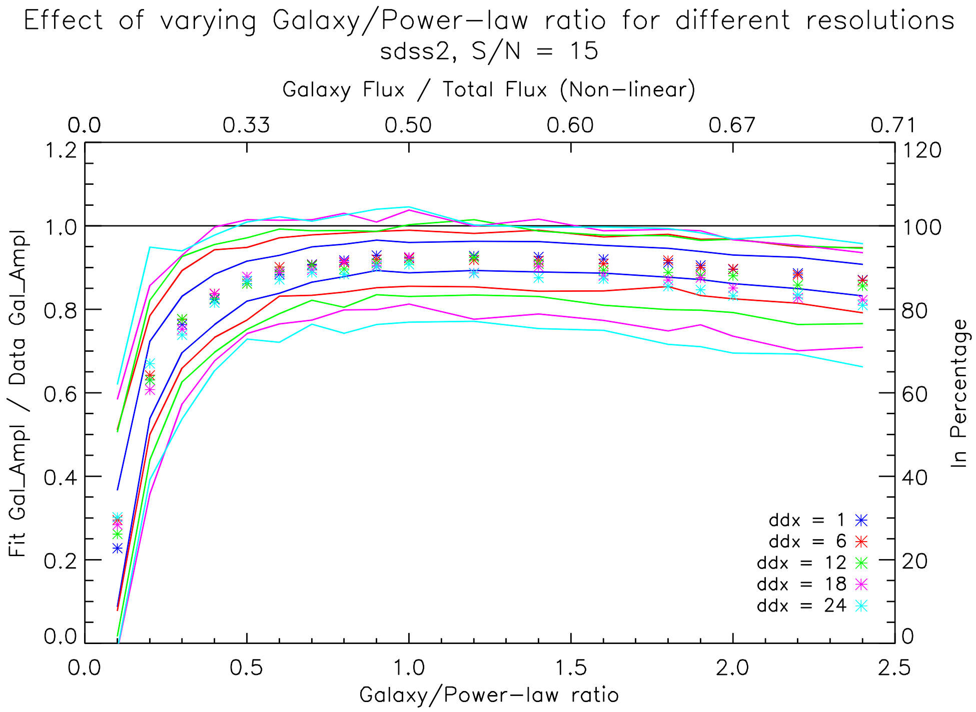



**Figure 4.5**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

The sdss1 plot in Figure 4.6 similarly looks a lot like the previous Figure 4.3, the big difference being that this time the different colored data points are clearly separated and sorted, apart from at low Galaxy/Power-law ratios. Ignoring the cyan colored data points, the other four are very consistently sorted as we would expect, with the highest resolution at the top and the lowest resolution at the bottom. The behavior when varying the Galaxy/Power-law ratio is the same as before, with a slight overestimation at low values turning into a bigger underestimation for high values.

In total, these two plots actually provide what might be the strongest indication of the effect of one of our parameters of all the plots, by quite clearly, particularly for the sdss1 plot, indicating that lowering the resolution causes the fitting routine to underestimate the host galaxy contribution, perfectly in accordance with what we originally expected. Once again, concluding anything in regards to the Galaxy/Power-law ratio is difficult, but a careful recommendation might once again be to avoid data with a very low galaxy contribution.

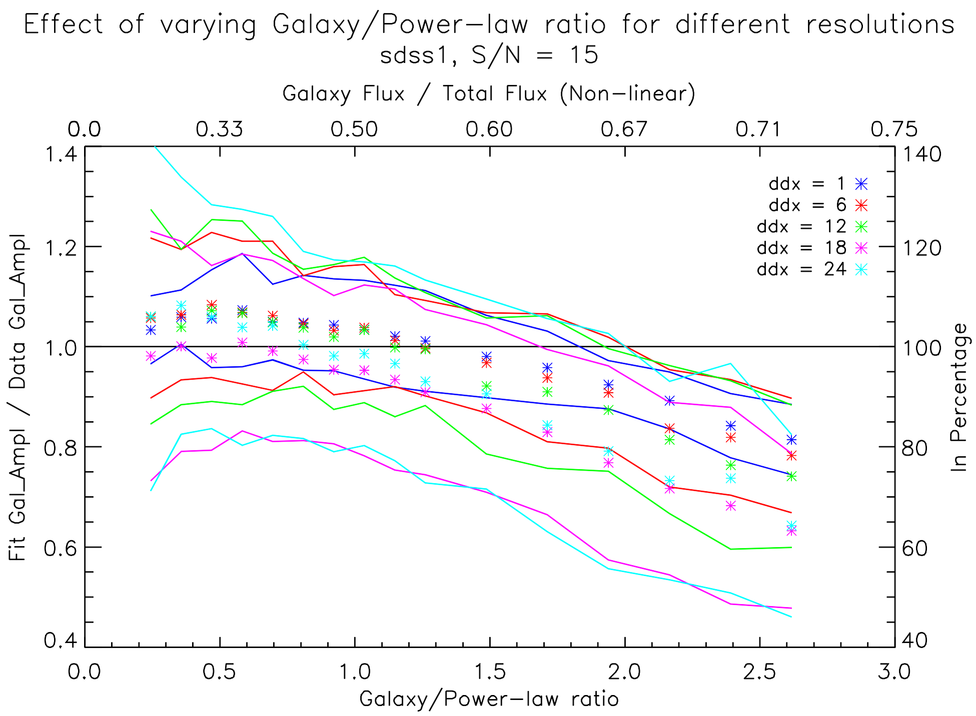



**Figure 4.6**
**Each asterisk represents the average of 200 datasets with the same values of the three parameters**
**The lines depict the average ± the standard deviation of the 200 datasets**

## 4.4. Predictions for NGC5548

Finally, in order to test our experiences from working with the Sloan data, we try to predict the result of doing spectral decomposition on a set of NGC5548 data. As mentioned before, this particular dataset is used because we know very accurately what the host galaxy contribution is for this dataset, and we also know the total flux. Using these values, we can then try to predict what values the fitting routine should return. The resolution of the NGC5548 data is 2 Å per pixel, and the signal-to-noise ratio is about 50. Due to this, the sets of simulated data with the original resolution and an S/N-ratio of 20 are used in the following calculations.

The galaxy amplitude is $4.341 \pm 0.434$ and the total flux is $6.630 \pm 0.243$ (Bentz et al. 2007), and we can then determine the Galaxy/Power-law ratio

$$rgalpow = \frac{gal\_ampl}{f_{total} - gal\_ampl} = 1.896 \pm 0.585 \tag{4.1}$$

The error on this is calculated using error propagation, where $g$ is the galaxy amplitude and $f$ is the total flux

$$\Delta rgalpow = \sqrt{\left(\frac{d}{dg}\left(\frac{g}{f-g}\right)\right)^2 * \Delta g^2 + \left(\frac{d}{df}\left(\frac{g}{f-g}\right)\right)^2 * \Delta f^2} \tag{4.2}$$

We get a Galaxy/Power-law ratio of around 1.90, and this is then the value we will use to choose which of our previous results to use.

The sdss1 set of data with a Galaxy/Power-law ratio of 1.939 then predicts that the fitting routine should return a galaxy amplitude $92.22 \pm 4.14$ % that of the value of the data, equal to a galaxy amplitude of $4.004 \pm 0.444$. Once again the error is calculated using error propagation, and based on the error on the galaxy amplitude of NGC5548 and the standard deviation of the 200 datasets.

For sdss2 we use the set of data with a Galaxy/Power-law ratio of 1.900, and this predicts a galaxy amplitude $90.76 \pm 2.88$ % that of the value of the data, equal to a galaxy amplitude of $3.940 \pm 0.413$. The error is calculated as before.

The actual galaxy amplitude returned by the fitting routine is $3.82$, with an error of $0.03$ that is probably a bit too good to be true, indicating that the fitting routine is overconfident in its ability to determine the galaxy contribution. The four different values are summed in Table 4.1. The result of the spectral decomposition of NGC5548 can be seen in Figure 8.4 in the appendix.

| Source | Galaxy Amplitude | Error | Error % |
|---|---|---|---|
| NGC5548[1] | 4.341 | 0.434 | 10.00 |
| Sdss1[2] | 4.004 | 0.444 | 11.09 |
| Sdss2[3] | 3.940 | 0.413 | 10.48 |
| Fit[4] | 3.82 | 0.03 | 0.52 |

**Table 4.1**
**1) The value of the galaxy contribution for NGC5548, from Bentz et al. 2007.**
**2) The predicted result of doing spectral decomposition of NGC5548, based on sdss1**
**3) The predicted result of doing spectral decomposition of NGC5548, based on sdss2**
**4) The actual result of doing spectral decomposition of NGC5548**

Our predictions of the galaxy amplitude found by the fitting routine for the NGC5548 are at first glance not completely off, although the fit routine underestimates the galaxy amplitude even more than we expected. But considering the 10-11 % uncertainties on our predictions, and the fact that we have only used a single NGC5548 dataset, which our previous experiences indicate might mean the values could be off by something like 10-20% compared to the average we would get from several datasets, it's not really surprising. And the Galaxy/Power-law ratio of the NGC5548 dataset could actually be anywhere from 1.3 to 2.5, and using sets of simulated data with other $rgalpow$ values might have provided different predictions.

# 5. Discussion

In summary, the results obtained from the analysis of the sets of simulated data based on the two Sloan datasets, sdss1 and sdss2, seem to be in disagreement as often as they are in agreement.

Particularly the effects of varying the Galaxy/Power-law ratio are quite different depending on whether we are looking at sets of data based on sdss1 or sdss2. For sdss2 the behavior is in accordance with our expectations, in so far as the fitting routine underestimates the galaxy contribution when it is low compared to the total flux, and seems better at determining the galaxy contribution when it is larger, although there seem to be a consistent underestimation of about 10 %. For sdss1 the behavior is more or less the opposite, with the fitting routine overestimating the galaxy contribution for low Galaxy/Power-law ratios, and underestimating it for high values. The overestimation at low ratios could possibly be because the correction applied to the sdss1 data was not sufficient.

This difference between the two datasets could possibly be at least partially due to the difference in the strength of the Fe emission lines of the two datasets. The sdss2 dataset has very strong emission lines, and the features of this emission might be obscuring the features of the galaxy spectrum. This in turn would explain why the spectral decomposition of the sdss2 dataset yields an initial galaxy amplitude of zero, and why spectral decomposition of the sets of simulated data based on the sdss2 datasets systematically underestimates the galaxy amplitude by around 10 %. The sdss1 dataset on the other hand has very weak Fe emission, probably not obscuring the galaxy spectrum significantly, and this then explains why determining and correcting for the initial galaxy amplitude of the sdss1 dataset was necessary.

The results are better when looking at the effect of varying the signal-to-noise ratio. On this, the results for sdss1 and sdss2 are in agreement on one thing; lowering the S/N-ratio increases the uncertainties and thereby decreases the precision of our measurements. The sdss2 plots also show that for very low S/N-ratios, the accuracy of our measurements is also worsened. A similar effect is not immediately noticeable in the sdss1 plots. These results are in pretty good agreement with our expectations, although we would have thought that lowering the S/N-ratio would have caused the fitting routine to underestimate the galaxy contribution more consistently than what is evident in our data.

The best results are from varying the spectral resolution. For both sdss1 and sdss2 it is quite clear that the effect of lowering the spectral resolution is twofold; the first effect is that, similarly to lowering the S/N-ratio, it increases the uncertainties of our measurements, and the second is that it causes the fitting routine to systematically underestimate the galaxy contribution. This is particularly true for the sets of data based on sdss1, but also for most of the sets of data based on sdss2. And both of these effects are very well in accordance with our expectations.

The outcome of our attempt to predict the result of doing spectral decomposition of the NGC5548 dataset was not very conclusive either, although the result is within the errors on our predictions. In order to really test our predictions, we would have to conduct several tests such as this, using different datasets from different objects.

One thing most of the results have in common though is the quite significant uncertainties that actually make any conclusions based on them shaky at best. For all the plots you could actually sort the different colored data points in the completely opposite way of what we see and of what we expect and still be within the uncertainties. And a lot of the data points could be placed both above and below the actual value of the galaxy contribution and also still be within the uncertainties. In order to make more significant conclusions about the effects of the three parameters then, additional datasets would probably have to be examined in a similar way as the sdss1 and sdss2 datasets.

Based on our results then, a very cautious recommendation for other people working in this field would be threefold. First of all, we would recommend choosing data with a high resolution, in order to have as many and as well defined features of the spectrum as possible for the fitting routine to use. The resolution of the Sloan data of 69 km/s seemed to produce good results, but also the sets of simulated data with a sixth of this, equal to 414 km/s, gave decent results. Second, we would recommend choosing data with a signal-to-noise ratio of no less than 10, and preferably 15 or better. This way, one can avoid the large uncertainties found for data with lower S/N-ratios. And third and finally, when working with data with a low galaxy contribution, and to a certain degree also data with a high galaxy contribution, we would advise people to be aware that there seems to be a systematic underestimation of the galaxy contribution, and that the actual contribution might be a lot bigger than what is found. One should generally keep in mind that there are a lot of uncertainties involved when determining the host galaxy contribution based on spectral decomposition.

# 6. Conclusion

In this thesis we have examined how accurately the host galaxy contribution to the AGN spectrum can be determined, and how it depends on technical properties such as the signal-to-noise ratio and the spectral resolution and more physical properties such as the strength of the Fe emission and the ratio between the galaxy contribution and the total flux. This was done using sets of simulated data based on two datasets from the Sloan Digital Sky Survey, with varying values of the different parameters.

The results are somewhat ambiguous, making it difficult to determine the exact effect of each of the parameters. In general, the results seems to indicate that low values of the signal-to-noise ratio and the spectral resolution results in a low precision of the determined value of the host galaxy contribution, and that using data with strong Fe emission decreases the accuracy of the values of the galaxy contribution. The results are particularly ambiguous on the effect of the ratio between the galaxy contribution and the total flux, and seem highly dependent on the other parameters. There does seem to be a slight indication that when the ratio between the galaxy flux and the flux from the AGN is low, the determination is particularly inaccurate, and when the ratio is high, the fitting routine systematically underestimates the galaxy contribution by around 10 %.

Based on our results we have come up with a few cautious recommendations for other people working with AGNs. At the technical level, we recommend using data with a signal-to-noise ratio no lower than 10 or 15, and to use data with a spectral resolution no larger than around 400 km/s. We also recommend people to be very cautious with the results from spectral decomposition, and expect large uncertainties and possibly a systematic underestimation of the host galaxy contribution of as much as 10 %, and to be particularly careful when dealing with AGNs where the host galaxy contribution is low compared to the total flux.

## 6.1. Future work

The subject we are working with in this thesis has a lot of potential for future work. In order to do a more thorough and conclusive examination of the effect of the parameters we have been working with, one would have to analyze a lot more datasets similarly to what we have done with our two datasets. In doing so it might be a good idea to use objects that have been thoroughly studied beforehand, in order to have a better idea of the different values of the original dataset, that can then be compared to the results of the simulated datasets. And it might also be a good idea to do a more controlled examination of each parameter separately, in order to get a better understanding of the effect of the individual parameter, before trying to understand the effect of combining some or even all of them.

# 7. References

Bentz et al. 2007, ApJ, 662, 205; http://iopscience.iop.org/0004-637X/662/1/205/

Bentz et al. 2009, ApJ, 697, 160; http://iopscience.iop.org/0004-637X/697/1/160/

Bentz et al. 2013, ApJ, 767, 149; http://iopscience.iop.org/0004-637X/767/2/149/

Bradley M. Peterson, *Active Galactic Nuclei*, Cambridge University Press, 1997

Kaspi et al. 2005, ApJ, 629, 61; http://iopscience.iop.org/0004-637X/629/1/61/

Peterson et al. 2004, ApJ, 613, 682; http://iopscience.iop.org/0004-637X/613/2/682/

Urry & Padovani, 1995, PASP, 107, 803; http://adsabs.harvard.edu/abs/1995PASP..107..803U

Watson et al. 2011, ApJ, 740, L49; http://iopscience.iop.org/2041-8205/740/2/L49

W. Kollatschny, M. Zetzl 2013, arXiv:1301.7704; http://arxiv.org/abs/1301.7704

# 8. Appendix

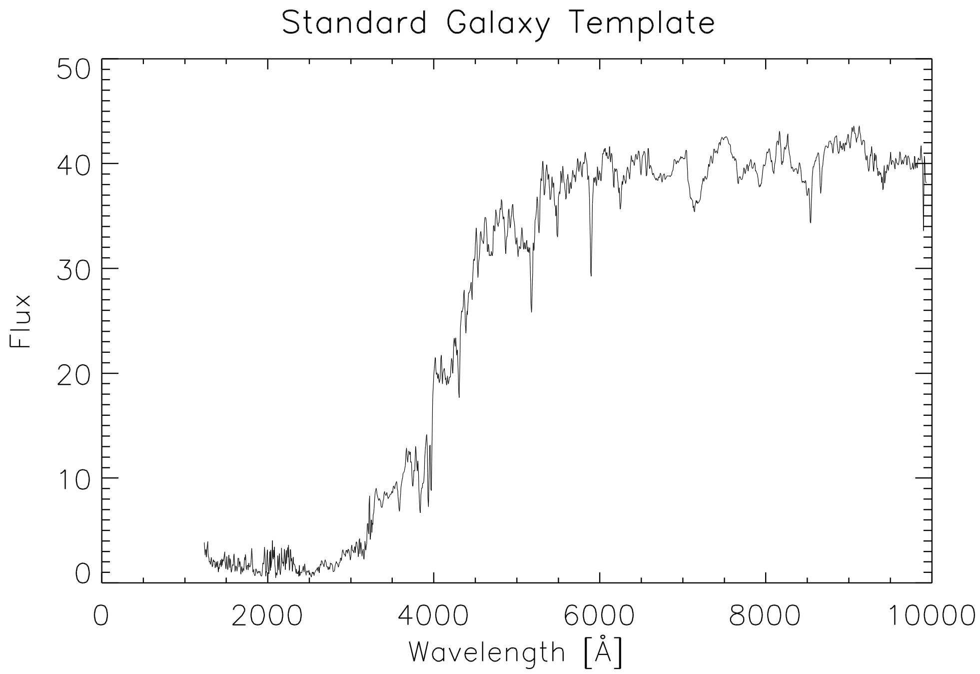



**Figure 8.1**

**The standard galaxy template used to add galaxy flux to the data.**

**The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ Å^{-1}$**

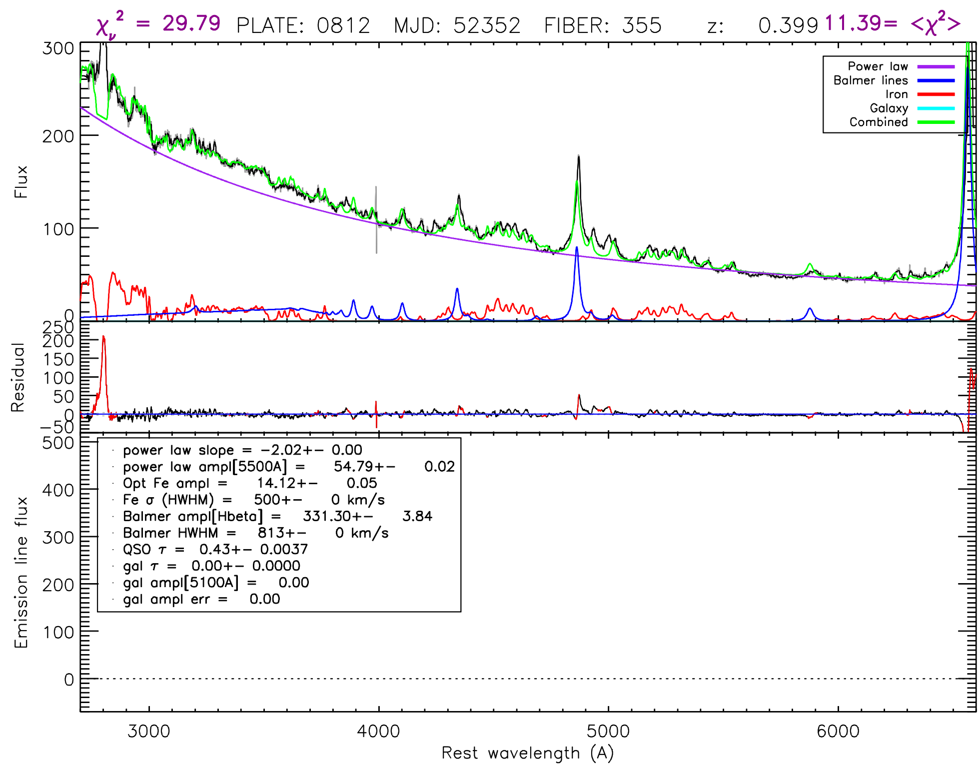



**Figure 8.2**

**The result of spectral decomposition of sdss2**

**The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ Å^{-1}$**

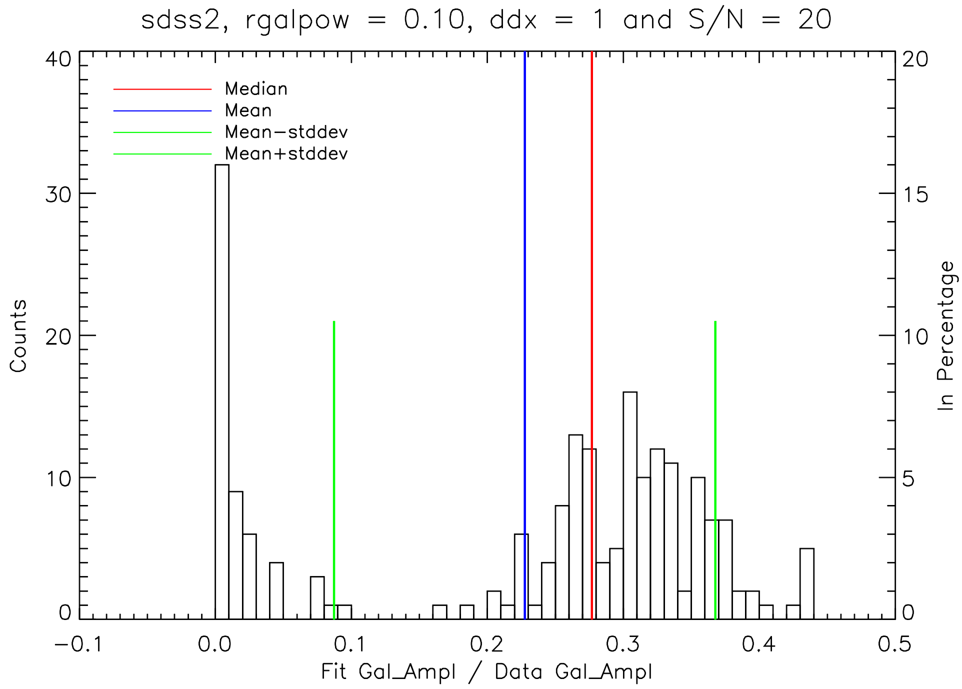



**Figure 8.3**
**Histogram of the galaxy amplitude found for the 200 datasets with the particular values of our three parameters described in the title**

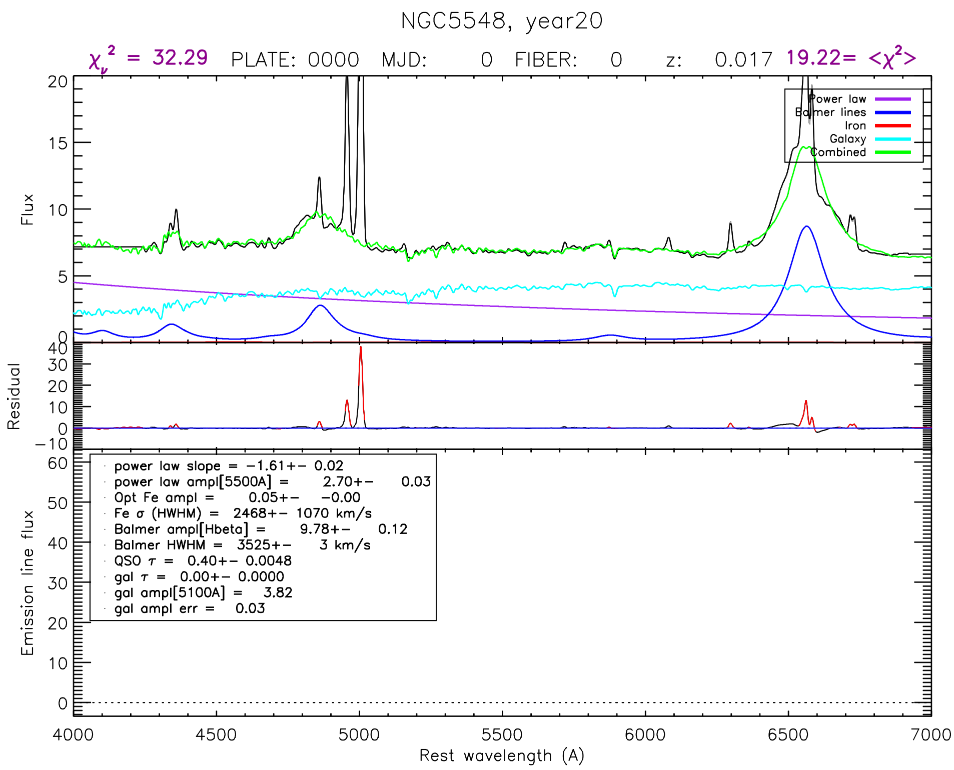



**Figure 8.4**
**The result of spectral decomposition of NGC5548**
**The flux is in units of $10^{-15}\ erg\ s^{-1}\ cm^{-2}\ \text{Å}^{-1}$**