UNIVERSITÉ PARIS-SUD XI

Ecole doctorale « Astronomie et Astrophysique d'Île-de-France »

Discipline : *Physics*

DOCTORAL THESIS

Defence 13th January 2015

by

DANIEL MACHADO

Subject :

Improving automated redshift detection in the low signal-to-noise regime for Large Sky Surveys

Members of the Jury :

Mr.	J.-L. STARCK	Research Director of Cosmostat, CEA	Primary Thesis Supervisor
Mr.	F. B. ABDALLA	Reader at University College London	Secondary Thesis Supervisor
Mr.	G. P. DES FORÊTS	Professor at IAS, Université Paris-Sud	Examiner
Ms.	S. MAUROGORDATO	Researcher at Observatoire de la Côte d'Azur	Examiner
Mr.	J. G. BARTLETT	Professor at Paris-Diderot University	Reviewer
Mr.	F. MURTAGH	Professor at De Monfort University	Reviewer

I look up into the night sky, I see a thousand eyes staring back,

And all around these golden beacons, I see nothing but black.

I feel the weight of something beyond them, I don't see what I can feel,

If vision is the only validation, then most of my life isn't real.

– Sam Sparro – Black & Gold

Abstract

Redshift is the primary measure by which astronomers can map the Universe in the radial direction. In order to test the assumptions of homogeneity and isotropy, accurate redshifts of galaxies are needed, and for a great many of them. Additionally different cosmological models can only be distinguished by careful observations of the large scale structure traced by these galaxies. Large sky surveys are the only mechanism by which redshifts for a large number of galaxies can be obtained. Accurate redshift estimation is additionally required for many other fields of astronomy including but not limited to: weak lensing, studies of dark matter haloes, galaxy morphology studies, chemical evolution studies, photometric calibration, and studies of large scale structure and galaxy clustering.

Problems exist in all surveys at the dim limit of observation, which usually corresponds to the higher redshift objects in the survey, where noise becomes problematic. Magnitude or signal-to-noise ratio cuts are often employed in order to eliminate potentially troublesome objects; such a procedure is a blunt tool for separating good redshift candidates from ones likely to be inaccurate.

In this thesis we develop an algorithm to tackle redshift estimation of galaxy spectra in the low signal-to-noise regime. The first part of this thesis introduces the concepts of denoising, particularly False Detection Rate denoising, wavelet transforms and redshift estimation algorithms. The second part details how these concepts are united into the Darth Fader (**d**enoised and **a**utomatic **r**edshifts **th**resholded with a **fa**lse **de**tection **r**ate) algorithm. The final parts of this thesis apply the algorithm both to idealised synthetic data generated from the COSMOS Mock Catalogue, and to a subset of real data from the WiggleZ survey.

We show that Darth Fader can operate effectively at low signal-to-noise given an appropriate choice of FDR parameter for denoising, and an appropriate feature-counting criterion. We also show that Darth Fader can remove the continua of spectra effectively at low signal-to-noise for the purposes of redshift estimation by cross-correlation. Additionally we show from tests on spectra from the WiggleZ survey that our algorithm has the ability to process a substantial subset of that data without the need for visual inspection (to which the entire WiggleZ spectral survey has been subjected), and to a high degree of accuracy. We conclude that the Darth Fader algorithm has potential to be used in large-sky survey pipelines, particularly where signal-to-noise is expected to be poor.

Acknowledgements

As I look back upon this doctoral thesis, it is with a sense of pride and relief that I have finally reached the end of it. I am reminded of my time in France, and the people whom, without their help, this thesis could not have been completed.

First and foremost, a thank you must go to Jean-Luc Starck and Filipe Abdalla, who have supervised me throughout this thesis, given me the opportunity to visit new places over the course of it, and have had to put up with me during that time! I couldn't have done it without them.

A special mention must also be given for Adrienne Leonard, without her significant contribution, astute guidance and ongoing encouragement, this thesis could not have been completed. Additionally sharing an office with me probably deserves an award in of itself! Thank you.

An additional thank you must go to Sandrine Pires, Paniez Paykari, François Lanusse, Simon Beckouche and Jérémy Rapin for making me feel welcome in France and at the CEA.

There are too many people to mention individually and if I were to try and mention everyone I'd invariably miss someone out, so I thank all my friends and colleagues that I've made during the course of this doctorate, particularly other colleagues and friends at the Service d'Astrophysique, CEA and all the great people of the Astronomy department at UCL.

Two more people I have to thank are Gavin Kirby for being there for me, and Hernán Furci for being a good friend and neighbour.

A final, big, thank you to my mum and dad for their continued support, both during the ups, and the downs.

List of Figures

Contents

Introduction

This thesis focuses on the study of redshift estimation of galaxies, and in particular the estimation of redshifts from galactic spectra in cases where the presence of noise is extensive. An empirical algorithm for this task is developed and tested, both against idealised simulations and real data from a spectroscopic sky survey.

This document is organised as follows:

⋆ In chapter 1 the current state of affairs in Cosmology is described. The concept of redshift is derived from Einsteinian General Relativity, homogeneity and isotropy considerations. The importance of redshift for tracing the history and evolution of the Universe over time, and constraining Cosmology, is highlighted. Differences in galaxy types and how these differences in morphology translate into differences in observations and affect subsequent analysis is explained.

⋆ Chapter 2 introduces the concepts of noise present within natural signals and the process of its removal (denoising), with particular attention paid to False Detection Rate (FDR) denoising. Additionally the concepts of sparsity and wavelet transforms are introduced, initially from an analogical viewpoint with respect to the more familiar Fourier transform. Particular attention is given to Discrete Wavelet Transforms as these form a significant part of later analysis. The usefulness of wavelet transforms for signal processing is highlighted.

⋆ The focus of chapter 3 is that of the primary methods for redshift estimation for different types of survey data: photometric and spectroscopic. The need for automatic algorithms to analyse these data is highlighted. Synthetic catalogue generation is investigated, and the COSMOS Mock Catalogue (CMC, which we use in further analysis) is described.

⋆ The development of the Darth Fader algorithm is the focus of chapter 4. Attention is drawn to the use of wavelet transforms for empirical continuum removal, and FDR denoising for empirical detection of likely true features. Worked examples are given for the same spectrum under white Gaussian noise regimes with different signal-to-noise ratio (SNR) values.

⋆ Results from simulated datasets generated from the CMC are presented in chapter 5. The impact of varying the levels of (white Gaussian) noise, and the varying of the FDR parameter, α, on three key statistics - catastrophic failure rate, retention rate, and capture rate - is investigated. Variable noise per pixel, in the form of an instrumental response from a realistic instrument, is additionally investigated as an attempt to have the simulated data approach more realistic data.

⋆ The Darth Fader algorithm is applied to real data in chapter 6. Initially a proof of concept on high SNR spectra from SDSS is shown. Lastly the algorithm is applied to (a subset of) real

spectra from the WiggleZ survey (which has a lower overall SNR), and the effects of varying the FDR parameter and peak-counting criterion is investigated.

Chapter 1

Background

Summary

1.1 Birth of Cosmology

Nothing in the Universe is stationary, from the coldest subatomic particles, through to planets, stars and even galaxies. This fact was lost on early astronomers who thought of only one galaxy – the Milky Way, existing within a sphere of fixed (background) stars, with some odd moving 'stars' that they called *'planetes'*.

Initially, Astronomy had the use of only one primary tool – the human eye; which, unfortunately, is far from optimal for observing the heavens. Any further breakthroughs in the subject, other than the cataloguing of various objects, had to await the advent of new technology; and it was Galileo Galilei whom, in 1609, brought that technology into Astronomy, changing the field forever – it was of course, the telescope. It was Galileo himself who in 1610, provided the definitive observations that the Milky Way was a collection of stars.

However, even with the advent of the telescope, it still required a considerable length of time before the next revolution in Astronomy took place. The first step was that of Fr. Secchi (1877), who contributed significantly to the field of astronomical spectroscopy, allowing us to understand that the Sun is indeed a star, and that different stars in the galaxy have different spectral absorption lines and thus different constituent elements and characteristics. Slipher (1915) then applied similar spectroscopic techniques to various visible 'nebulae' (now known to be galaxies, but at the time thought to be satellite nebulae of the Milky Way), showing that, in many cases their spectra were indicative of large recessional velocities, and thus these 'nebulae' were unlikely to be gravitationally bound to the Milky Way. Indeed, it was not until Hubble (1925), using the then largest telescope in the world, at Mount Wilson Observatory, that these 'nebulae' were conclusively shown to be *other* galaxies, and far more distant than had previously been imagined. Hubble had provided the world with the first definitive evidence suggesting that the Universe was far larger than just our own Milky Way galaxy. Later, in 1929 Hubble proposed from observations that (virtually) all of these galaxies were moving away from us, and in particular, it was the distance to the galaxy that was directly proportional to how quickly it would appear to be receding – and observational cosmology was born.

1.1.1 Redshift

What Hubble and Slipher had discovered was that the majority of galaxies had spectra that were *redshifted* with respect to standard reference spectra in the lab. The common way of describing a redshift, is through the Doppler effect (Doppler 1842). The Doppler effect occurs when a source of waves (light or sound) moves relative to an observer. If, in the rest frame of the source, the waves are being emitted at a constant wavelength, then to an observer the received wavelength will be distorted. How the wavelength is distorted is dependent upon the relative motion being toward or away from the source.

If the motion is toward the source, then the waves are observed to be compressed, and appear to have a shorter length, relative to someone in the rest frame of the source. If the source were emitting monochromatic visible light, the observed light would be skewed towards the blue end of the spectrum, it would be *blueshifted* (assuming the relative motion between source and observer is large enough). Conversely, if the relative motion is away from one another, the observed spectrum would be *redshifted*. This effect is represented by figure 1.1.

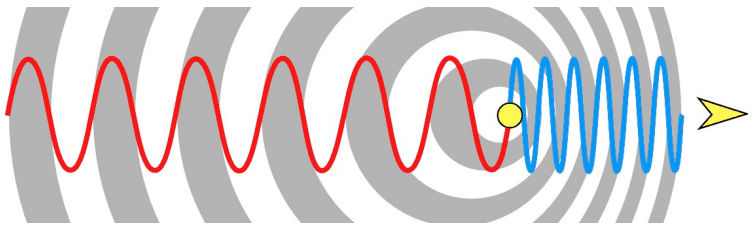


Figure 1.1: Doppler effect – The diagram represents a source (yellow) moving to the right, causing compression of the waves in front of its path, and extension of the waves behind it. These correspond to a blueshift to an observer situated in front of the path and redshift to and observer situated behind the path, respectively. Waves emitted perpendicular to the relative motion of the source are unaffected. Adapted from http://m.teachastronomy.com/astropediaimages/Doppler_effect_diagrammatic.png.

The kind of redshift that Hubble had discovered, was *not* a Doppler shift, however the conclusion that galaxies were moving apart was still a valid one; these effects have a different origin altogether, within the theory of Relativity.

1.1.2 Special Relativity

Observational cosmology was not of course, developed independently of theory, and parallel theoretical developments emerged alongside, the most significant and enduring of which were from Einstein in his groundbreaking work on both Special and General Relativity (Einstein 1905, 1915).

Starting with a few postulates, Einstein put forward a model in which space and time no longer retained their absolute nature as everyday experience might suggest, instead they are intimately linked, malleable and dynamic. The invariance of the laws of Physics in non-accelerating frames of reference, and the constancy of the speed of light for all observers were these two postulates.

The first of these postulates proposes that all inertial frames of reference are in constant rectilinear motion with respect to one another and are equivalent (there is no overall 'stationary' or preferred reference frame). An inertial reference frame can be thought of as a coordinate system within which events occur, such that they can be completely pinpointed by a set of spatiotemporal coordinates; in addition an accelerometer at a fixed position in any inertial frame would detect zero acceleration. Implicit in this postulate is the assumption that the Universe does not have a preferred direction (since the rectilinear motion of an inertial frame can be in any direction), and that its properties remain the same everywhere (since all inertial frames are equally valid and are relatable via Lorentz transformations irrespective of where they are), these properties are termed isotropy and homogeneity respectively. Experiment (Michelson and Morley 1887) and observation (Zeldovich et al. 1982; de Lapparent et al. 1986) have both been found in support of isotropy and homogeneity (and in the case of homogeneity, on large enough scales such that galaxy clustering is smoothed out).

The second of these postulates states that the speed of light (in a vacuum) is constant and measured to have the same value for all (inertial) observers. In order for the measurement of the speed of light to yield the same result in different inertial reference frames, the intuitive (but incorrect) assumptions of absolute space and absolute time must be abandoned, and thus measured lengths and times will be tied to an individual's frame of reference.

To illustrate an example of relative time, a thought experiment is considered where a hypothetical 'light clock' bounces a photon between two perfectly reflecting mirrors in order to track the passage of time, and is observed by two independent observers, A and B. When the clock and the observer (A) are at rest with respect to one another, a photon travels between the mirrors, traversing a length L, in a time, $t_A = {}^{L}/_{c_A}$. A second observer (B) moving with a fixed velocity, v, relative to A (parallel to the mirrors of the clock) observes the clock to move as the photon bounces between the mirrors, B thus sees the photon traverse a longer, diagonal path, whose distance is D, traversing it in a time $t_B = {}^{D}/_{c_B}$. In this time span, observer B sees the clock travel a linear distance of vt_B, hence the distance D represents the hypotenuse of a right angled triangle whose other sides measure vt_B and L, thus $D = \sqrt{(vt_B)^2 + L^2}$. Postulating the constancy of the speed of light for all inertial observers implies that $c_A = c_B (= c)$. This agreement on c additionally shows that t_A and t_B – both measures of time for the photon to bounce between the same two mirrors – cannot be the same for both observers, since the distances traversed are not equal. The time measured for both observers, for the same process – a photon bouncing between two mirrors – will be different dependent upon their relative motion. Substituting for D and L, and letting $t_A = t$, and $t_B = t'$, allows the relationship between the two measured times to be quantified, yielding the equation for time dilation,

$$t' = \frac{t}{\sqrt{1 - \frac{v^2}{c^2}}}\,. \tag{1.1}$$

Similar considerations can be employed to reach conclusions of length contraction in addition to time dilation; space and time are no longer immutable concepts, but intimately linked, changing to relative observers such that they disagree on lengths and durations but always agree on the measured value of the speed of light. The definition of 'distance' between two points has to change as a consequence of this, and it must also include a temporal component. Recall the definition of an arbitrary distance element, dl, in Euclidean 3D space and Cartesian coordinates, representing the length of the shortest path between the origin and destination (these points also being arbitrary),

$$dl = \sqrt{dx^2 + dy^2 + dz^2}\,. \tag{1.2}$$

Due to length contraction (and time dilation), this distance will not always be observed to have the same value between different observers in different inertial frames from a relativistic perspective, and as such is no longer a useful definition. In Relativity the concept of path and its associated distance in a Euclidean space is replaced with that of a *geodesic* (which is the shortest journey between two points in spacetime) as measured by a *metric* which, within a (flat) Minkowski (1907) spacetime, is as follows,

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2\,, \tag{1.3}$$

where the convention is $(-+++)$ and that the time coordinate is the 'zeroth' coordinate[1]. It is this quantity that replaces the traditional concept of distance, and will be measured to be the same for all inertial observers. The Minkowski metric[2], η, is therefore often described in matrix form as,

$$\eta = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\,. \tag{1.4}$$

equation (1.3) can be adapted to spherical polar coordinates with a standard change of variables from (x, y, z) to (r, θ, ϕ) (the temporal coordinate remains unchanged), yielding,

$$ds^2 = -c^2dt^2 + dr^2 + r^2\left[d\theta^2 + \sin^2\theta\, d\phi^2\right]\,. \tag{1.5}$$

This definition of the metric lends itself better to describing homogeneity and isotropy than one expressed in (x, y, z) coordinates, however, it still represents a flat and static spacetime.

1.1.3 General Relativity

Einstein had developed Special Relativity in full knowledge that it was only valid for the special case of inertial reference frames (hence the 'special' in Special Relativity), and over the next decade, extended the idea of a unified spacetime and the postulates of relativity for the more general case of

[1]The sign convention of $(-+++)$ is not the only possible formalism, and a choice of $(+---)$ is equally valid but does not affect the mathematics. In addition the time coordinate being the first coordinate (labelled with 0) is an arbitrary (though customary) choice.

[2]It is common practice, for convenience, to normalise time coordinates such that $c = 1$, and hence the first entry in equation (1.4) would be -1.

non-inertial reference frames (hence 'general' in General Relativity, or simply GR). The basic principle in GR is the equivalence between inertial mass and gravitational mass being due to acceleration and gravitational force being fundamentally indistinguishable - being at rest in a gravitational field with a uniform strength, g, is equivalent to acceleration, a, in the absence of a gravitational field (ie: an inertial frame), provided the magnitude of a & g are the same.

This equivalence between gravitational and accelerative forces, when combined with the model of Special Relativity, leads to an interesting conclusion: all acceleration and gravitation are a result of smooth non-linear transformations between a succession of locally inertial frames – in other words, curvature of spacetime. The behaviour of spacetime was succinctly described by John Wheeler (Wheeler and Ford 1998):

> "Spacetime tells matter how to move; matter tells spacetime how to curve."

The initial formulation of GR was expressed by Einstein as follows,

$$\mathbf{G}_{\mu\nu} = \frac{8\pi G}{c^4}\mathbf{T}_{\mu\nu}\,, \tag{1.6}$$

where $\mathbf{G}_{\mu\nu}$ is the Einstein tensor,

$$\mathbf{G}_{\mu\nu} = \mathbf{R}_{\mu\nu} - \frac{1}{2}R\,\mathbf{g}_{\mu\nu}\,, \tag{1.7}$$

and $\mathbf{R}_{\mu\nu}$ is the Ricci curvature tensor; R is the scalar curvature; $\mathbf{g}_{\mu\nu}$ is the metric tensor; G is Newton's gravitational constant, c is the speed of light in a vacuum and $\mathbf{T}_{\mu\nu}$ is the stress-energy tensor.

The Einstein tensor as a whole, describes the curvature of spacetime in a way that is consistent with energy considerations. The stress-energy tensor can be considered to be a quantity that describes both the energy and momentum density and flux for a region of spacetime.

Einstein then, following a similar prescription in equation (1.5) proposed an initial, homogeneous and isotropic model for a static Universe,

$$ds^2 = -c^2dt^2 + dr^2 + R^2\sin^2\left(r/R\right)d\Omega^2\,, \tag{1.8}$$

where R is the radius of curvature, and the angular part is condensed into a single term $d\Omega^2 (= d\theta^2 + \sin^2\theta\, d\phi^2)$.

The motivation for a static Universe at the time was based on very limited observational evidence particularly since the confirmed existence of other galaxies was not yet established, and the internal motions of the Milky Way suggested a contained and bounded existence for the Universe (which at the time was limited to the Milky Way plus satellite nebulae). Einstein found that the simple model he proposed was not a solution to the very set of equations he had derived (equation (1.6)). Einstein subsequently modified his equations for GR to include an additional term – the Cosmological Constant – such that a static Universe could be maintained,

$$\mathbf{G}_{\mu\nu} + \mathbf{\Lambda}\,\mathbf{g}_{\mu\nu} = \frac{8\pi G}{c^4}\mathbf{T}_{\mu\nu}\,, \tag{1.9}$$

where $\mathbf{\Lambda}$ is the cosmological constant, and the symbol and name are used interchangeably.

Einstein's original intention for introducing the cosmological constant was purely for maintaining a static Universe, and one which he later retracted on the observations and conclusions of Hubble (1929). The basic principle is that empty space itself can contain vacuum energy which by its nature exerts a 'negative pressure', causing expansion of the Universe if left unchecked, however, if this were in precise

opposition to the gravitational attraction exerted by the matter and energy content of the Universe, then it would remain static over time. Such a situation would be unstable however, the energy density of the vacuum is constant, whereas the energy density of the photon and matter components to the Universe are not, and hence a slight expansion from equilibrium would quickly result in $\mathbf{\Lambda}$ dominance, similarly a slight contraction from equilibrium would result in matter dominance. Whilst there is no evidence to suggest a cosmological constant such as the one Einstein had proposed, there is however no physical motivation for the cosmological constant to be zero either.

1.1.4 The Friedmann–Lemaître–Robertson–Walker Model

The Friedmann–Lemaître–Robertson–Walker (FLRW) model is an exact solution to Einstein's field equation (1.9) that describes an homogeneous and isotropic Universe undergoing simple, metric expansion (or contraction). Developed initially by Friedmann (1922, 1924) and later, independently, by Lemaître (1927) the model was originally developed in the absence of any cosmological constant (but generalisable to be solutions inclusive of $\mathbf{\Lambda}$), and as such naturally described a dynamic Universe that was either in contraction or expansion, but not static. Further contributions, including a proof of uniqueness for the metric under the criteria of isotropy, homogeneity and curvature, were developed by Robertson (1935, 1936a,b) and Walker (1935). The resultant metric is similar to one obtained by Einstein for his static model (equation (1.8)), with the key difference being the inclusion of a new term, $a(t)$, called the *scale factor* and a curvature term $S_k(r)$,

$$ds^2 = -c^2dt^2 + a^2(t)\left[dr^2 + S_k^2(r)\, d\Omega^2\right], \tag{1.10}$$

where $S_k^2(r)$ is a term dependent upon curvature, $a(t)$ is the (dimensionless) scale factor, and all other terms are identical to equation (1.8). $S_k(r)$ is a term that is dependent upon curvature, where k represents the type of curvature involved, as follows,

$$S_k^2(r) = \begin{cases} R_0^2 \sin^2(r/R_0), & \text{if } k = +1, \quad \text{(Closed)}, \\ r^2, & \text{if } k = 0, \quad \text{(Flat)}, \\ R_0^2 \sinh^2(r/R_0), & \text{if } k = -1. \quad \text{(Open)}. \end{cases} \tag{1.11}$$

where R_0 is the radius of curvature where $a(t_0) = 1$, with t_0 taken to be the present.

It should be clear that the scale factor a, in order to preserve both homogeneity and isotropy can at most, be a function of time only, since to be otherwise would imply a dependence upon position and/or direction, with the sign of a associated to expansion if positive and contraction if negative[3]. Important to note is that the value of k determines the curvature that is applicable to the space part of the metric only, hence a flat model for this metric does not mean a flat *spacetime*, but merely a Euclidean behaviour (i.e.: flat) for the space part of the metric. Indeed setting $k = 0$ yields equation (1.5) with the addition of the extra term $a(t)$, with this model describing a Minkowski-like metric where the (infinite) space part expands (or contracts) uniformly and identically at all points, with this being termed *metric expansion*. Similarly setting k=1 yields equation (1.8) with the incorporation of the scale factor, this describing a Universe in which space is finite (and initially parallel light rays eventually converge) and undergoes metric expansion in an identical way. Setting $k = -1$ yields a hyperbolic (saddle) shape to the space part of the metric, and in this space initially parallel light rays eventually diverge, it retains the property of metric expansion.

[3]For a negative time, positive space convention.

A useful analogy to picture the metric expansion of space is to imagine uniformly spaced points on the surface of a balloon. For any choice of point on the surface, all the other points become progressively further away as the balloon is inflated. This is true for **any** of the points chosen, and the effect is greatest for the most distant points from the one chosen. The points, being stuck to the surface of the balloon do not move relative to its surface (i.e.: across its surface) – they are said to comove with the expanding surface of the balloon. In the same way that the expansion is a property of the balloon's surface, metric expansion is an intrinsic property of the space itself, and not the constituents within it.

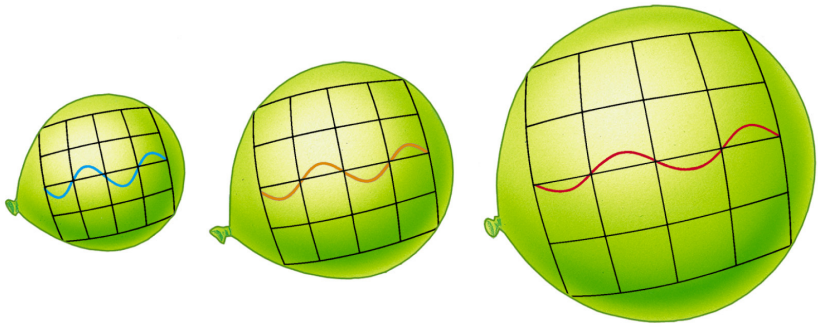


Figure 1.2: Balloon Analogy — A simple analogy of an expanding Universe as a balloon, with light being stretched from blue to red, thus becoming redshifted. All comoving points with respect to the surface of the balloon during its expansion and become progressively, simultaneously, more distant to one another. The real Universe of course has higher dimensionality than this analogy. Adapted from http://www2.astro.psu.edu/users/raw25/astro1h09/Images/FG17_04.JPG.

1.1.5 Cosmological Redshift

As mentioned previously in section 1.1.1, the type of redshift Hubble had found was not strictly a Doppler redshift, but instead it was a cosmological redshift, and whilst these are observationally similar[4] they have a profoundly different origin. Cosmological redshift has at its core, the metric expansion of space as described in section 1.1.4.

Hubble (1929) proposed that observed (recessional) velocity of (nearby) galaxies was directly proportional to the distance they were away from us, proposing the equation $v = H_0\, d$. Hubble had proposed this law on the basis of nearby galaxies and thus the Hubble parameter, H_0 is constant over that region, in general, however, the Hubble parameter is time-dependent. This is a consequence, not of galaxies moving through a static space, but galaxies moving *along* with a dynamic, expanding space (i.e.: 'embedded' and comoving with the metric expansion of space).

Consider the FLRW metric in equation (1.10) for the case of a comoving point (where a galaxy may be situated) at a radial coordinate r (such that $d\Omega^2 = 0$), we can define the proper distance, D_p as the physical distance in an instant of time between the observer at that point; this is equivalent to what a ruler would measure as an instantaneous measurement between these two points (i.e.: $dt = 0$ in equation (1.10)). This is defined as follows,

[4] In fact, in the case of a single observation these two origins for redshift would be indistinguishable. It is only via the multitude of redshifted objects, each with redshifts in proportion to their distance, that allow us to distinguish between the two interpretations.

$$D_p = \int ds = \int_0^{r'} a(t)\, dr = a(t)\, r \quad , \tag{1.12}$$

where r represents the comoving distance to the point, and via the definition of $a(t_0) = 1$. This comoving distance (expansion independent), is defined to be equal to the proper distance (expansion dependent) today (i.e.: at t_0). The proper distance is thus variable, dependent upon at what time the measurement is taken.

In general, the scale factor is time-dependent, and as such the proper distance will also have a time dependence as follows,

$$\dot{D}_p = \dot{a}\, r\,. \tag{1.13}$$

Hence the change in the proper distance with time can be related to the proper distance at the present time via,

$$\dot{D}_p = \frac{\dot{a}}{a}\, D_p\,. \tag{1.14}$$

equation (1.14) is the same as the one Hubble proposed if we make the association that the change in the proper distance over time is simply the recessional velocity of the galaxy, and the the Hubble parameter, H is equal to $\dot{a}/a$.

In addition it is possible to relate the redshift relation to the scale factor by considering once again the FLRW metric, with an object at a distance r, emitting two pulses of light in quick succession, noting that photons travel on paths in spacetime such that $ds = 0$, and again setting $d\Omega^2 = 0$. Defining the time the first pulse is emitted as t_e and the second pulse at $t_e + \delta t_e$, and the associated observations as t_o and $t_o + \delta t_o$, we obtain,

$$c^2 dt^2 = a^2(t)\, dr^2\,,$$

and hence, recalling that r is the comoving distance and as such it does not change with expansion,

$$r = \int_{t_o}^{t_e} \frac{c}{a(t)}\, dt = \int_{t_o+\delta t_o}^{t_e+\delta t_e} \frac{c}{a(t)}\, dt\,,$$

which becomes (since we can switch the limits of integration),

$$\int_{t_e}^{t_e+\delta t_e} \frac{1}{a(t)}\, dt = \int_{t_o}^{t_o+\delta t_o} \frac{1}{a(t)}\, dt\,. \tag{1.15}$$

Next, we make the assumption that these pulses are close enough together such that any change in $a(t)$ over their duration is small, and as such equation (1.15) can be evaluated as follows,

$$\frac{\delta t_e}{a(t_e)} = \frac{\delta t_o}{a(t_o)}\,,$$

$$\frac{\delta t_o}{\delta t_e} = \frac{a(t_o)}{a(t_e)}\,. \tag{1.16}$$

Recall that the redshift, z, is defined as follows,

$$1 + z = \frac{\lambda_o}{\lambda_e}\,, \tag{1.17}$$

where λ_e is the emission wavelength in the rest frame of the source, and λ_o is the wavelength of the same light as observed in the rest frame of the observer. Hence, combining equations (1.16) and (1.17), it can be seen that the redshift is related to the scale factor,

$$1 + z = \frac{\lambda_o}{\lambda_e} = \frac{c\,\delta t_o}{c\,\delta t_e} = \frac{a(t_o)}{a(t_e)} = \frac{1}{a(t)}\,, \tag{1.18}$$

if we maintain the definition that $a(t_o)$ is the present day and defined to be 1, and note that our choice of $a(t_e)$ was an arbitrary one and hence can be taken to have been at any time t.

It is important to note that objects in the Universe (i.e.: galaxies and galaxy clusters) are *dynamic* and not-necessarily perfectly comoving objects, in general they will have a composite velocity with two individual components, the comoving velocity due to the expansion of the Universe (with space), and velocities in response to their local gravitational environment, termed their *peculiar* velocity, v_{pec} (through space). This can in principle complicate the redshift measurement, recalling the definition of (cosmological) redshift from equation (1.17), and noting that the difference between the emitted wavelength we would assume it to have, λ_e and its true emitted wavelength, λ'_e is related by $1 + {}^{v_{pec}}/_{c}$, hence the measured redshift can be defined as,

$$1 + z_{meas} = \frac{\lambda_o}{\lambda'_e} = \frac{\lambda_o}{\lambda_e}\frac{\lambda_e}{\lambda'_e} = \left(1 + z_{cosm}\right)\left(1 + \frac{v_{pec}}{c}\right) = (1 + z_{cosm})(1 + z_{pec})\,, \tag{1.19}$$

where z_{pec} is defined to be ${}^{v_{pec}}/_{c}$. For greater cosmological distances, the recessional velocity due to the metric expansion of space will dwarf any peculiar velocities due to the local gravitational environment and thus $z_{meas} \approx z_{cosm}$.

1.1.6 Convergence on the ΛCDM Model

Cosmic Microwave Background

As soon as the expanding Universe model had been generally accepted after the observations of Slipher (1915); Hubble (1925, 1929), it became clear from a theoretical point of view that returning to points earlier in the history of the Universe would imply that the Universe would have been smaller in the past. Taking this basic idea to its natural conclusion, it was realised by Lemaître (1931) that at some point in the distant past, all the matter content of the Universe must have been much closer and therefore denser, increasing the frequency of collisions between particles and raising the temperature. The expansion of the Universe ought to have had a beginning as an extremely dense plasma of energetic subatomic particles and photons - the Big Bang[5].

An important outcome of such a beginning to the Universe would have been a background of radiation from all directions (i.e.: isotropic), with said background radiation possessing a blackbody spectrum relevant to the uniform (i.e.: homogeneous) thermal temperature at which it was emitted and to which it subsequently cooled as the Universe expanded. The origin of this phenomenon is the expansion of the Universe to such a point were the energy density of the plasma it contained dropped sufficiently so as to make it energetically favourable for electrons to couple to protons, thereby decoupling matter and radiation (photons), and allowing the Universe to become transparent to photons for the first time. Such a phenomenon, now called the Cosmic Microwave Background (CMB),

[5]The term was not itself used by Lemaître, and is said to have originated from Fred Hoyle - a prominent Steady-State cosmologist - in a 1949 radio broadcast.

was predicted by Alpher and Herman (1948) and correctly identified in observations by Penzias and Wilson (1965)[6]; the temperature of which has been measured to be $2.72548 \pm 0.00057\,\mathrm{K}$ (Fixsen 2009, for the WMAP instrument).

Fluctuations at the quantum scale, during the phase of decoupling of CMB photons from matter, are expected to have been the progenitors of the first over-densities of the Universe (via initial rapid growth called 'inflation' (Guth 1981)), seeding the first dark matter haloes and hence the first compact structures. These tiny fluctuations ($\sim 10^{-5}$) are indeed seen in the map of the CMB sky, as in figure 1.3,

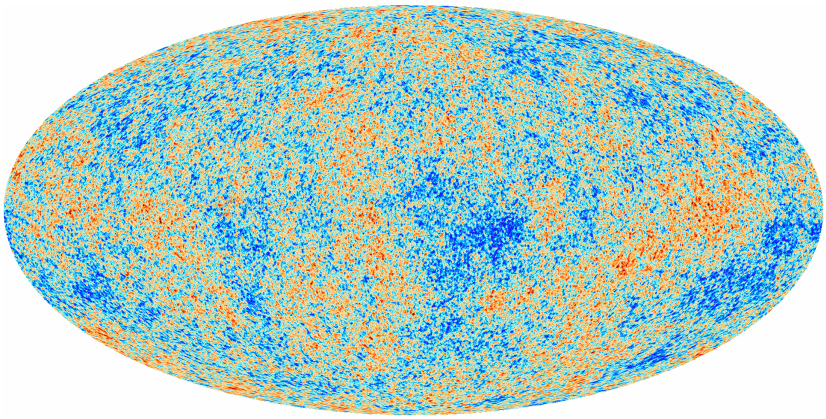


Figure 1.3: All-sky image of the CMB as seen by the Planck satellite — the red and blue regions indicate minutely ($\sim 10^{-5}$K) hotter or cooler regions; image from http://www.esa.int/Our_Activities/Space_Science/Highlights/Planck_s_Universe.

The Big Bang is today an integral part of the 'Standard Model' of Cosmology – the $\mathbf{\Lambda}$CDM Model. Unfortunately the theories of physics break down at scales where General Relativity and Quantum Field Theory both become important (the Planck scale), as such, the nature of the progenitor of the Big Bang, and whether or not it marks the beginning of the Universe, or just of the Universe's expansion, remain open questions.

Dark Matter

Dark matter (DM), and specifically cold dark matter (CDM), is another of the integral components to the $\mathbf{\Lambda}$CDM model. It has been known to astronomers that dark matter (in the visible spectrum) exists in space; cold gas, dust and low albedo planets and planetary bodies for example. All these objects are made of baryons (protons, neutrons, and other quark-composed particles) and emit in other wavelengths of light that are not part of the visible spectrum (e.g.: in infra-red) in addition to blocking light from background sources behind them. Truly dark matter has also been known to exist since the discovery of neutrinos (Cowan et al. 1956) which don't interact with the electromagnetic force, and hence are inherently dark. Neutrinos are considered 'hot' dark matter since they can traverse very large distances before interacting and hence do not interact particularly strongly through gravity (their travel speed is close to the speed of light and they are of very low mass, and so they cannot 'clump'). CDM is a hypothetical form of non-baryonic matter that interacts only weakly with normal matter, and is non-luminous and non-reflective since it does not interact with the electromagnetic force, but 'slow' enough (compared to the speed of light) so as to allow for time to bind gravitationally and clump

[6]The CMB phenomenon had been observed prior to the observations of Penzias and Wilson, however, the previous observations had not been conclusively identified with the phenomenon as predicted by Alpher and Herman.

(hence 'cold'). Hot DM models fell out of favour when predicted large scale structures could not be made to resemble observations from the initial conditions set by the very small CMB temperature deviations, it remains however one of the minor components of dark matter.

The clues as to the existence of dark matter (in particular cold DM) came from observations of galaxy clusters and galactic rotation curves as performed by Zwicky (1937) and Rubin and Ford (1970); Rubin et al. (1980), in which it was observed that objects (particularly in the outer regions where they were expected to be slower) were moving at speeds much faster than could be accounted for if the gravitational attraction present were due to that of the observable luminous matter only; even accounting for potentially obscured matter was not sufficient to explain the discrepancy. The most significant evidence for the existence of DM comes from observations of the merger of galaxy clusters, in particular that of the Bullet Cluster (Clowe et al. 2004) where the (free) baryonic components (mainly compressible dust and gas) have collided and heated up (the pink regions in figure 1.4) and the compact stellar components have moved past one another (the mean distances involved are very large, so they do not collide). The significance is that in the regions highlighted in blue, the background galaxies have been lensed (this is a property of massive objects in GR, which can bend light around them much like a lens), and the luminous foreground galaxies do not have sufficient mass in order to produce the extent of the observed lensing. Additionally, the centre of mass of the visible components, being in the approximate centre of the collision, should in principle be where the most lensing occurs, but this is not what is observed. Hence there must be missing mass, that is incompressible (since if it were not it would collide like the gaseous components), non-luminous and gravitationally interacting, in addition it must be sufficiently slow so as to maintain a presence on the timescales of galaxy cluster mergers – Cold Dark Matter.

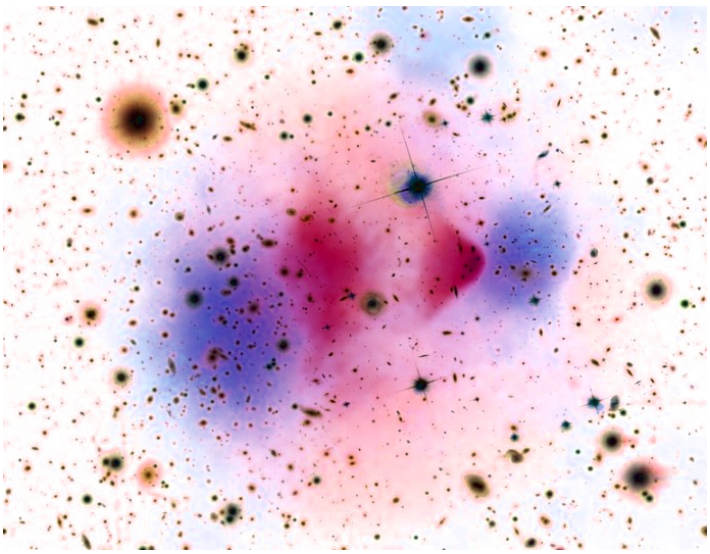


Figure 1.4: False colour image of the Bullet Cluster — A merger of two galaxy clusters, the colliding and heated gas is shown in pink, but the majority of the mass associated with the lensing is highlighted in blue; adapted from http://images.astronet.ru/pubd/2008/08/24/0001229277/bullet_cluster_c60w.jpg.

The latest results from the Planck Collaboration et al. (2013a,b) suggest that dark matter is just under 5.5 times more abundant (in mass-energy terms) than ordinary (baryonic) matter, and it is reasonable to assume that, given the nature of gravitational attraction, dense regions of matter (galaxies and galaxy clusters) are accompanied by dense regions of CDM.

Currently it is not known what CDM actually *is*, though there are candidates with the current favourite being an as yet undiscovered variety of 'WIMP' (Weakly Interacting Massive Particle)[7]. WIMPs are hypothesised particles from the particle physics sector, and are expected to have the correct properties for being a cold dark matter candidate, namely they are expected to be neutral, heavy, only weakly interacting with baryonic matter, and stable. Favourite candidates are neutral supersymmetric partner particles that arise from natural extensions of the Standard Model of particle physics under a symmetry of fermion-boson exchange (these may interact both via the weak force and gravity); and/or sterile neutrinos that arise naturally from (non-supersymmetric) extensions to the Standard Model from chirality/handedness considerations and provide a mechanism for regular neutrinos to acquire mass (sterile neutrinos interact only gravitationally with regular matter). The most recent (lack of) results from the LHC appear to rule out many of the simplest supersymmetric extensions to the Standard Model of particle physics[8], though these results do not preclude the existence of more complex supersymmetric extensions, or other varieties of extensions that do not include supersymmetry; hence from the standpoint of particle physics, the exact nature of CDM remains an open question.

The Cosmological Constant - $\mathbf{\Lambda}$

With the observations of an expanding Universe, the prevailing thought at the time was that the expansion would be slowing down over time as the mass-energy content of the Universe would slow the expansion. The Universe was assumed to be FLRW with some dark matter component, and a potential very early inflationary period. The primary debate at the time was whether or not this mass-energy content was sufficient to halt the expansion, and subsequently reverse it, or whether it would never quite be enough, and the Universe would simply keep expanding ever more slowly.

As mentioned in section 1.1.3, the cosmological constant is an additional term, introduced by Einstein into his gravitational field equation (1.9), with the intention of producing a static model for the Universe. The static model was soon discarded in favour of a dynamic, expanding model in light of observations, however the $\mathbf{\Lambda}$ term remained plausible, and there was no physical motivation to require it to be zero. $\mathbf{\Lambda}$ is a property of spacetime itself, it behaves in a *repulsive* manner and crucially, if its value is large enough, results in the accelerated expansion of the Universe.

Observations by Riess et al. (1998); Perlmutter et al. (1999) of type Ia supernovae, a type of standardisable candle whose intrinsic brightness (and thus distance) can be known, found that supernovae that were further away had their light less redshifted than expected, indicating that, at the time of emission, the Universe was expanding more slowly than it does today and hence the expansion of the Universe had increased over that time frame – the expansion of Universe is indeed accelerating.

$\mathbf{\Lambda}$ is associated to the energy of the vacuum of space, or simply 'vacuum energy'. We know that $\mathbf{\Lambda}$ has certain properties: it has negative 'pressure' (if positive in value) and so behaves like a repulsive counterpart to gravity, its energy density remains constant (in space) under expansion, and it comprises a very large fraction of the total mass-energy content of the Universe ($\sim 68.3\%$, Planck Collaboration et al. 2013b).

Specifically, the cosmological constant, $\mathbf{\Lambda}$ is a special case of vacuum energy in which the energy density remains constant over time, a time-dependent cosmological constant is not admissible by Ein-

[7] Other, non-matter, alternatives such as modified theories of gravity exist, however, these often have outcomes which contradict established observations, or properties considered unsatisfactory (loss of the equivalence principle, for example).

[8] In particular the observation of B_s meson's CP violating decay into a muon-antimuon pair, at a rate exactly predicted by the Standard Model, without enhanced loop-contributions from supersymmetric particles, puts strong constraints on the nature of supersymmetry (Aaij et al. 2013).

stein's field equation (1.9). Other, more exotic, possibilities could produce effects like a cosmological constant, but allow the possibility to vary over time.

The source of such a vacuum energy is not known, this lack of knowledge is embodied in the name given to it – Dark Energy (DE), though this in part stems from the energy not being electromagnetic in nature (and hence dark). Virtual quantum particles, appearing for the briefest of moments from the vacuum, borrowing the energy to do so from the vacuum itself, before mutual annihilation and restoring that energy back to the vacuum (as a consequence of Heisenberg's Uncertainty Principle) are one possible contributor, and could produce such an effect. Theoretical predictions from QFT however predict a value for the vacuum energy that completely dwarfs what is observed by 120 orders of magnitude, this discrepancy is so large it is termed the *vacuum catastrophe*, (a detailed review from both perspectives can be found in Carroll 2001). Evidence for vacuum energy of some kind has been observed in measurements of the Casimir Effect (an effect that relies on virtual particles in the vacuum being excluded between parallel metal plates resulting in attraction between them, as measured by Sparnaay 1957; Lamoreaux 1997), though it is not yet known if there is a mechanism that may make it possible to reconcile the discrepancy between QFT prediction and cosmological observation.

ΛCDM Universe

The **Λ**CDM Universe is therefore an expanding Universe, with the rate of that expansion increasing over time. This expansion had a beginning with the Big Bang, and a period of rapid inflation producing a (very nearly) spatially flat metric as in section 1.1.4. Quantum fluctuations at the subatomic scale produced minutely over-warm and under-warm regions on large-scales. A radiation-dominated phase where relativistic matter and radiation were coupled together as a hot and superdense plasma spawned, upon expansion and cooling, a transparent and matter-dominated era where initially under-warm regions experienced local over-densities and began to clump cold dark matter (CDM) and eventually hydrogen and helium gas to form the first large compact structures under the influence of gravitational attraction. The expansion of the Universe slowed during this matter dominated phase, however the continual expansion progressively reduced the energy density of both the matter and radiation content of the Universe, resulting in balance tipping toward the current dominance of Dark Energy. The model is isotropic and - on large enough scales such that local clumping of matter is negligible - homogenous.

The **Λ**CDM model for the Universe is in good agreement with current observations, and simulations of such a model produce Universes that resemble our own (Springel et al. 2005b, the Millennium Simulation).

1.2 Measuring Cosmological Redshift

The accurate measurement of the redshifts of a large number of galaxies is an important goal of Cosmology. Redshift is the best proxy we have for the large-scale distances to galaxies, and it is a direct indicator of the scale factor of the Universe at earlier times. It is only through redshift that we can begin to untangle the otherwise 2D night-sky into a 3D and temporally-dynamic map of the Universe. The location, distribution, history and evolution of large-scale structures are critically important to our understanding of the formation and evolution of our Universe. Unsurprisingly, good redshift measurements are needed in many other fields of Cosmology, in order to help cosmologists answer the still many unanswered questions about the **Λ**CDM model. Redshift is useful for but not limited to: studies of large-scale structure; galaxy formation, evolution and distribution; weak lensing; studies attempting to determine the nature of dark energy; the distribution, location and evolution of

CDM in the Universe; and constraining Cosmological parameters.

1.2.1 Galaxy Types

To measure the cosmological redshift of different galaxies, we need to collect their light and identify distinguishing features. For the purposes of the rest of this chapter (and indeed all of this work), the term redshift is assumed to mean cosmological redshift.

A galaxy is a dynamic and varied collection of many gravitationally bound stars. Each star in the galaxy will have a characteristic emission spectrum and luminosity, due to having unique chemical compositions and masses. The total light observed coming from a galaxy (treating it as a point source) is the sum of all the light being emitted from its constituent stars. Hence, with each galaxy having a unique population of stars, the sum of all the light it emits will also be unique; having different fluxes/intensities at particular wavelengths. This sum total of emitted light is called a Spectral Energy Distribution (henceforth SED) if it is a plot of energy versus wavelength as in figure 1.5, or simply a spectrum if it is a plot of flux versus wavelength.

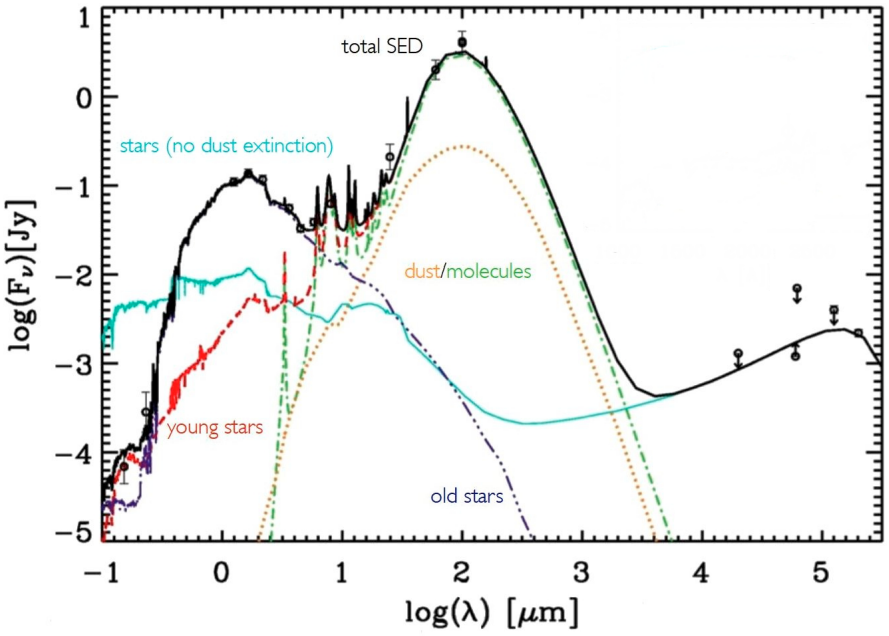


Figure 1.5: An Example of an SED Profile — The diagram represents an SED profile of a galaxy, depicting in particular the contributions that the different sources (coloured) have on the total output at each particular wavelength (solid black line). The y-axis represents units of energy and the x-axis is the wavelength. Adapted from Panuzzo et al. (2007)

There are 3 broad types of galaxies, first classified by Hubble (1926), spiral (also called late-type), elliptical (also called early-type), and irregular. The terminology of late and early type is potentially misleading since there is no evidence to suggest that elliptical galaxies evolve into spirals (on the contrary, there is evidence to suggest the reverse is true). The galaxy type will have a bearing on the characteristics displayed by its spectrum, so much so that galaxies can be grouped by spectral type.

Elliptical galaxies are normally found in populous regions such as clusters, and in particular towards the centre of clusters (Loveday 1996). How they form has been a matter of some debate: they may have begun forming in much the same way as spirals but have been prevented from forming a

disk due to their local perturbative environment (Dressler 1980), or they may have formed directly as a consequence of the 'warmth' of their dark matter environment (Lake and Carlberg 1988a,b). The leading hypothesis however is the alternative mechanism that they are constructed from mergers of similarly sized spirals where the tidal disruption is large enough to effectively disorder the disk structure resulting in a structure-deficient spheroid (Barnes 1992; Kauffmann et al. 1993; Springel et al. 2005a, and for a review see, Barnes and Hernquist (1992)). Studies of distant galaxy clusters by Dressler et al. (1994) suggest a comparatively reduced population elliptical galaxies, and a comparatively enhanced population of interacting galaxies in relation to lower redshift (and hence more evolved) clusters, with this further indicating that tidally dynamic environments, and in particular mergers, are likely responsible for generating elliptical galaxies. This hierarchical merging of smaller progenitors into larger galaxies originates from the ongoing merging of (cold) dark matter haloes (Press and Schechter 1974; White and Frenk 1991). It should be noted that there is no fundamental reason that 'primary' ellipticals and ellipticals arising from mergers should be mutually exclusive, and it remains possible that both these mechanisms are at work.

Ellipticals are characterised by very low star-formation rates (SFRs) since they have exhausted most of their supply of stellar raw materials – gas and dust, with these having been catalysed into star formation or otherwise ejected from the system from the merger of the progenitors or strong tidal interaction. The stars within them tend to be on average older, cooler, redder, and less massive but with higher metallicities, this is because the high mass and short lived blue stars have long since expired (their lifetime being much shorter than the age of their host galaxy). Together this results in ellipticals, in their own rest frames, being generally redder overall and frequently showing a strong jump in the spectrum at ∼4,000Å which is mainly due to the presence of metals[9], with a contribution from the Balmer break.

Spiral galaxies are usually found in more isolated regions, (said to be in the *field*) or toward the periphery of clusters (Dressler 1980). Spirals are more structurally complex, possessing regions of denser material in a disk possessing coplanar arms that spiral away from the centre. The arms are the primary location for star formation, and are thus typified by a young, hot and blue stellar population. The origin of the spiral arms, and whether they persist, with stars and dust/gas components moving along with them, on lifetimes comparable to the age of the galaxy (the winding problem), or instead whether the arms move relative to the stars themselves, compressing and expanding regions of dust/gas into star formation as they travel around the centre is not known (Toomre 1977; D'Onghia et al. 2013). Further components of spiral galaxies can include a central bulge (itself reminiscent of elliptical galaxies on a smaller scale), bar features, and a population of diffuse older, redder, and metal-deficient stars in a halo not otherwise associated to the arms/disk. The galaxy itself will likely contain significant amounts of dust and gas, potentially causing absorption of part of the spectrum. Spirals are characterised by high SFRs (Kennicutt 1998). They may also host Active Galactic Nuclei (henceforth AGN) which are hypothesised to be strongly accreting black holes at the galactic centre (Lynden-Bell 1969), causing an increase in luminosity at certain, and in some cases all, wavelengths. Also, under the unified model of AGNs, different galaxy types (eg: 'normal', Seyfert I & II, Quasar, etc) based on their properties (such as being radio-loud, emission-line widths, excesses in different wavelength ranges, etc) are essentially the product of the orientation of the AGN at the galactic centre, with the jets that the AGNs produce being of any inclination, and in particular if along the line of sight, the overall galactic luminosity can be greatly increased (Antonucci 1993; Urry and Padovani 1995; Bianchi et al. 2012; Kazanas et al. 2012).

[9]Astrophysicists term elements heavier than Helium as 'metals', though strictly speaking, some are non-metal: C, N, O etc.

Irregular galaxies are defined mostly by not being of either elliptical or spiral type. They may be pairs of galaxies (or more) in different stages of the process of merging, or very recently merged, or if very early on in the Universe, a disorganised collection of stars yet to take shape. Depending on the progenitors and the progression of the merger, irregulars can exhibit some of the characteristics of both spirals and ellipticals or no strongly defining characteristics at all; they are generally less well defined than either of the previous classes. Irregular galaxies may also be produced by strong disruptive tidal interactions with neighbours disrupting the shape, but not causing any merging, this can trigger new star formation and add or remove components from the galaxy (e.g.: gas, stars). This complexity can on occasion make it more difficult to produced generalised spectra for irregular type galaxies when compared to the other two types.

Broad classifications of colour and magnitude for galaxies are not as precise at pinpointing type, evolution and other properties as for stars, but such classifications do reveal a bivariate distribution corresponding to two main groups, a 'blue cloud', corresponding mainly to spirals; and a 'red sequence', corresponding mainly to ellipticals (Baldry et al. 2004; Bell et al. 2004). Luminosities tend to increase as a function of metallicities and hence age, populations within the 'red sequence' being predominantly ellipticals, thus show a natural progression towards the more luminous members being most red (since the interstellar medium becomes more enriched over time, and star formation remains relatively flat). The 'blue cloud' shows no such definitive sequence and is most likely complicated by gas and dust (reddening factors), inclination (affecting both luminosity and, to a lesser extent, colour) and alternating quiescent and starburst phases (resulting in a comparative lack of blue emission during quiescent phases an excess of blue emission during active phases).

There exists a third, much less numerous population between these two termed the 'green valley', which predominantly consist of galaxies transitioning from the blue cloud to the red sequence. It is much less populated since the transition from blue to red happens comparatively quickly. After a sudden starburst phase in catastrophic spiral mergers into a single spheroidal galaxy, the galaxy quickly reddens as metallicity is boosted from the remnants of high mass stars, and the gas/dust content is depleted; the brief lifetime of the high mass stars ensuring that the blue excess from rapid star formation is transient in the absence of further raw materials. Other members of the green valley population might include high redshift blue galaxies, quiescent blue galaxies with a large proportion of dust, quiescent evolved spirals that have exhausted their gas/dust supply, and red galaxies that have had a sudden burst of star formation (possibly by cannibalising a smaller, gas-rich satellite galaxy for example).

As well as the stellar component, galactic light can also be contaminated (which may be augmentative or subtractive) with other sources, including but not limited to: noise; intrinsic sources such as gas and dust/interstellar medium (ISM), supernovae, AGN; and extrinsic sources such as other luminous objects along the line of sight, IGM, terrestrial atmosphere (for ground based observations). For nearby galaxies, the light of individual stars can be isolated (and measured); however if the galaxy is very distant, it will appear point-like, and the individual stars within it cannot be resolved.

1.2.2 Spectroscopy

The principal method for estimating redshifts of galaxies to date, has been that of spectroscopy. Spectroscopy relies on the fact that even though separated by millions of parsecs, the constituents of galaxies remain predominantly the same, namely being Hydrogen and Helium dominated, with the addition of further heavier elements – metals.

When undergoing excitation, chemical elements will absorb (or emit if undergoing de-excitation)

photons of very specific wavelengths dependent upon electron-shell transitions, resulting in a chemical fingerprint that is unique to that element.

Instrumental investigation involves light taken from the source galaxy, and spread out very widely (diffracted) with the aid of diffraction gratings, so that a spectrum is produced of the source's light separating the wavelengths over a wide area, the general technique having been invented by Fraunhofer in 1814, identifying absorption and emission lines for the first time. Secchi (1877) pioneered the use of spectroscopy in astronomy, by classifying stellar spectra and identifying the Sun as a star. The same technique was later adapted to study galaxies, and used to great effect by Slipher (1915) in identifying local galaxy (recessional) velocities. Emission and absorption features in the spectra can be found, for example resembling the characteristic lines of Hydrogen lines here on Earth, but the lines themselves are found to be of longer wavelength – they have been redshifted. The difference in where the lines are measured to be and where they are expected to be is precisely related to the redshift by equation (1.18).

As may be evident, the more wide the spectrum can be made to be, the more precisely these characteristic wavelengths and wavelength gaps can be measured, particularly in regions where many emission/absorption lines can be very close to one another. This means spectroscopy can be very precise in obtaining a value for the redshift of a galaxy.

There are however problems with spectroscopy. To obtain a sufficient flux of photons, the target galaxy must be observed for a considerable length of time thus making spectroscopy very slow. The reason why a large flux is needed is due to the fact that in the spectrum forming process, the flux is rarified to observe the lines. This in turn limits spectroscopic estimation of redshift to closer or particularly luminous galaxies.

One further complicating factor is what is known as the 'redshift desert', which limits the usefulness of spectroscopy in the redshift range of $1.4 \lesssim z \lesssim 2.5$ (Steidel et al. 2004), particularly for ground-based observations. The reason for this is an unfortunate overlap of technological and physical limitations making the redshifts of galaxies significantly more difficult to identify, rather than representing a true absence of observable galaxies.

From a technological standpoint, ground based observations currently have their most significant sensitivity in the 4,000 to 9,000 Å range when compared to the regions just outside this range. The principal reasons being that within this range, CCD technology is comparatively very efficient, whereas beyond this range, cooling and different materials for CCD design would be needed. The night-sky background that would otherwise be a contaminant to feature-identification is comparatively very low in this region, and the atmosphere is highly transparent to visible and near-infra-red light. However, in excess of 9,000 Å, both background sources and atmospheric transmissivity become more problematic: the night-sky background is brighter in this region and thermal noise in the IR necessitates cryogenic instrumentation with cooled optics/focal planes (Steidel et al. 2004), additionally the atmosphere is punctuated by windows of opacity and regions of reduced transmissivity in the mid-IR mainly due to the (often variable) presence of water vapour, and is completely opaque for large sections of the far-IR at ground level. The atmosphere is also opaque to near-UV and narrower wavelengths, limiting the shorter wavelength detection range of ground-based spectrographs.

Independent of these technological issues however, there remains a further, physical, problem when moving up in redshift to $z = 1.4$ and beyond. Approaching this redshift value results in each of the principal identifying features – characteristically strong emission lines such as the [O II] doublet, [O III], H_β, and H_α – at low-z, moving far enough to the long wavelength end of the spectrum so as to lie beyond detectability within an 'observing window' of 4,000 to 9,000 Å; the [O II] line marks

the final detectable line at a redshift of $z \sim 1.4$. At $z = 2.5$ and beyond however, new features from the far UV (in the emission rest-frame) begin to move into this observing window, allowing redshift identification for galaxies beyond this redshift range (if they are sufficiently bright). Between these redshift values however, no significant features typically exist within the observing window, thus rendering this particular redshift range a 'desert' with respect to redshift identification, and thus poorly mapped in relation to the regions just below and above this range. Good mapping of this range is critical to our understanding of the global star formation rate, which is currently believed to have peaked sharply at a redshift of $z \approx 1.5$, however decreasing only slowly beyond this redshift to about half its peak-value at a redshift of 3 (Connolly et al. 1997). Hence the redshift desert makes this peak (and associated high region) in global SFR – indicating that this particular time period was significant in terms of galaxy formation, and potentially very active with respect to galaxy mergers – in part more difficult to observe.

Even though spectroscopy is a very accurate method, its negative points make spectroscopy unsatisfactory for probing the deep Universe; it is too slow, it will only pick out the very luminous galaxies which are unlikely to be representative of the galactic population at large, and there is a markedly reduced performance in the 'redshift desert' ($1.4 \lesssim z \lesssim 2.5$). There is however another method for estimating redshift that provides an alternative solution.

1.2.3 Photometry

The concept of using photometry to measure redshift is not new, and dates back to Baum (1962), who proposed that by using broadband colour filters, it would be possible to identify features that, depending on which band they fell into, would reveal their redshift. These features are usually taken to be a large decrease/increase when moving from one range of wavelengths (i.e.: the colour band) to another, or alternatively, a *break* in the SED profile. Often this will be the Balmer break or the 4,000Å break (which is mainly due to metals, and often overlaps with the Balmer break), or less commonly, the Lyman break.

In rare cases, a sufficiently active (and fortuitously orientated) AGN can supplement an SED with emission lines, provided this contribution is at least comparable in magnitude to the photometric errors involved, as shown by Hatziminaoglou et al. (2000). The presence of these emission lines can also assist in targeting photometric redshift calculations at particular SED features.

Photometry has, in recent years, become a more prominent tool for measuring redshift. Compared to spectroscopy it is considerably faster, (by orders of magnitude) since it does not involve many hours collecting light from the target galaxy due to the fact that it does not require the forming of spectra. Photometry also does not require the same high luminosities that spectroscopy does, and thus can provide redshift measurements for many more, as well as a more representative sample of, distant galaxies.

There is, however, one major drawback to Photometry as a redshift estimation tool, if a break is detected in a broadband colour filter, or even between filters should they not be contiguous, it is impossible to identify precisely where in that band (i.e.: at what wavelength) the break occurs. Also the observed SED may not have any prominent breaks at all, or more than one, making redshift estimation much more complicated. Evidently photometric redshift (also known as photo-z or z_{phot}) estimation methods are not as precise as those that use spectroscopic redshift (also known as spec-z or z_{spec}). Spec-z and photo-z methods are not mutually exclusive, indeed, they are often used in conjunction, with the spec-z estimation operating as a backup, checking or training method for photo-z estimation.

The optimal way to isolate breaks would be to have many narrowband filters, since this would make it easier to track breaks and their locations. Unfortunately, making the filter bands narrower has the adverse effect of requiring more flux to obtain the same strength of signal. There exists a trade-off: too narrow, and the filter bands will struggle to detect signals against noise or take longer to gather enough flux to do so; too wide and the bands will lose sensitivity to both the location and indeed presence of crucial identifying features. It is not known what width of filter is optimal for this task, it is likely to be dependent upon SED type (those which are smoother and with less prominent breaks will not need filters to be particularly narrow), and the magnitudes of the galaxies observed (with dimmer galaxies being more quickly imaged with wider filters).

Most of the photometric redshift codes available operate by taking magnitudes in different colour banded filters obtained from Large Sky Surveys (such as the Sloane Digital Sky Survey, SDSS, Fukugita et al. 1996) and computing the redshift, z_{phot}, in one of two possible ways.

One method involves using the survey data to then construct an observed SED (redshifted) and matches it to a known (local or model) SED type or template, looking specifically for any distinct breaks. The other alternative is that the code can take a subset of those galaxies for which pre-recorded spectroscopic redshifts are available, and then 'learn' the relationships between magnitudes in each filter (i.e.: the photometry) and the redshift (as determined by the spec-z), and after having learnt that 'formula', applies it to the rest of the galaxy data available (ones without any spec-z data) in order to determine the redshift. These two methods are called the *'template-matching'* (or sometimes template fitting) method and the *'training'* or *'empirical'* method respectively. Both approaches were, in part, developed by Connolly et al. (1999) for template matching and Connolly et al. (1995) for empirical methods.

However, independent of the method chosen, there remains the same raw material to work with, namely magnitude measurements in various broad colour bands from surveys; they thus will incur the same errors associated with this process. A large number of astronomical observations are made using the filters that were used for the SDSS survey which were extensively defined by Fukugita et al. (1996).

In general there are additional effects on the measurement of the SED of a galaxy due to the 'hardware' used to perform the measurement. These are the transmission of the filters used for each colour band, the quantum efficiency of the CCD device used to record the photons, and the Point Spread Function or PSF of the telescope.

In principle a filter *should* allow all photons through so long as they are within the wavelength range of that filter, with photons that lie outside this range being absorbed (such filters are called *'top-hat filters'*). In practice however, filters will absorb some of the photons they are designed to transmit, and hence real transmission values do not reach 100%, with some being as low as 75%. If one filter has a considerably larger transmission that the one immediately preceding or following it, this may affect the breaks of an SED by making them appear larger or smaller then their actual value. This is simple to account for if the properties of each filter are known beforehand. An example of the transmission curves for filters are shown in figure 1.6.

The quantum efficiency arises from the nature of quantum mechanics as a random, probabilistic process; the photons striking the CCD plate will have an associated probability to produce interactions (electron-hole pairs), however, there also exists a corresponding probability to not produce interactions, therefore some photons will simply be absorbed and not register in the CCD. The efficiency is strongly dependent upon the photon wavelength, and of course, the type of CCD used. Again, if the CCD is much more efficient to detections of one wavelength range than to another, it will distort the SED in

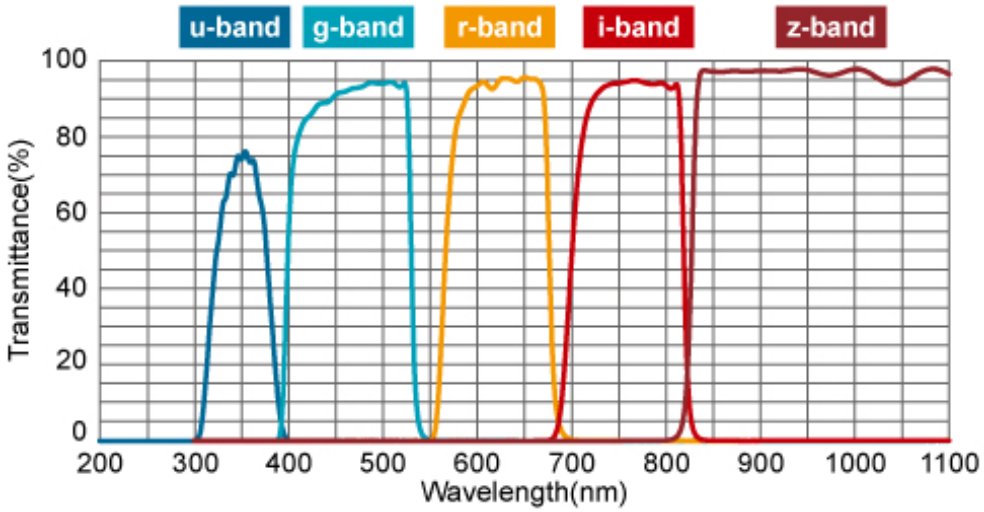


Figure 1.6: Filter Transmissions — These graphs show the filter bands of the ugriz filters used on the Sloan Digital Sky Survey, along with their approximate colours and transmissions. Image from: http://www.asahi-SeminaireBourbaki.spectra.com/opticalfilters/astronomical_filter.html

that range.

Both these effects will combine to give a response curve for each filter. Generally these responses will be different from one another, often significantly so. The responses can, in principle, be corrected for with suitable normalisations, if we know a priori the transmission of the filters and the CCD efficiency. Figure 1.7 gives an example of total response curves for the SDSS filter system.

Lastly, the PSF is a mathematical characterisation of what the imaging instrument (i.e.: telescope) does to a point source of light (which, depending on the observational nature of the instrument, may or may not be inclusive of atmospheric effects). In general, telescopes do not give us an exact representation of the object it is trained on; instead, point-like objects will appear blurred, and this would incur errors in photo-z estimation. As an example consider two galaxies, at very different redshifts that happen to be aligned along the line of sight, if those galaxies are sufficiently close together, the PSF of the telescope would smear the two objects into one, and our estimate for the redshift of this 'galaxy' would be completely incorrect. In principle PSFs can be corrected for by deconvolving out their contribution to a particular image. This is a non-trivial process, since the PSF is often not a simple function, varies across the image (generally asymmetric), and in some cases varying over time as well (particularly for ground-based observations with atmospheric contributions). This deconvolution is usually performed by training the imaging system on known point sources, and attempting retrieval of the original image, in order to deduce a good model for the action of the PSF on the image.

1.3 SEDs & Spectra

As mentioned previously in section 1.2.1, the constituents of a galaxy have a direct impact on what its SED or spectrum will look like. Although SEDs and spectra are subtly different - one is defined in units of energy (strictly energy density) per wavelength bin, whereas the other if in units of flux (strictly flux density) per wavelength bin - the principles and physics behind them remain the same so much so that the two terms are often used (though it is technically erroneous) interchangeably. It

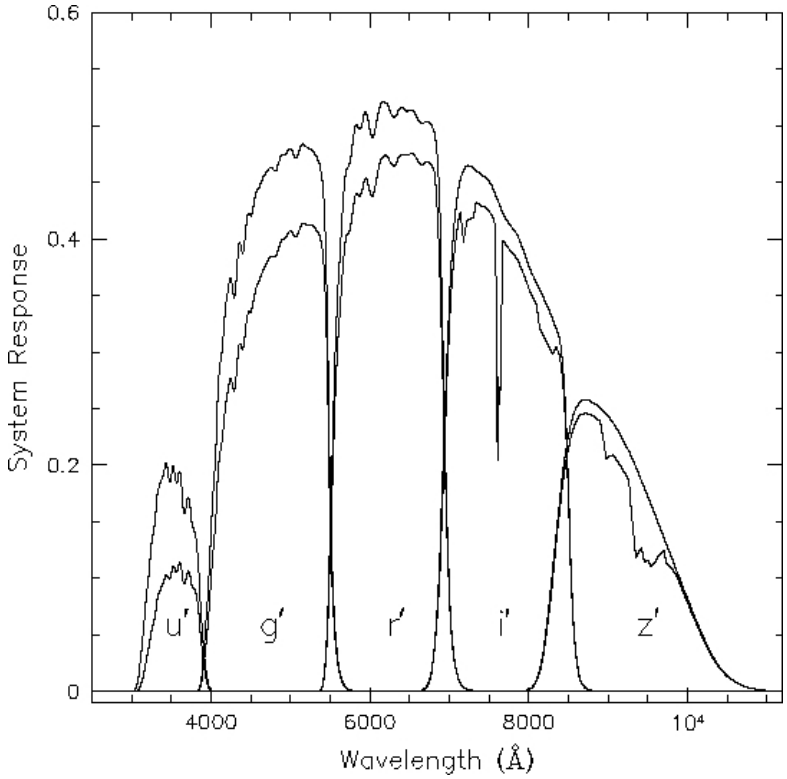


Figure 1.7: Response Curves — These graphs show the total response curves for each of the SDSS ugriz filters after both transmission and CCD quantum efficiency have been taken into account. The lower curve is the same as the upper curve except for the inclusion of atmospheric absorption. Image from: http://www.cfht.hawaii.edu/Science/mswg/filters.html

was also noted that the different constituent parts of a galaxy contribute to different features in the spectrum.

1.3.1 Stars

All galaxies contain a stellar population, and depending on the initial conditions during the formation of the galaxy, and its subsequent evolution, the types and abundances of these stars will differ significantly. Stars of course, come in different types, characterised primarily by their luminosity and chemical composition, from which temperature, mass, size and age can then be determined.

Secchi (1877) was the first person to begin to classify stars according to features in their spectra (predominantly the width of hydrogen lines), however at the time, concepts of stellar evolution based on nuclear fusion of light chemical elements could not be proposed without the advent of mass-energy equivalence (Einstein 1905), its association to energy generation in stars (Bethe and Marshak 1939) and theories of stellar nucleosynthesis (Hoyle 1946). As such, the initial classes developed by Secchi did not always correspond well to physical properties or individual stellar type.

A later classification system - Harvard spectral classification (based on the Draper catalogue, Pickering 1890) - classifies stars primarily on their temperature as inferred from their observed colours. In order of decreasing temperature these classes are O, B, A, F, G, K, and M, (with further subdivisions)

ranging in colour from blue, white, yellow, orange and red[10]

A further classification is the MKK system[11], and later refined to just the MK system (Morgan et al. 1943; Morgan and Keenan 1973), in which stars are classified primarily on their luminosities with some sensitivity to both temperature and surface gravity (and thus radius and mass). Stars are classified from **0** to **VII** (with further subdivisions), with **0** being the most luminous (hypergiant) class. The class denoted **V** represents main sequence stars like our Sun, and **VII** represent white dwarf stars. The various Harvard class spectra, in conjunction with their approximate colours are shown in figure 1.8 for stars all of the same MK type, in this case type V, main sequence stars.

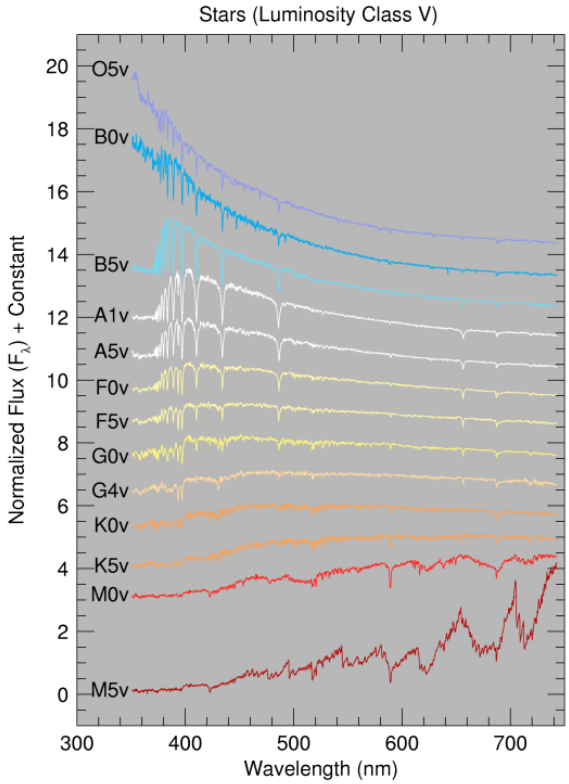


Figure 1.8: OBAFGKM Spectra — Different Harvard class spectra along with approximate colours for stars of the same MK type (type V, main sequence stars). Image adapted from: http://www.astronomy.ohio-SeminaireBourbaki.state.edu/~pogge/TeachRes/Ast162/SpTypes/OBAFGKM_scans.jpg

As can be seen the general trend is from red to blue with increasing stellar temperature (from M to O in the Harvard scheme). Class M can be highly variable (due to the cooler temperature allowing for molecular chemical species to become important) as demonstrated by the two, evidently dissimilar, M type spectra, however a general trend of excess in the red is true for all M-type spectra. The skew toward excess in either the blue or the red wavelength regions is dominated by temperature. A galaxy composed primarily of blue stars (O and B), will thus have an overall bluer colour (in its own rest frame); and if otherwise composed of primarily of red M type stars, it will instead have an overall

[10]The reason for the apparent disordered labelling is historic, with stars originally being labelled on their hydrogen line widths with A being the most intense. Unfortunately the hydrogen line widths peak at a certain stellar temperature (that of A-type stars) and subsequently decrease as temperature increases or decreases from this value, resulting in the rearranged and somewhat obscure ordering. Classes L, T and Y also exist, however, they refer to sub-stellar mass objects that do not undergo nuclear fusion in their cores, and hence are not particularly luminous and do not impact sufficiently on a galactic scale to alter its spectrum.

[11]Also known as the Yerkes system.

redder colour.

The general population of stellar types and the evolution of stellar properties are often conveniently summarised on a Hertzsprung-Russell (HR) diagram (figure 1.9). The different MK sequences are labelled, demonstrating their relative luminosities, note however that each MK class straddles a large temperature range, and thus a large number of Harvard classes (as in figure 1.8). Furthermore, a single Harvard class, may have stars of different masses, radii, and luminosities and as such don't represent a homogeneous group of physically similar bodies, even if the spectra share similar characteristics. Properties such as stellar mass and stellar radius can be seen to track with increasing luminosity, however for relatively fixed luminosities (giant branches), red stars have significantly larger radii than blue stars.

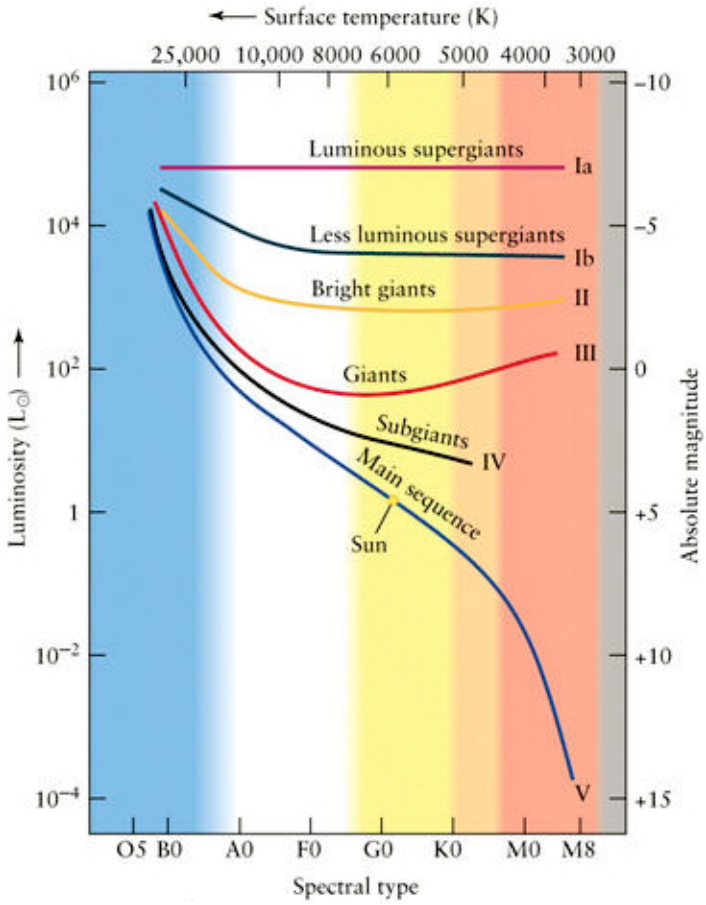


Figure 1.9: The HR Diagram — Stars are arranged according to their spectral type as a function of their luminosity and surface temperature. Both stellar mass and radius tend to increase with increasing luminosity. Image from: http://www.daviddarling.info/encyclopedia/L/luminosity_class.html

There exists a relationship between mass and luminosity (the mass-luminosity relationship Eddington 1924) that allows for an estimate of the mass of a star from its luminosity as follows,

$$\frac{L}{L_\odot} \approx \beta \left(\frac{M}{M_\odot}\right)^\alpha \tag{1.20}$$

where the subscript $_\odot$ denotes the mass/luminosity with the respect to the solar mass/luminosity. The constants alpha and beta, as well as the approximate nature of the equality, change dependant

upon the mass regime of the stars in question,

$$\begin{cases} M < 0.43M_{\odot}, & \beta = 0.23,\ \alpha = 2.3 \\ 0.43M_{\odot} < M < 2M_{\odot}, & \beta = 1,\ \alpha = 4 \\ 2M_{\odot} < M < 20M_{\odot}, & \beta = 1.5,\ \alpha = 3.5 \\ M > 20M_{\odot}, & \beta = \alpha = 1 \text{ (only a proportional relationship).} \end{cases}$$

The origin of the different formula for different mass ranges is due to the primary mode for energy transport from the centre of the star to its surface being either radiative-convective, convective-radiative or entirely convective.

During the initial period of star formation in a galaxy, with an ample supply of raw materials (dust and gas), many stars across the range of possible masses will be produced. Precisely how many of each type are produced depends upon the initial mass available to collapse into stars, dependent upon the initial mass function (IMF). The IMF is an empirical function determining the likelihood of generating stars of a particular mass-range in a given volume of space, relating that likelihood to the inverse of the mass content in that region to some exponent, α, as determined by observation. The first person to perform this empirical derivation was Salpeter (1955) for the case of stars more massive than our Sun, working with globular clusters. The IMF is only valid for hydrogen-burning stars entering the main sequence at the beginning of their life-cycle, and is not intended to apply to brown dwarf formation, however, Kroupa (2001a) tentatively extend the IMF into the sub-stellar region. The topic of IMFs will be explored in further detail in section 3.2.1. Evidently the initial masses and populations of stars will evolve over time into other stellar types (giants) and hence evolve other spectral types accordingly. The resultant galaxy spectrum will be, in its most significant part, an integral combination of the spectra of its constituent stars, dependant upon the age of the galaxy.

A complicating matter is that galaxies do not necessarily have one initial burst of star formation and fall silent, but instead have multiple phases of star formation, or prolonged star formation, as such there is an associated star formation *rate* (as described in section 1.2.1). This rate of star formation is related to infall times for the reservoirs of dust and gas that remain unused (still in the halo) during the initial burst of star formation, and feedback processes from stellar winds, supernovae and galactic winds that can disrupt star formation for example (Heckman et al. 1990).

Knowing the IMF, SFR, the mass-luminosity relation and the overall age of the galaxy can allow us to model the relative proportions of each type of star at any point in its history, and how much each star contributes to the overall luminous output of the galaxy.

1.3.2 Gas & Dust

Gas and dust are the initial raw materials with which stars are constructed during periods of star formation within a galaxy. Depending on the 'birthdate' of the galaxy this may represent an enriched medium with high metallicity (galaxies forming comparatively more recently in the history of the Universe) or a comparatively low metallicity medium consisting primarily of hydrogen and helium gas (galaxies that formed early on in the history of the Universe/protogalaxies) with otherwise little enrichment. Gas and dust are, in the absence of fresh sources (companion galaxies/mergers), limited and eventually used up in star formation, or otherwise ejected from the system through dynamic gravitational interactions. Gas and dust are therefore predominantly of importance for spiral (and irregular) type galaxies, and hence ellipticals – with significantly depleted supplies of both – will be less susceptible to their effects.

The largest factor affecting the luminosity of a spiral galaxy is the inclination. By their nature, edge-on galaxies will be significantly dimmer than face-on galaxies since light from the stars of the trailing edge will be absorbed/attenuated by the stars in the leading edge and any intervening gas and dust within the disk.

Gas can be a source of light within a galaxy, if it is warmed by stars, and can contribute to the makeup of an SED. The specific location of both gas and dust are important, and its distribution can significantly affect the overall SED makeup, in particular the extent of reddening (Calzetti et al. 1996).

Reddening is a composite effect caused by the presence and location of gas and dust within a galaxy that leads to the selective attenuation of blue light. Two components are particularly important to reddening, those of scattering and absorption. The scattering component arises from a process similar to Rayleigh scattering but with a larger particle size (a type of Mie scattering), where particle sizes are similar to the wavelength of blue light. This scattering results in preferential attenuation of part of the blue spectrum, resulting in an overall excess of red. Though the wavelength dependence of this process is significantly weaker than the equivalent for Rayleigh scattering (λ^{-1} as compared to λ^{-4}).

The second important component to reddening is the absorption of higher energy photons by cool dense gas/dust that then re-emits thermally, releasing lower energy photons, thereby skewing/converting light from bluer wavelengths to redder ones (Witt et al. 1992).

The standard way to measure reddening is by the comparison of standard line ratios where these ratios are known (eg: Hydrogen), the differences between the true ratio and the measured one, when combined with the measure colour excess in a particular choice of filter bands, it is possible to estimate the reddening that has occurred (Calzetti 1997).

It is important to distinguish reddening from redshift, whilst both result in a more red than expected spectrum, they originate from significantly different physical effects. A redshift is a translation of the spectrum along the wavelength axis, the locations of spectral lines change; reddening however is a systematic reduction at blue wavelengths or an amplification of red wavelengths and is 'in situ', in the sense that spectral lines are *not* shifted. In general an observed distant galaxy can be observed to be more red due to both reddening and redshift, and in particular cases where there are few/weak emission lines, can be a complicating factor.

1.3.3 Constructing SEDs

Given full knowledge of stellar, gas and dust constituents, inclination, and morphological type – and how much each aspect contributes overall – it becomes possible to construct integrated SEDs/spectra for nearby galaxies where each component can be observed (approximately) separately. However it is generally not possible to observe all these components individually with reference to distant galaxies, since these usually appear point like, and hence the individual effects of each component is by default integrated into a single spectrum.

Given only the integrated spectrum of distant galaxies, it is possible to make an assessment of the likely makeup of the spectrum by observing the proportions of these components in nearby galaxies and comparing when the integrated spectra of these are similar. Hence it is possible to use data from nearby galaxies as templates for more distant galaxies (e.g.: Kennicutt 1992) or in order to generate semi-empirical SEDs for more distant galaxies. Semi-empirical techniques associate regions of the SED to particular properties of the galaxy – for example hot stars (O and OB types) and neutral hydrogen (**HI**) regions are strongly associated to UV features (both emission, absorption, and generalised continuum), cooler stars (F/G/K/M) and **HII** regions (ionised Hydrogen) are associated

to the optical region and IR regions – from which extrapolation to higher redshift can be performed. Such techniques were used by Coleman et al. (1980); Kinney et al. (1996); and Calzetti et al. (1994) to generate a set of SED templates that were accepted as a good set of standard templates (known to the community as CWW-Kinney templates) for cross-correlation redshift techniques such as those by Glazebrook et al. (1998); Connolly et al. (1995).

A further method to construct SEDs for distant galaxies is to do so by entirely synthetic modelling methods. The basic principle generates an integrated spectrum of the galaxy derived from a set of initial conditions (such as the IMF, the quantities of available gas and dust, infall times, star formation rate, metallicities, etc.), that are then evolved to resemble actual galaxy spectra. Many such synthetic models exist, including but not limited to, GALEXEV (Bruzual and Charlot 2003b), GALEV (Kotulla et al. 2009) and PEGASE (versions 1, 2 and HR respectively; Fioc and Rocca-Volmerange 1997, 1999; Le Borgne et al. 2004). The general procedure will be explained further in section 3.2.

1.4 Conclusion

In this chapter, we have introduced the background physical concepts to redshift and spectra. We have traced the history of the discovery of redshift, and derived the origin of cosmological redshift (section 1.1) from the principles of General Relativity as applied to an isotropic and homogenous Universe. We have indicated the key pieces of evidence that have led to the convergence on the standard model of Cosmology, namely the Lambda-CDM model.

In section 1.2, we identified the different morphological types of galaxies (elliptical, spiral and irregular), their principal properties, and their formation histories. The two principal methods by which galactic redshift can be measured were also explained. Additionally, the importance of redshift data from large sky surveys in constraining the Lambda-CDM model was highlighted, some examples of large sky surveys are given in section 3.1.

Lastly, in section 1.3 we isolated the main different components that contribute to the overall spectral output of a galaxy, and how morphological types, galaxy properties (such as age and metallicity) and galaxy constituents (gas, dust and stellar population types) can be associated to these components and vice-versa. The relationships behind a galaxy's properties and the resultant spectrum of that galaxy are expanded further in section 3.2.

Chapter 2

Wavelet Analysis

Summary

2.1 Noise & Denoising

For any real-world measurement, *noise* is property of that measurement and often serves to complicate the process of measuring; it is manifested as a random addition to the signal that has a different origin from the signal itself. The most common form of noise is one where the measurement $\mathbf{M}$, is composed of the true signal, $\mathbf{S}$ and an additive noise term, η that is Gaussian in nature (has a statistical Gaussian distribution; equation (2.1)). The reason that this is the most common form of noise stems from the Central Limit Theorem, which states that the cumulative independent effects of random variables (each with their own distribution, and not necessarily Gaussian) will approximate to a Gaussian distribution if there are a sufficient number of them. In general the measurement, true signal and noise can be of any dimensionality, however their individual size and dimensionality must correspond to one another. The majority of the work presented here will focus on 1D signals, i.e.: spectra, but is in most cases generalisable to higher dimensions.

$$\mathbf{M} = \mathbf{S} + \eta \tag{2.1}$$

Noise in astronomical observation can come from many sources, including but not limited to: thermal photons, atmospheric effects (for ground-based observations), interstellar radiation (more important for space based observations), electronic noise (due to circuitry), quantum effects, background

sources, and terrestrial sources. In some cases these can be controlled (for example terrestrial observations can be done during the new moon phase in order to minimise the effect of moonshine), but in most cases they cannot be controlled, or are very difficult or otherwise impractical to control. It is important to be able to quantify how much noise is present in any given measurement, relative to the true signal content and commonly this is done via the signal-to-noise ratio (SNR) relating some property of the signal (such as the flux in a specified band, or the integrated flux of an emission line) to some property of the noise (usually the standard deviation). Hence, in order to obtain the best quality signal, particularly when the noise is non-negligible relative to the true signal, we need to be able to remove the noise accurately from our measurements and this is termed 'denoising'.

Denoising procedures attempt to recover the true signal $\mathbf{S}$, given the measured spectrum $\mathbf{M}$, and some knowledge or a model of the noise (η) generally leading to a estimate of the signal, $\mathbf{S}'$ which - if the denoising has been successful - will be very similar to the true signal, i.e.: we wish to minimise $\delta\mathbf{S} = \left| \mathbf{S} - \mathbf{S}' \right|$.

The denoising methods presented below are performed in the space where the signal is measured ('direct' or 'real' space), however, due to the properties of white Gaussian noise, any transformed space that is related to real space by a linear transformation will by definition continue to have white Gaussian noise, and these methods would be equally applicable in that space also. Therefore there exists a potential for exploiting this property if we can find a space where the signal is made more compact and larger (the power is confined to fewer, but larger coefficients) than in real space and where the noise remains non-compact and confined to very many – but crucially also very small – coefficients. Ideally we wish to find some transform, $\widehat{\mathcal{T}}$ such that,

$$\widehat{\mathcal{T}}^{-1} \left| \widehat{\mathcal{T}} \mathbf{M} \right|_\theta = \mathbf{S}' \approx \mathbf{S} \,, \tag{2.2}$$

leaves $\mathbf{S}'$ better approximated to $\mathbf{S}$ (gives a smaller $\delta\mathbf{S}$), under some thresholding condition $\left| \quad \right|_\theta$, than simply performing this same thresholding procedure in direct space. For a real-world measurement however, we will not know the true signal, $\mathbf{S}$, and these methods must be tested and optimised through simulations – where the true signal can be known – prior to application to real data.

2.1.1 K-σ Denoising

One simple denoising method is to perform a K-σ denoising on the spectrum. The principle is simple and relies on the assumption (in the simplest example) that the noise is white and Gaussian, and flat across the entire spectrum. If this is the case, then the standard deviation of the noise, σ, can be estimated, and each point on the spectrum can be compared to some test statistic based on this value, and subsequently accepted as likely to belong to the true signal or rejected (thresholding). This procedure is commonly used in other applications to remove outliers from a data-set, however it can be adapted such that it keeps outliers (since we expect these to be signal) and rejects all other pixels.

The general procedure is to calculate a smoothed version of the spectrum, this can be done by running a window across the spectrum and calculating the mean within that window (other statistical averaging methods with their own advantages and disadvantages can be used instead, e.g.: the median). The window is generally chosen to be an odd number of consecutive pixels on the spectrum in order to have a well defined centre and the mean at each stage is assigned to this central pixel. An issue arises at the edges where the first few pixels cannot have a whole window calculation, this is often resolved by an interpolation/fitting or a simple replication/mirroring of the first full-window

values, an equally valid technique is truncation, though this does reduce the size of the data being processed. It is common for analysis methods to struggle at edge-regions though this is due to the deficiency of the data at & 'beyond' the edge, rather than the method itself.

The standard deviation of the mean-subtracted spectrum (the original spectrum minus the smoothed spectrum) gives an estimate for the noise standard deviation (i.e.: an estimate of σ) of the signal. Then, the mean-subtracted spectrum can be thresholded with respect to a critical threshold, θ. Different types of thresholding can generally be used at this point, however the most common method is that of a hard thresholding were pixels in the mean-subtracted spectrum are set to zero if they are less than the stipulated threshold, θ. The denoised spectrum can then be reconstructed by adding the now thresholded mean-subtracted spectrum to the mean/smoothed spectrum.

This threshold is most often defined in multiples of the (estimated) noise standard deviation, hence $\theta = \mathrm{K}\sigma$. The reason for this is to better associate individual pixels as to how likely they are to have come from the distribution of the noise, and to have a value that is determined by the data, relative to the data. For example, a threshold of 2σ would leave unthresholded those pixels that have at most a $\sim 5\%$ individual chance of belonging to the noise distribution (and would thus be very unlikely to be noise pixels), and threshold everything else to zero. Anything below the threshold is undecidable as to whether it belongs to the true signal, or merely to the noise (all of these pixels have a probability of belonging to the noise distribution that is greater than 5%), and so we are forced to assign those pixels to the noise.

The value of K, which determines where this threshold is placed is chosen at the discretion of the user, a larger value indicating a more stringent requirement for accepting a pixel to belong to the signal (it has to be extremely unlikely that the pixel could belong to the noise distribution), and a smaller value conversely indicating a more relaxed criterion which would allow more true signal pixels to be accepted as such, however this is at the expense of also accepting spurious noise pixels.

K-σ denoising is a relatively blunt tool since a signal containing any features within the boundaries specified by K-σ, will naturally lose these features in the process. Additionally, the features that remain cannot be guaranteed to belong to the true signal, since, by the nature of K-σ denoising, there is still a small probability that they belong to the noise distribution and have been erroneously accepted as signal. Evidently K-σ denoising becomes less effective with an increasing noise component to the measurement (for the same relative strength of true signal), and will increasingly fail to discriminate features that lie within this threshold.

A simple K-σ denoising algorithm performed in pixel-space:

1. For a specific window length (an odd number of pixels wide, for example, 3), w, calculate the local mean, m_w (alternatively, calculate the local median).

2. Assign m_w to the central pixel within this window.

3. Shift the window 1 pixel further along the spectrum and repeat steps 1 - 3 until reaching the end of the spectrum, producing a spectrum that is a smoothed version of the original (noisy) spectrum, c - the mean-spectrum, $\overline{m}$.

4. Extrapolate the edge regions where the initial and final few pixels were beyond the first full window, or truncate.

5. Compute the mean-subtracted spectrum, $\delta = c - \overline{m}$.

6. If $\left|\,\delta_i\,\right| \leq \mathrm{K}$ then we cannot confidently identify if this pixel is noise or signal, so we assume it to be noise and set the i^{th} pixel to zero.

7. If $\left|\,\delta_i\,\right| > \mathrm{K}$ the we can assume the i^{th} pixel is likely to be signal, and leave it unchanged.

8. Reconstruct the denoised signal, $c' = \left|\,\delta\,\right|_{\mathrm{K}\sigma} + \overline{m}$

With the K-σ method, each pixel is tested separately, and at a threshold of 2σ for example, the individual probability of a misclassification (of a noise pixel mistakenly being accepted as signal) is small ($\sim 5\%$) for any given pixel; however, when considering all the pixels in a given spectrum the likelihood of misclassification increases significantly. Increasing the threshold will result in fewer misclassifications of noise pixels as signal, the strong disadvantage being the necessary classification of further signal pixels as noise. There will always be an overlap where pixels of similar values are the result of noise in some cases, but signal in others (or indeed combinations of both signal and noise), and thresholding results in an inevitable compromise between incorrectly accepting noise pixels as signal, or incorrectly declaring signal pixels as noise.

2.1.2 Non-stationary Noise

Frequently with real data (as in the case of spectrographs) whilst the noise is often Gaussian (or at least to a very good approximation, Gaussian), the intensity of the noise varies in each pixel as a consequence of the instrumental response and background sources, and as such the noise is more complicated. Unlike the K-σ method, a local standard deviation of the measurement within a sliding window will not provide a good estimate of the local standard deviation of the noise. This apparent stumbling block can however be overcome if the instrumental response is known. In general manufacturers of spectrographs can test and model the response properties of their instruments, yielding an 'error-curve' which can be taken to be a good representation of the standard deviation of the noise at each pixel, which can in turn be substituted in step 4 in the previous algorithm in order to perform a similar K-σ denoising. Similarly, the error-curve can be used whenever a denoising requires an estimate of the variance of the noise in any given pixel.

2.1.3 The False Detection Rate Method

The False Detection Rate (FDR) method is an altogether more sophisticated option for denoising. This method, first developed by Benjamini and Hochberg (1995)[1] (and introduced to the Astrophysics community by Miller et al. (2001)), allows us to control the average fraction of false detections obtained, over the total number of detections, through the use of an adaptive threshold, α. A false detection, is defined simply: when a pixel containing predominantly noise is incorrectly accepted (under some test statistic and predefined threshold) as being a pixel that contains predominantly signal.

Spectra can contain many pixels, and for any given spectrum we do not know how many of them are signal pixels and how many are noise. Using the K-σ approach described previously, even with each individual pixel only having a probability of being a false discovery of 5% (at 2σ), a situation where there are for example 300 signal pixels together with 2,700 noise pixels we could still expect to encounter 150 false detections in the 3000-pixel measurement, which leads to a very high overall number of false detections (particularly compared to the number of signal pixels) despite each pixel individually having a low probability of being a false detection, leading to unreasonably high

[1] FDR is termed here, and in many other places in the literature as *False Discovery Rate*. The name used here is intended to be a more intuitive substitute, and does not represent a difference in the method.

contamination of the recovered signal. In practice we often do not know how many pixels are signal and how many are noise, however, we can say that a simple K-σ thresholding will generally result in a rate of false detections for the entire spectrum that is significantly larger than might be expected given a reasonable limit on an individual pixel. One can alternatively set a threshold so that it is scaled relative to the number of pixels ($\alpha' = {}^{\alpha}/_{N}$), with more pixels requiring a stricter threshold (the Bonferoni method). This guarantees a maximum at α for the number of false detections; however this results in the unnecessary rejection of a large number of signal pixels, with the situation progressively degrading with increasing numbers of pixels.

The False Detection Rate (FDR) is defined as the fraction of pixels incorrectly accepted as signal relative to the total number of pixels accepted as signal. A threshold α is chosen such that its value is between 0 & 1. The FDR is then guaranteed to be less than or equal to alpha *on average* (i.e.: $\langle \text{FDR} \rangle \leq \alpha$). The general procedure is to construct p-values from the test statistic for each of the N pixels. The p-value is the area under the probability distribution curve of the null hypothesis (that the pixel is noise) that extends beyond the measured pixel, alternatively it is the probability that the measurement in that pixel could have been more extreme than the measured value (as a result of chance). Evidently measured pixels that are already at the tail of the null distribution (and are therefore more likely to be signal) will have smaller p-values, it is comparatively less likely that that pixel could have had a more extreme value than the one measured.

Denoting the list of *ordered* p-values as $p_1, \ldots, p_N$, we can define θ such that,

$$\theta = \max \left\{ n \,\middle|\, p_n < \frac{n\alpha}{N\mu_N} \right\} \tag{2.3}$$

where $\mu_N = 1$ in the case of statistically independent tests, or $\sum_{k=1}^{N} k^{-1}$ when the tests are dependent, and alpha is our chosen FDR threshold (the allowed fraction of false detections).

Upon obtaining a value for θ we can then declare all pixels whose p-values are less than or equal to p_θ to be signal pixels, irrespective of their individual p-values with respect to the threshold, α. The proof of this (and that this leads to a guaranteed bounding on $\langle \text{FDR} \rangle$ of α) is technical and not necessary to show here, but it can be found in Benjamini and Hochberg (1995); Benjamini and Yekutieli (2001).

It should be noted that whilst it is compelling to associate the FDR threshold α with the equivalent K-σ threshold in a K-σ denoising, they are fundamentally different and have no basis for a direct association. A '2σ' thresholding is *not* equivalent to setting an FDR threshold of $\alpha = 0.0455$.

2.2 From Fourier Series to the Continuous Wavelet Transform

In order to understand the origin of wavelets and wavelet analysis, it is first necessary to briefly look at its natural ancestor, Fourier Analysis. Fourier analysis was originally developed by Joseph Fourier, in the early 1800s, as a tool for representing periodic functions as a sum over a basis set of sine and cosine functions. He demonstrated that any real, integrable and periodic function, $f(x)$, could be expressed as the sum of a series of sine and cosine functions:

$$f(x) = a_0 + \sum_{n=1}^{\infty} \Big(a_n \cos(\phi_n x) + b_n \sin(\phi_n x) \Big), \tag{2.4}$$

where a_0, a_n, and b_n, are constants which can be obtained from integrations of the function $f(x)$, and $\phi_n = 2\pi\, {}^{n}/_{L}$ with L representing the period of the function $f(x)$. In the general case this expression

is only exact for an infinite number of terms.

Utilising Euler's formula, equation (2.4) can be rewritten as follows,

$$f(x) = \sum_{n=-\infty}^{\infty} A_n \, e^{i \, \phi_n \, x} \tag{2.5}$$

where $i = \sqrt{-1}$, and A_n is related to the previous constants (a_0, a_n, and b_n) as

$$A_n = \begin{cases} (a_n - ib_n) & n > 0 \\ (a_n + ib_n) & n < 0 \\ a_0 & n = 0. \end{cases}$$

and can be derived from

$$A_n \equiv \frac{1}{L} \int_{-L/2}^{L/2} f(x) \, e^{-i \, \phi_n \, x} \, dx$$

In the limit of taking continuously valued periods of sine and cosine basis functions (instead of integer periodicities as determined by n), equation (2.5) becomes the (inverse) Fourier Transform,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(k) \, e^{\, i \, kx} \, dk \tag{2.6}$$

with the associated (forward) Fourier Transform,

$$\widehat{f}(k) = \int_{-\infty}^{\infty} f(x) \, e^{\, -i \, kx} \, dx \tag{2.7}$$

The effect of a Fourier Transform (FT) is to take any real, continuous and integrable function and express it in a different basis, characterised by *frequencies* in a sinusoidal basis instead of the usual coordinates (multiples of the basis vectors $[1, 0]$ and $[0, 1]$) in a Cartesian basis. Assumed properties are the stationarity of the signal, and that the basis functions (in real space), as well as the signal, are *infinite* in extent. This transformation yields a different *representation* of the function, and whilst FTs will not change the information content of the function or signal, they can often highlight information that is not easily seen or understood in the original basis. A simple example of this can be seen in figure 2.1.

There exist multiple limitations to the Fourier Transform that resulted in its adaptation into the Short-Term Fourier Transform (STFT; equation (2.8)). The STFT attempts to overcome the inherent problems with the general FT by the addition of a 'window' to the transform (utilising a window function, $g^*(x - \varepsilon)$, centred on ε), in order to limit the region of the signal being transformed such that within that window, stationarity can be assumed (the FT is not generally applicable to signals that are not stationary). In general signals may contain many basis frequencies, but they do not necessarily contain these basis frequencies in all regions of the signal, indeed most real signals are generally non-stationary and finite in extent, making a simple FT sub-optimal. Indeed the FTs of a pair of signals containing the same basis of sinusoids of differing frequencies – one such that all the basis sinusoids are present in all regions (such as in figure 2.1), the other such that each of the basis sinusoids exist in independent regions, but adjoin with one another (i.e.: non-stationary) – will share near-identical Fourier Transforms.

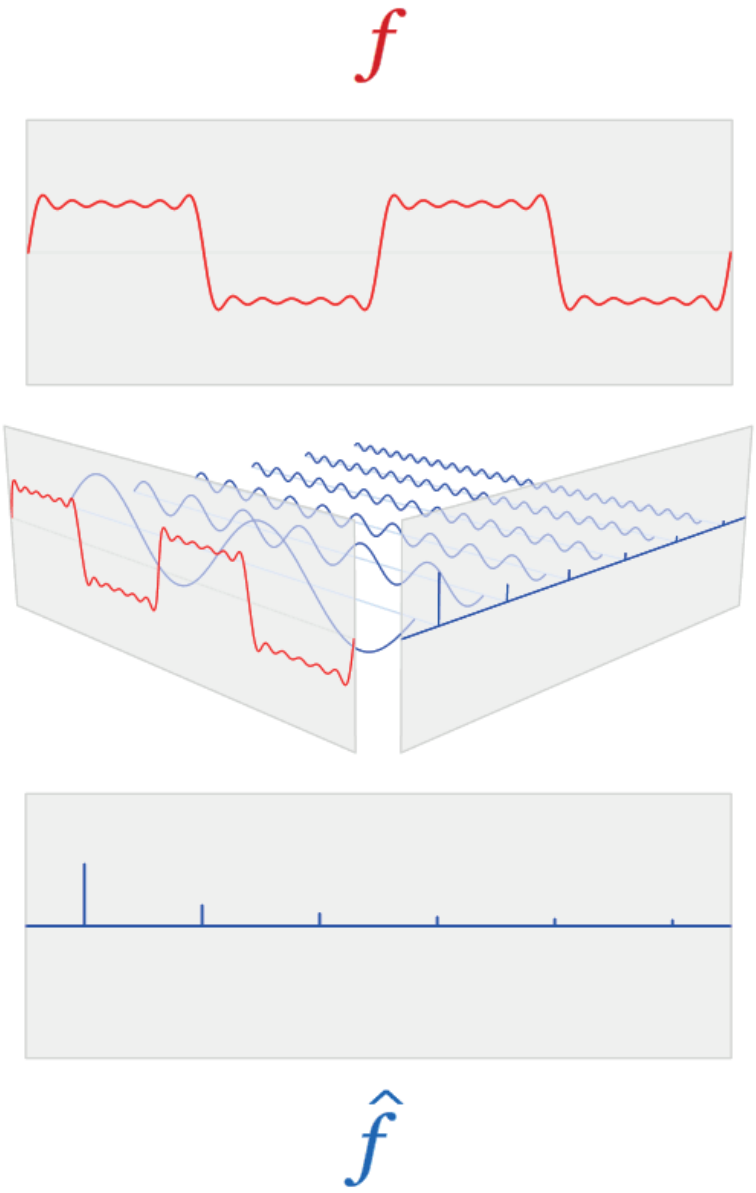


Figure 2.1: This figure shows a simple function, f, constructed from the addition of a handful of sinusoidal functions in real space, and its resultant Fourier Transform, $\widehat{f}$, depicting the individual frequencies (and the magnitude) of each of the sinusoids in Fourier space which compose the original function. Adapted from: http://upload.wikimedia.org/wikipedia/commons/5/50/Fourier_transform_time_and_frequency_domains.gif

$$\widehat{f}_{STFT}(\varepsilon, k) = \int_{-\infty}^{\infty} f(x)\, g^*(x-\varepsilon)\, e^{-i\,kx}\, dx \tag{2.8}$$

The STFT attempts to give a picture of the signal that is dual in its representation, containing simultaneous information from both the x and k domains. The size of the window has important implications for the informativeness of the resulting transform. Wide, and the transform approaches the original Fourier transform without any window, meaning the result will have excellent resolution

in the k domain, but very poor resolution in the x domain; narrow and the transform does little to modify the original signal, and it retains excellent resolution in the x domain, but poor resolution in the k domain. For many classes of signal this fundamental limitation to the STFT is problematic: to obtain good resolution in either domain, correspondingly requires the sacrificing of resolution in the opposite domain, and such a choice is static over the entire transform and signal. For some signals (multiple oscillatory features in different locations, for example) the STFT remains the better choice, however, for the signals we are considering – galaxy spectra – the STFT is far from ideal.

This fundamental problem of simultaneous precision in corresponding domains was made famous (from a physical perspective) by Heisenberg (1927) and was to become known as the Uncertainty Principle. Simply, the Uncertainty Principle states there exists a fundamental limit in precision between two 'conjugate' (physical) variables (in this case x and k), with it being impossible to specify both simultaneously (to an arbitrarily high precision), $\Delta_x \Delta_k \geq constant$, with the exact value of this constant dependent upon the different domains chosen. There is always a limitation to signal *intervals*, and frequency *bands*, and in signal-processing this same principle is termed the *Gabor Limit*.

This unsatisfactory quality of the STFT arises from its application of a fixed window combined with the infinite-in-extent basis functions (in real space), which result in an egalitarian gridding of identical 'tiles' on the x-k plane (a fixed window size necessarily implies a fixed Δ_x and Δ_k), termed the *localisation*. This necessarily means that signal will suffer significant losses in resolution if this window size is not made appropriate for all the signal content.

In general components of the signal that are present for its entire span do not need fine resolution in the x domain; similarly, transient features will require sharp resolution in the x domain. What is required is a transform that possesses a degree of selective localisation, with the tiling changing to suit the situation. By definition the area of each tile in the grid must remain the same (due to the Uncertainty Principle) but we are not restricted in how that area is constructed (the tiles need not be identical, however, they must fully cover the plane without overlapping) - we can selectively favour finer resolution in one domain at the expense of the other and vice versa, leading to a gridding that is no longer egalitarian, covering a greater expanse in the one domain whilst the other is small, trading the respective lengths and widths of the tile such that the situation becomes reversed when the other domain becomes large, thus obtaining a *multiresolution* analysis of the signal. This idea is depicted pictographically in figure 2.2.

The Continuous Wavelet Transform (CWT) can obtain a a simultaneously dual representation of an input signal (if the signal is 1D, then the output of the transform, called the *scalogram* will be 2D) resulting in an output that is similar in nature to figure 2.2, but continuous. This dual representation is particularly useful since it can illustrate features more easily from the data than a complete representation in either domain respectively, this can be seen in figure 2.3.

The CWT can be analogised to the STFT by associating terms between them, yielding the *mother wavelet*, ψ under a substitution from the frequency, k, to its inverse, the **scale**, s,

$$g^*(x-\varepsilon)\, e^{-i\,kx} \Rightarrow \frac{1}{\sqrt{s}}\, \psi^*\Big(\frac{x-\varepsilon}{s}\Big) \tag{2.9}$$

giving the CWT,

$$W_{CWT}(s,\varepsilon) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(x)\, \psi^*\Big(\frac{x-\varepsilon}{s}\Big)\, dx \tag{2.10}$$

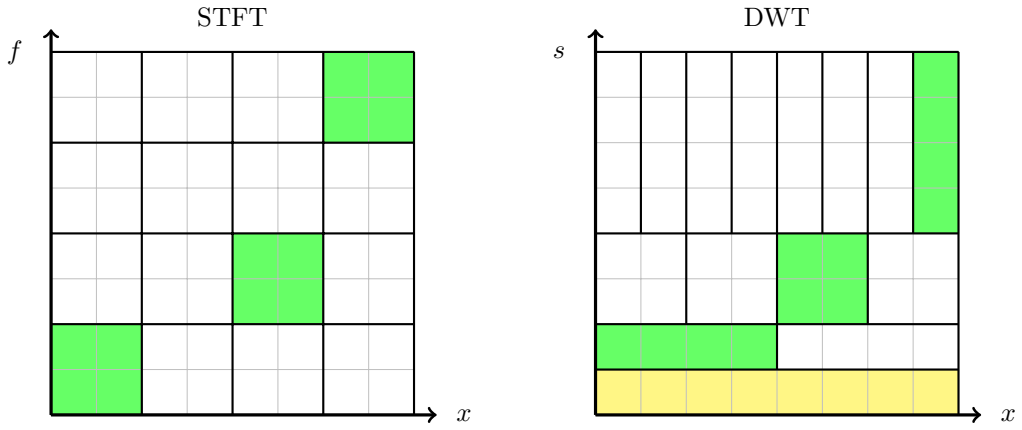


Figure 2.2: This figure shows a schematic of a tiling of the time-frequency plane for a STFT, and the equivalent tiling for a Discrete Wavelet Transform (DWT). Note that in a DWT the concept of frequency is exchanged for one of *scale*, which can be considered as the inverse of frequency. In the STFT case, each tile has the same area, and the same proportions, meaning that for some regions in the plane, the resolution in one or other domain will be insufficient to well describe the signal. In contrast the DWT manages to ameliorate this by changing the proportions of the tiling, and thereby optimising the localisation, but with each tile maintaining the same area (shaded green) offering a multiresolution analysis for the signal. The yellow shading corresponds to the original signal at a scale value of 0.

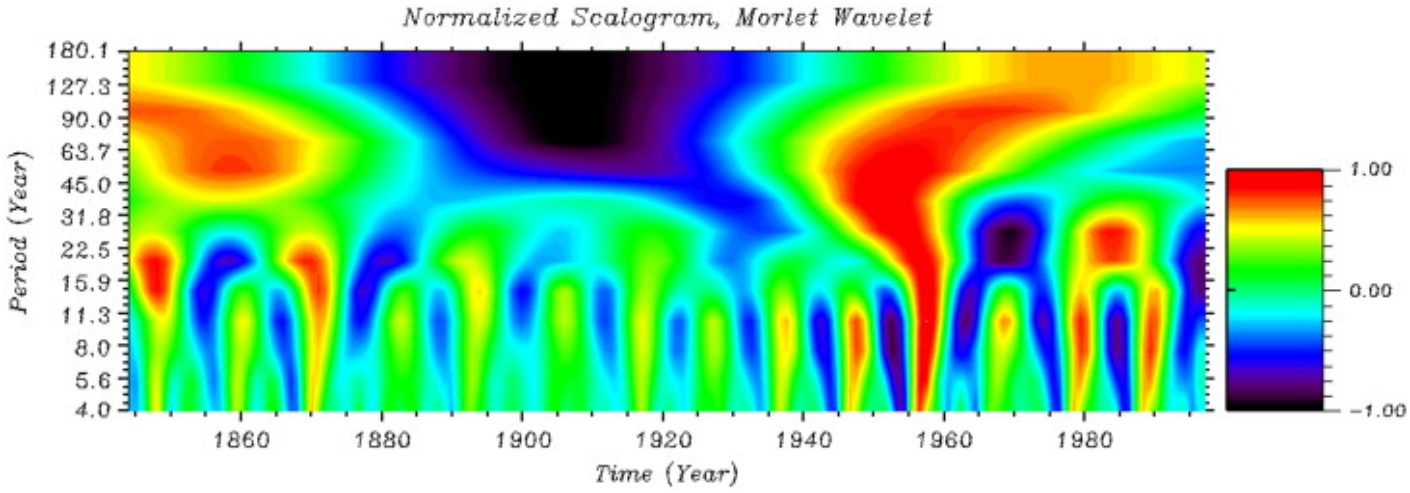


Figure 2.3: This scalogram shows the sunspot activity as a function of time (in years) and scale (as period/inverse of frequency) and is the result of a CWT (in this case using a Morlet mother wavelet). Easily seen is the approximately 11-year solar activity cycle, less easily seen, but highlighted with the scalogram is a longer-period cycle (on the order of a century) that appears to couple with the 11 year cycle. Image from: de Jager et al. (2010).

where s and ε represent scale and translation parameters. The inverse CWT is similarly defined,

$$f(x) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sqrt{s}} W_{CWT}(s,\varepsilon)\, \psi\Big(\frac{x-\varepsilon}{s}\Big)\, \frac{1}{s^2}\, d\varepsilon\, ds\,, \tag{2.11}$$

where C_ψ is a constant derived from the integral of the Fourier transform of the wavelet, Ψ, and as such is dependent upon the wavelet used. Proceeding with the inverse transform (i.e.: reconstruction)

is only possible if this constant is finite, with this being termed the admissibility condition. Tied into this condition is the property that $\Psi(0) = 0$, and hence the wavelet must have zero mean (hence oscillatory). Due the fact that the CWT is *redundant* the inverse transformation in equation (2.11) is not unique.

The CWT can readily be represented as a convolution between the signal and the mother wavelet, where $\overline{\psi}_s(\varepsilon) = (1/\sqrt{s})\psi^*(-\varepsilon/s)$,

$$W_{CWT}(s,\varepsilon) = f * \overline{\psi}_s(\varepsilon) \; . \tag{2.12}$$

which can also be expressed as direct multiplication in the Fourier domain as,

$$\widehat{W}_{CWT}(s,\mu) = \sqrt{s}\,\widehat{f}(\mu)\,\Psi^*(s\mu) \; . \tag{2.13}$$

where μ is the Fourier domain counterpart to s, in an analogous way as x and k.

Many different mother wavelets exist, each with different properties and yielding different wavelet 'families', however, they must all obey certain conditions: they must be square integrable/have finite energy (i.e.: $\int \psi^*(x)\psi(x)\,dx < \infty$, generally normalised such that this value is equal to 1.), they must be oscillatory (have zero mean), and they must be compact (not infinite in extent) both in real and Fourier space. The scale parameter allows the scanning of the wavelet to pick out details at different resolutions in an analogous way to scale on a map: at large scale the resolution on structures at the level of streets and buildings is irrelevant, the only relevant information is the large scale behaviour of the terrain; conversely, at small scales, the most important features are streets and buildings, terrain ceases to be important since at these scales it will generally change very little. A CWT is analogous to being able to perform a continuous zoom on a map, whereas a DWT is more similar to a collection of individual maps of the same region, but each at different (fixed) levels of zoom. Since many real signals possess this property of not requiring a high level of detail for a global view of the signal, wavelet analysis has many real-world applications.

2.3 Discrete Wavelet Transforms

In order to obtain a discrete version of the wavelet transform (for practical computational purposes), one option would be to compute the CWT with selected choices of scale and translation parameters. This is equivalent to a discrete sampling of the CWT in the time-frequency plane. This sort of discretisation, however, is not ideal for computational purposes since many of the transforms are redundant. The origin of this redundancy arises from the fact that any discrete sampling of the signal decomposition in the time-frequency plane cannot guarantee that those particular choices of scale and translation yield an *orthogonal* basis of wavelet functions. It is this orthogonality condition (specifically, orthonormality) that is the key to obtaining a non-redundant transform, and consequently a more compact representation.

In order to obtain a non-redundant transform therefore, the wavelet basis must be carefully selected beforehand such that – much like with a Fourier series – the individual basis functions can be selected to be orthonormal. Orthonormality is defined on (basis) functions when their inner-product is zero if they are different, and identical to 1 if they are the same basis function,

$$\langle \phi_i \, , \, \phi_j \rangle = \delta_{ij} \; , \tag{2.14}$$

where ϕ_i & ϕ_j represent basis functions (of the same set) and they can only be considered to be a set

of basis functions if they span the entire space in which we want to decompose a function.

Recall that the goal for the DWT is to obtain a representation of our signal that has the property of multiresolution, we wish the coarsest scale – the 'terrain' – to be poorly detailed (since generally, it is not necessary that it be highly detailed), but the finest scales to be highly detailed. Alternatively, we wish to isolate *particular* tiles in the x-s plane, as shown in figure 2.4

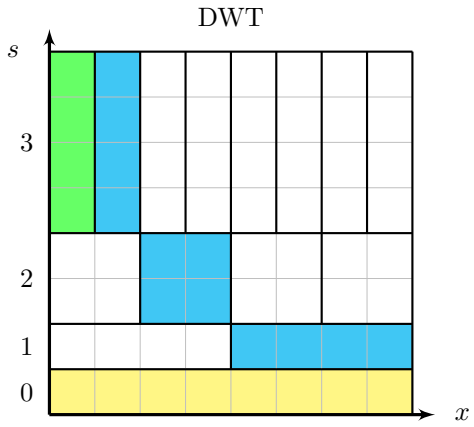


Figure 2.4: The blue and green highlighted tiles ($s = 1, 2, 3$) in this partitioning of the x-s plane are the ones that correspond to a *multiresolution* picture, and are therefore the tiles we are interested in. The remaining tiles are *redundant*, in the sense that all their information content is reproduced in other tiles; specifically, these tiles contain all the necessary information to construct a full picture of the signal (highlighted in yellow, $s = 0$). This particular tiling approach is termed *'dyadic'* since the lengths and widths of each tile are related by factors of two to the ones preceding or succeeding it.

The goal for a discrete wavelet transform therefore is to have a basis set of functions based on a *mother* wavelet, where each member of that basis set can be constructed from the mother wavelet via a choice of scale and translation parameter, with the condition that all members of the basis set are mutually orthogonal, and normal. One common way of achieving this (though this is not the only way) is via a dyadic construction, where wavelet bases each have a length set to an integer power of 2, achieved by (multiple-)halving of the mother wavelet (itself with a length expressible as 2^n). Ideally the function which we wish to decompose should have a length of 2^n, with the size of the mother wavelet then being chosen to match; evidently this is not always possible, however, the original signal can always be modified to fit such a length with resampling and padding being possible examples of how to achieve this.

The discrete wavelet is then of the form,

$$\psi_{j,a}(x) = \frac{1}{\sqrt{s_0^j}}\,\psi^*\Big(\frac{x - a\,\varepsilon_0\,s_0^j}{s_0^j}\Big)\,, \tag{2.15}$$

where a and j are integer values of translation and scale, and s_0 and ε_0 are often chosen to be $s_0 = 2$ and $\varepsilon_0 = 1$ in order to result in the aforementioned dyadic sampling of the x-s plane.

With this sort of construction however, continuous halving and translating of the wavelet bases cannot encompass the entirety of the signal with a finite number of bases; this is equivalent to being unable to cover the entire length of the x-axis in figure 2.4 with blue tiles alone, being restricted to halving and translating each time there is always a remainder (green tile). The solution to this problem is via a *scaling function* which acts as a 'bridging' mechanism to compensate for the remainder of

the transform after a finite number of wavelet scales have been computed, with this being dependent upon the size of the initial signal, the scaling function thus corresponds to coverage of the green block in figure 2.4. The scaling function can be derived from the wavelet (or vice-versa) since they must form a corresponding pair of transforms to partition the x-s plane, i.e.: at scales greater than 1, the blue blocks corresponding to the wavelet transform must first have passed through a scaling block (the block immediately beneath it) in order to shrink the wavelet to fit the new, shorter input (in this dyadic system, halved) hence the 'scaling' in scaling function.

A common method of interpreting the DWT is to revisit the concept of associating the transform to a (discrete) convolution as in equation (2.13). A discrete convolution is equivalent to a frequency-pass filter, and a multiresolution wavelet transform can then proceed as a series of bandpass filters as in a *filter bank*. Filters are generally termed high-pass if the result is comprised of high frequencies, or low-pass if the result consists of low frequencies; as such we can consider a high-pass filter in frequency as a 'small-pass' filter in scale and a low-pass in frequency as a 'large-pass' in scale.

The general construction for a multiresolution transform as a filter bank (assumed dyadic) is shown in figure 2.5, where the filtering shown represents what is known as *analysis* (i.e.: a forward transform/decomposition into a wavelet basis). The inverse transform (reconstruction) is easily obtained by repeating the procedure in reverse, also known as *synthesis*, depicted in figure 2.6. The wavelet coefficients, ω, may be treated prior to reconstruction, with for example, a particular choice of thresholding.

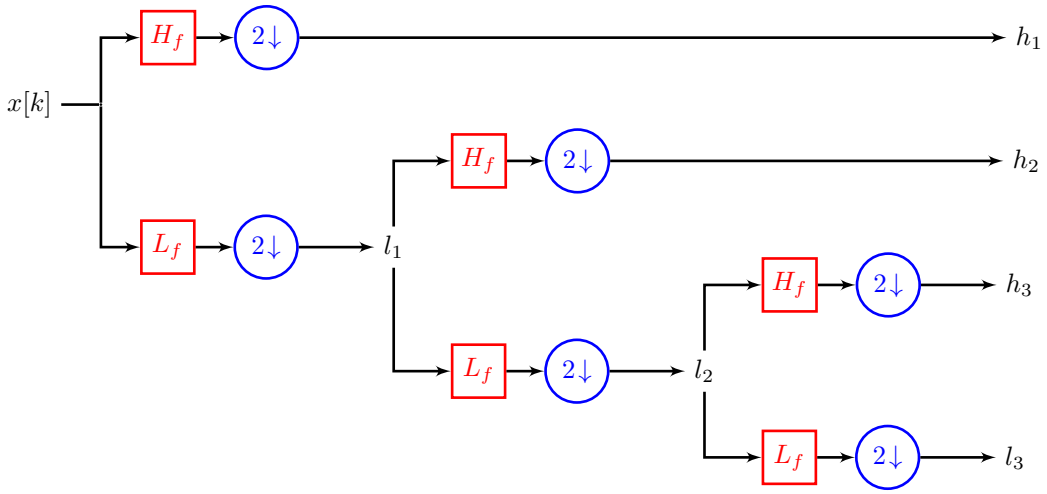


Figure 2.5: Analysis Filter Bank – The DWT proceeds through a series of filters, here labelled L_f, a low-pass (i.e.: high scale) filter and H_f, a high-pass filter. The $2\downarrow$ operation is termed *downsampling* and involves the alternating selection of half of the entries (with the others being discarded, resulting in a net shrinkage of the size of the output) of the result of the transform since this result would otherwise be oversampled. The outputs of the transform, the set $\omega = \{h_1, h_2, h_3, l_3\}$ are termed wavelet coefficients.

The filter bank shown in figure 2.5 is a 3-tiered filtering, thus representing a wavelet transform with 3 scales, the H_f filter is equivalent to a discrete wavelet transform similar to the one in equation (2.15) and the L_f filter is the accompanying *scaling function*, ϕ. The process can generalised to any number of scales, however, the downsampling steps restrict this to an absolute maximum dependant upon the size of the input signal, $x[k]$.

The wavelet coefficients, ω, produced by the DWT analysis bank (figure 2.5) possess a direct correspondence to the blue and green highlighted tiles in figure 2.4, namely the blue tiles correspond to the coefficients h_i and the green tile to the final coefficient l_n where n represents the number of scales, which in this case is 3. Furthermore this architecture for the transform is one with only a minimal coverage (maximally non-redundant) of the x-s plane, and different architectures, where the branchings and filter blocks are located, can be constructed depending on the desired application with the possibility of accepting some redundancies in exchange for other useful properties such as isotropy or the forgoing of downsampling/decimation.

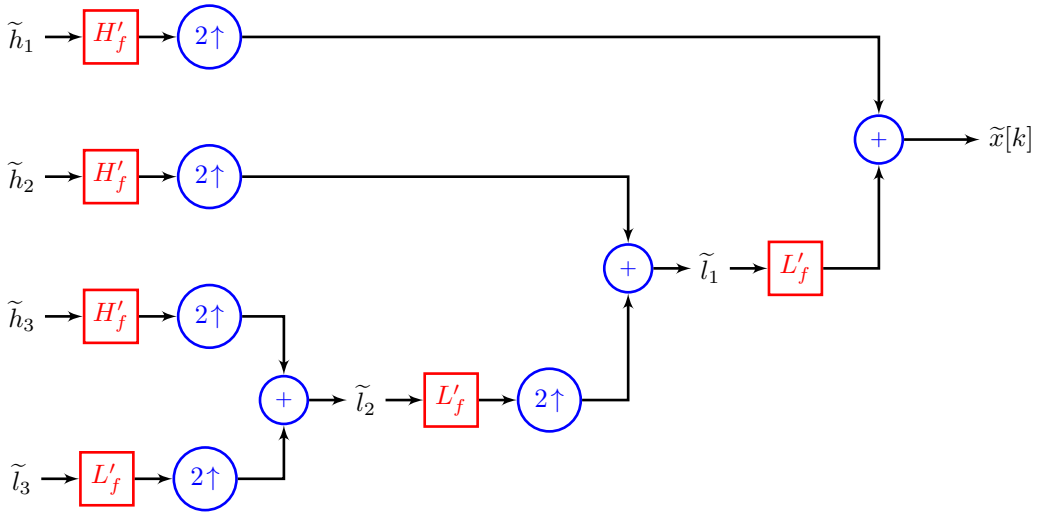


Figure 2.6: Synthesis Filter Bank – Reconstruction of the signal from the set of wavelet coefficients, ω. L'_f and H'_f represent the inverse filters of L_f and H_f respectively. The $2\uparrow$ operation is an upsampling and involves the simple insertion of zeros between each entry; the subsequent coaddition of the outputs requires the recombination of odd and even samples. The inputs to the reconstruction (the wavelet coefficients, ω) may be treated prior to the reconstruction process, generally with some sort of thresholding.

2.3.1 The Haar Wavelet Transform – A Simple Example

In order to demonstrate the general process of a discrete wavelet transform (DWT), a simple example using a *Haar* wavelet is shown below. Haar (1910) invented the first (and simplest) wavelet function, a compact step-function on the $[0, 1]$ interval as in figure 2.7.[2] An immediate and important point to note is that this wavelet has discontinuities and, as such, no continuous wavelet analogue exists (being discontinuous the Haar wavelet would violate the condition required for a finite extent in Fourier space). The Haar wavelet remains applicable to discrete cases, and although simple, it is often regarded as inferior to other wavelets for most practical applications where signals are continuous.

In order to demonstrate the application of the Haar wavelet transform, we use the following 8-entry signal, $x[k] = \{-1, -1, 4, 4, -5, -3, -2, -1\}$ as shown in figure 2.8. Convolving this signal with the wavelet results in finding halves of the differences between entry pairs in the original signal. This pairing of entries to yield one output automatically gives a halved number of output entries

[2]Haar's function was not recognised to be a wavelet until the later realisation that many different transforms had similar properties and were rooted in the same mathematics and the subsequent development of wavelet theory many decades later.

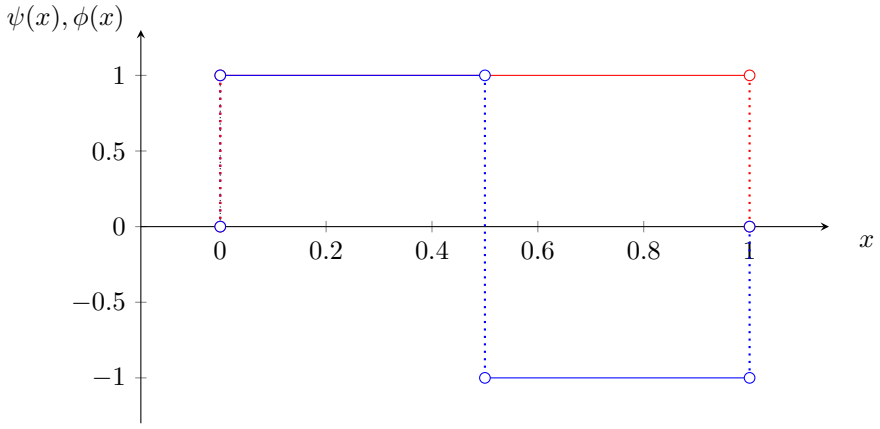


Figure 2.7: The mother wavelet of the Haar wavelet transform (blue) and the scaling function (overlaid in red). Note that the wavelet itself is discontinuous, and thus non-differentiable.

(downsampling is implicit). Convolution with the scaling function returns the mean between pairs of values. The example chosen here is a modified version based on one by Selesnick (2007).

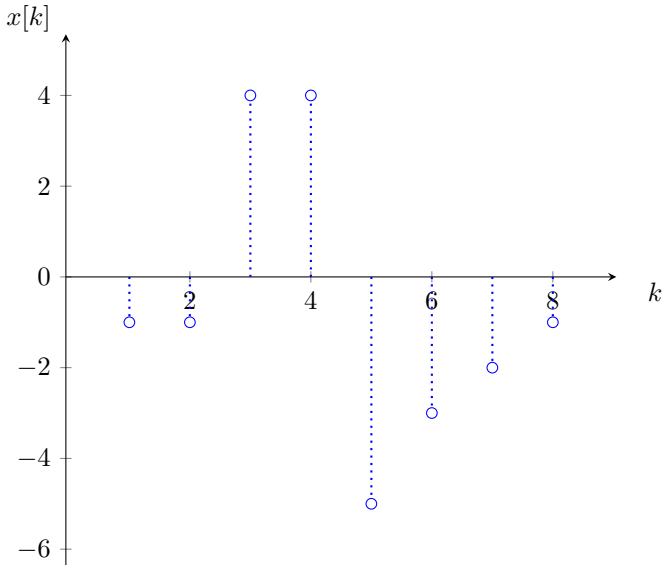


Figure 2.8: The example input signal to the Haar analysis filter bank, indexing runs from $k = 1$ to 8.

Pushing this signal through the analysis filter banks yields $h_1 = \{0, 0, -1, -0.5\}$, $h_2 = \{-2.5, -1.25\}$, $h_3 = \{2.125\}$ and $l_3 = \{-0.625\}$, which taken together yield the full set of wavelet coefficients, $\omega = \{h_1, h_2, h_3, l_3\} = \{0, 0, -1, -0.5, -2.5, -1.25, 2.125, -0.625\}$. It is possible to then impose a thresholding on these coefficients prior to reconstruction, in this example where $\left|\omega_i\right| < 2.125$ we set to zero prior to reconstruction, this is a type of hard thresholding. This yields a reconstruction (by alternate addition and subtraction of wavelet coefficients proceeding through the synthesis filter bank) from a modified set of wavelet coefficients, and is shown in figure 2.9. It is important to note that the entire 8-entry signal has been reconstructed from only two non-zero wavelet coefficients (after thresholding), preserving the principal features in the original signal – the step up and step down.

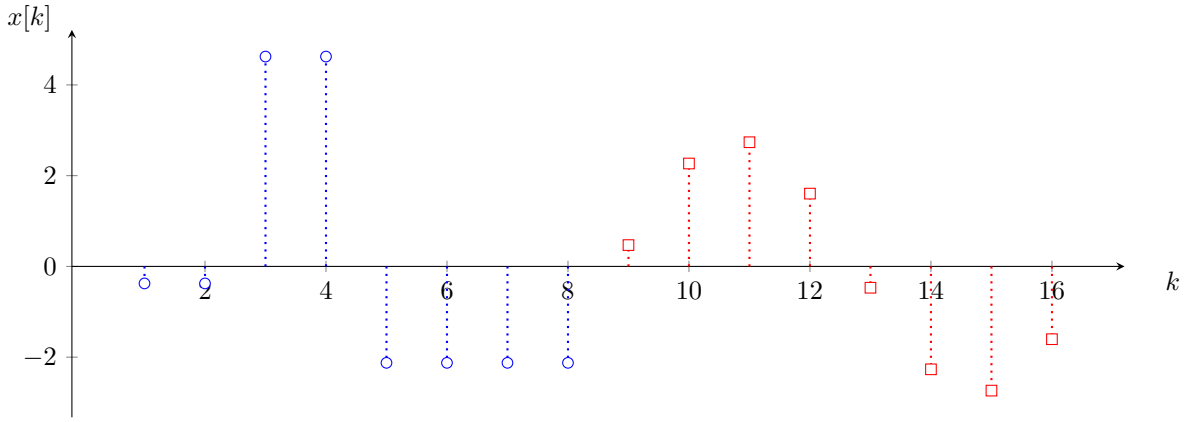


Figure 2.9: The example reconstruction of the original signal figure 2.8, after a thresholding of the coefficients ω_i and passing them through the synthesis filter bank (blue circles). Shown to the right of wavelet transform reconstruction is the equivalent discrete Fourier transform reconstruction (shifted to the right) where again, the components in Fourier space have been thresholded to retain only the two largest values (by magnitude) prior to reconstruction.

The Haar wavelet is well-suited to discontinuous functions, and as such picks out the discontinuities in our signal from $k = 2$ to $k = 3$ and again from $k = 4$ to $k = 5$ very well, however, the section of the signal running from $k = 5$ to $k = 8$ is better represented as a (discrete sampling of) a continuous curve, and as such the reconstructed signal does not perform well in picking up this region, and (as is the tendency for Haar wavelets) the smearing of the region into a single step is the outcome. We can see that the application of a similar process with a Fourier transform instead of a wavelet transform performs quite poorly in comparison, with the result resembling a discretely sampled sine-wave far more than the original input signal.

Often, it is more desirable to process the signal content (such as denoising) via its wavelet representation than via a similar process in the direct space of the signal. The reason for this is *sparsity* – (natural) signals are regularly found to be more compactly represented in wavelet space than direct space: only a handful of the wavelet coefficients, ω_i are found to be large, the others being small or zero. Crucially, these smaller coefficients will often correspond to noise features, and with appropriate thresholding, denoising with wavelets can be very effective.

In this simple example we can already see that the wavelet implementation has the potential to provide significant advantages: feature extraction and compressibility of signals. Both these desirable features arise from the natural property of *sparsity*. As a result of being sparse in an appropriate basis, a continuous signal can be sampled with fewer points (undersampled) than that which could be possible (conventionally) in real space via the Nyquist-Shannon sampling limit (Shannon 1949; Nyquist 2002, a classic paper reprint). This limit states that a signal limited to a frequency band $\pm f_B$ (i.e.: the Fourier transform of the signal exists within these bounds and is zero elsewhere) can be accurately reconstructed provided it is sampled uniformly with a sampling frequency, f_s where $f_s > 2f_B$, below this value aberrations of the signal will occur when reconstruction is attempted due to *aliasing*. Aliasing occurs when an undersampled signal has too few samples in order to definitively identify it, such an undersampling could be the result of many signals each with the same sampled points ('aliases'), but whose overall character is different - initially different and distinguishable signals become indistinguishable if undersampled. As such, undersampled signals when reconstructed will have undesirable components from these aliases.

A sparse representation in a wavelet basis is not constrained by the Nyquist limit in direct space, as such significantly higher compression rates can be found, such as the JPEG 2000 digital image format (Skodras et al. 2001). Sparse properties of real signals has also led to the emergence of entirely new fields such as *compressive sensing* where the signal is recorded in situ in a sparse manner, resulting in far fewer samplings required for the same measurement (resulting in significant savings in both resources and costs) and by extension an already compressed (sampled) signal requiring less storage. Compressive sensing has even led to the surprising development of new devices such as a *single-pixel* camera (SPC, Huang et al. 2013), which operates with multiple randomised sparse samplings with a single pixel which are then combined to form an image. The significant advantage of the SPC is that far fewer samplings are needed: a conventional digital camera contains millions of pixels all of which need to be simultaneously activated to take one image, that data is then compressed for better storage in the camera's memory; an SPC takes multiple measurements (many thousands) and the data collected is already in a sparse format and so crucially does not require energy-intensive compression.

2.3.2 The Starlet Transform

The Starlet wavelet transform, SWT (also known as the isotropic undecimated wavelet transform, IUWT) is another type of DWT, whose scaling function is based on the B_3-spline (Starck et al. 2007, 2010). The Starlet transform uses both different filters (the B_3-spline as the scaling function, and a wavelet function that is derived from the difference of two scaling functions, figure 2.10) and a slightly different architecture in the transform itself, differing from the basic Haar model shown in figure 2.5.

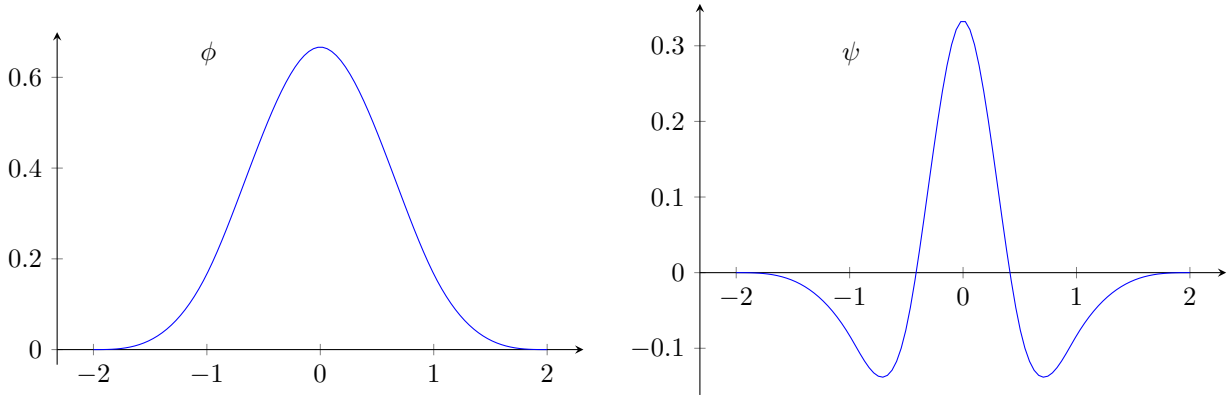


Figure 2.10: These figures show the B_3-spline scaling function (left), and the mother wavelet (right), that are used in the Starlet Wavelet Transform.

The SWT sacrifices both orthogonality and redundancy in order to obtain isotropic properties to the transform. The 'tree' of filters is replaced by a simple and repeated application with the L_f filter however this does not result in the expected downsampling, instead the input signal is left the same length, and gaps are formed between entries, increasing with increasing scale, leading to the 'à trous' algorithm (Holschneider et al. 1989; Shensa 1992).[3] This procedure obtains all the l_i coefficients at each scale in an analogous way to the transform in figure 2.5; the h_i coefficients however are obtained differently: no direct convolution with an H_f filter is necessary, instead these coefficients can be computed directly from the differences between the l_i coefficients, namely $h_{i+1} = l_i - l_{i+1}$.

[3] With 'à trous' being French for 'with holes'.

The scaling and wavelet functions are defined as follows,

$$\phi(x) = \frac{1}{12}\Big(\,|\,x-2\,|^3 - 4|\,x-1\,|^3 + 6|\,x\,|^3 - 4|\,x+1\,|^3 + |\,x+2\,|^3\Big)\,, \tag{2.16}$$

$$\frac{1}{2}\psi\Big(\frac{x}{2}\Big) = \phi(x) - \phi\Big(\frac{x}{2}\Big)\,, \tag{2.17}$$

For this wavelet transform, reconstruction of the signal is extremely simple, for n scales we obtain,

$$x = l_n + \sum_{i=1}^{n} h_i\,, \tag{2.18}$$

where each of the h & l coefficients have a pixel size, k, such that they are the same length of the original signal $x[k]$, similarly to figure 2.4 the signal occupies the '0^{th}' scale.

The advantages of the SWT, given that the transform no longer retains orthogonality nor non-redundancy are significant and render the transform highly applicable to astronomic data sets. These advantages are laid out by Holschneider et al. (1989) and include: a reasonable computation speed; trivial reconstruction (simple summation); the lack of wavelet coefficient shrinkage retaining pixel information at each scale thus making it possible to detect positional information in data without any associated errors or interpolations from the transform; there is a logical progression between scales which is easy to follow and determine; the transform is translationally invariant; and the transform is isotropic. It is these features, with particular importance to isotropy, that make the Starlet Wavelet Transform particularly suited to astronomical data (images in particular) where objects themselves are to a good approximation isotropic.

2.3.3 The Pyramidal Median Transform

The Pyramidal Median Transform (PMT) is a further transform that shares many properties with the DWT, and though it is a multiresolution transform, and has a similar construction to the SWT, it is not strictly a wavelet transform since it employs the use of median functions and is thus rendered non-linear.

The general procedure involves the construction of a median window, d, whose size is chosen to be 3. The first scale is treated as a copy of the original signal (thus $l_1 = x[k]$ and no h_1 is defined). The pre-downsampled coefficients, l'_{i+1} are computed from running this median window across the input spectrum, $l'_{i+1} = med\,(\,l_i, d\,)$, and thus have a size equal to k (the size of the input signal). The h_i coefficients are once again derived from the difference such that $h_{i+1} = l_i - l'_{i+1}$. The l_{i+1} coefficient is then computed from a downsampling, $l_{i+1} = \big(l'_{i+1}\big)_{2\downarrow}$, of the l'_{i+1} coefficient. The procedure is then repeated using the same fixed window until reaching the chosen number of scales.

Reconstruction can proceed with the use of B-spline interpolation provided that good reconstruction is possible (though this assumption does not necessarily always hold). The general procedure involves interpolation in order to 'undo' the decimation step and obtain an output that is twice as large as the input (obtaining the l_i coefficients), and by adding the result to the h_i coefficients one can obtain the preceding l_{i-1} coefficients.

The nonlinearity of the PMT makes noise properties more complicated than in the simpler linear wavelet transforms, and more care needs to be taken in order to denoise properly, however, the nature of the median function automatically provides an inherent robustness to outliers.

2.4 Conclusion

In this chapter we have introduced the concepts of noise and denoising (section 2.1), and presented two methods for denoising, the frequently used K-σ denoising (section 2.1.1) and the less prevalent False Detection Rate denoising (section 2.1.3). The advantages of FDR over K-σ in the extraction of true signal features, particularly in the limit of many pixels, were demonstrated.

Additionally, we introduced the concept of (continuous) wavelet transforms approached from the context of the similar properties that they share with the more commonly used and understood Fourier transform (section 2.2). We demonstrated the efficiency of the wavelet transform in yielding sparse representations of signals, and how it can exceed the Nyquist-Shannon sampling limit for that signal in real space. Furthermore a handful of the many 'real-world' examples of wavelet applications were given, such as the JPEG 2000 standard image format and the single-pixel camera.

The concept of wavelets was then extended to discrete wavelet transforms (DWTs, section 2.3), with examples given for the Haar, Starlet, and Pyramidal Median transforms in sections 2.3.1 to 2.3.3. The tiling of the x-s plane, with only dyadic examples being considered, and their relationship to the coefficients of the DWT was shown in figures 2.2 and 2.4. The integrals of the signal with the wavelet function of continuous transforms become replaced by a series discrete convolutions of the signal with the wavelet as a series of filter-banks, with these being constructed from the wavelet function and an additional scaling function, whose purpose is necessary for discrete wavelet transforms.

We saw how signal-processing - in particular, denoising - could be optimised by being performed in wavelet space, whereby signal content is confined (mostly) to a few large coefficients, and (Gaussian) noise retains its distributed nature across all the coefficients. The total 'power' of the signal is not changed by such a transform, it is merely reorganised to be contained in fewer coefficients - making these coefficients larger overall when compared to pixels in real space; noise does not undergo this compression, retaining its distributive nature, and thus maintains smaller coefficient values.

Chapter 3

Automated Redshift Estimation

Summary

3.1 Catalogues from Large Sky Surveys

In order to perform automated redshift estimation on galactic spectra it is necessary to first have a large data set for training purposes. Recent surveys, such as the ongoing and successful series of SDSS/BOSS surveys (York et al. 2000; Stoughton et al. 2002; Ahn et al. 2013), and future related surveys such as BigBOSS (Schlegel et al. 2011, BigBOSS was subsequently merged with the DESpec survey (Abdalla et al. 2012) resulting in the DESI spectroscopic survey, (Levi et al. 2013)) have and will provide vast amounts of both raw spectral data and photometric data.

The advances in modern technology, particularly over the last decade, have allowed data acquisition and storage to far outpace data analysis. Traditional analysis methods that may once have relied at least in part on human assessment are no longer viable methods for estimating the redshift from these vast data sets. It has become necessary to develop data analysis pipelines to perform certain tasks automatically. Included in these pipelines are algorithms to 'clean' the raw data (involving steps such as artefact removal, denoising, calibration, data rejection etc) and algorithms to process the data once it has been cleaned in order to obtain further information such as redshift estimates, morphological type, age etc.

3.1.1 SDSS

The Sloan Digital Sky Survey (SDSS) is a ground-based telescope large-sky survey collecting primarily three types of data: images, spectra and photometric magnitudes from galaxies and quasars. Each individual object may not have all three data-types available and some objects are present in the catalogue (eg: stars) but were not the primary focus of the survey. There have been three main phases (SDSS-I, 2000-2005; SDSS-II, 2005-2008; and SDSS-III, 2008-2014) of data collection at the dedicated telescope at the Apache Point Observatory, New Mexico, USA. SDSS-III contains various sub-surveys operating simultaneously, they are: BOSS (Baryon Oscillation Spectroscopic Survey), APOGEE (Apache Point Observatory Galactic Evolution Experiment), MARVELS (Multi-Object APO Radial Velocity Exoplanet Large-area Survey) and SEGUE-2 (Sloan Extension for Galactic Understanding and Exploration 2). APOGEE, MARVELS and SEGUE-2 (an extension of SEGUE-1) are concerned with stellar spectra and objects within the Milky Way and are not relevant for galactic redshift estimation; this work will use the terms BOSS and SDSS-III interchangeably.

The photometric filter-system (Fukugita et al. 1996) used by SDSS has a series of 5 filters (u, g, r, i, z) as shown previously in figures 1.6 and 1.7, that span a wavelength range of $\sim$ 3,000Å to $\sim$ 11,000Å. The accompanying spectrograph spans a wavelength range of 3,800Å-9,200Å (later extended to 3,600Å-10,400Å in the final (BOSS) phase). The main galaxy survey extends to a redshift of $z \sim 0.8$, and the quasar survey to $z \sim 2.2$. The resultant data is then given a (Petrosian) magnitude cut at r=17.77 for the main galaxy sample, and these spectra have a median SNR of 4.2 per pixel in the g-band. The full SDSS galaxy catalogue (currently) covers 14,555 square degrees and contains 1,880,584 galaxy spectra and 312,309 quasar spectra.

3.1.2 DESI

The DESI survey (Levi et al. 2013) is a planned sky survey due to commence operations in 2018 for a 10 year run, designed to collect spectra for a minimum of 18 million emission-line galaxies, 4 million luminous red galaxies and 3 million quasars in a redshift range of $0.5 < z < 3.5$. Thus making the overall catalogue around 10 times the size of the current SDSS catalogue. The goal of DESI is to determine redshifts at an error of $\sigma_z = 0.001(1+z)$ or better and estimate the large-scale structure of the Universe to within an accuracy of 1%, doing this primarily by means of Baryon Acoustic Oscillation (BAO) measurements.

BAOs are the imprints left upon large scale structures due to acoustic waves travelling in the primordial plasma, initiated by cosmological perturbations, that became 'frozen-out' when this plasma subsequently recombined into neutral atoms as the Universe expanded and cooled. BAOs are both a function of time elapsed since the Big Bang, and of redshift and can be used as a 'standard-ruler' for cosmological distance measurement. BAOs were first seen in the SDSS survey data (Eisenstein et al. 2005), and BOSS has shown the BAO technique to be both feasible to detect & measure, and robust to systematic error.

The spectrograph itself will observe in a range of 3,600 Å$< \lambda <$9,800 Å, and have a resolution between $R \gtrsim$1,500 in the blue increasing to $R \gtrsim$ 3,000 in the red and further still, $R \gtrsim$ 4,000, for the near infra-red. The choices of varying resolutions in each region are designed to specifically target the Ly-α forest, the 4,000 Å break and the [O II] doublet respectively.

3.2 Mock Catalogues

As previously introduced in section 1.3, galaxy spectra can be generated synthetically from initial conditions and criteria based on various parameters such as the initial mass function, star formation rates, infall times, etc. In general to create a galaxy catalogue requires a process termed 'Stellar Population Synthesis' (SPS) whereby model spectra are generated to resemble (sets of) individual galaxies. Spectra can also be generated from full hydrodynamical simulations of baryonic and dark matter, though this is often computationally demanding.

Further steps that are often employed are population mechanisms: whereby n-body simulated Dark Matter haloes are selectively populated with galaxies, such that the galaxy type conforms to the merger history and location of the halo, and the prior cosmology imposed in generating those haloes (for example, the Millennium Simulation Springel et al. 2005b); or with the use of a Luminosity Function (LF) that stipulates how galaxies must be placed in space in order to resemble a realistic population. (Merson et al. 2013).

A final step is required to make the catalogue resemble a realistic survey and that is the application of selection functions that precisely detail what proportions of galaxies can be observed, and where, and at what limiting magnitudes that will be no longer possible. The application of selection functions is wholly dependent upon the survey being modelled, and is extrinsic to the galaxy generation itself.

The three steps to generate a full galaxy catalogue can thus be summarised as: generation, whereby realistic galaxy spectra are constructed; localisation, whereby galaxies are distributed in space; and visualisation, whereby these galaxies are placed into the context of a particular survey .

In general, the spectral flux, F, of any galaxy at a time, t, and wavelength λ can be obtained from equation (3.1), as described in Charlot and Bruzual (1991); Fioc and Rocca-Volmerange (1997),

$$F_\lambda(t) = \int_0^t \int_\mu^{\mu+\delta\mu} \Gamma(t-\tau)\, f_\lambda(m,\tau)\, \xi(m)\, dm\, d\tau \;, \tag{3.1}$$

where $\xi(m)$ is the Initial Mass Function (IMF) defined in the interval $[\mu, \mu+\delta\mu]$ and expressed in terms the solar mass ($M_\odot$); $\Gamma(t-\tau)$ is the Star Formation Rate (SFR) at a time $t-\tau$, expressed in multiples of the solar mass per billion-year period ($M_\odot\, Gyr^{-1}$). Lastly, the term $f_\lambda(m,\tau)$ is the spectral flux of a star of initial mass, m, and age, τ, since ZAMS (the Zero Age Main Sequence). ZAMS is the point at which a stellar mass object enters the Main Sequence (MS) and becomes nuclear fusion dominated; this is considered as the 'birth' of the star, and any prior existence leading up to the formation of the star (as a gas cloud/protostar) is neglected. The $f_\lambda(m,\tau)$ term is taken to be zero once the star exceeds its lifetime and has become a post-fusion compact object (white dwarfs, neutron stars etc).

Evidently, to generate a representative mock catalogue, appropriate choices for the IMF, SFR and stellar spectra are needed. The majority of these choices reduce to either accurate observations of stars or accurate theoretical prediction. In cases where the parameters may be poorly understood, uncertainty will creep into the modelling, and may lead to unrealistic catalogues.

Not explicitly included in equation (3.1) are feedback processes that enhance metallicity, infall times for gas/dust reservoirs, and the influence that galactic winds have on both infall and metallicity considerations. Metallicities can be implicitly included in the stellar spectral flux, $f_\lambda(m,\tau)$, considerations.

A further and alternative approach is to bypass the complexities inherent to SPS by effectively 'ignoring' stars and to use pre-existing galactic spectra (from a real survey), and 'reverse-engineer' galactic properties from these, in order to construct randomised galactic spectra that share those

properties. This is achieved by random samplings of the measured statistical distributions associated with the properties describing the galaxy population. This has the distinct advantage that the errors and assumptions, both implicit and explicit, in equation (3.1) can to a large extent be avoided; furthermore the various populations of galaxy types, will by construction, be representative of a real galaxy survey, and the simulated properties will share the same distributions as those from which they were derived (by construction). There exists a strong disadvantage in that if the initial galaxy spectra on which such a catalogue is based are flawed, the resultant catalogue will inherit those flaws and not be representative of a real survey; additionally, real spectra are not as readily measured for higher redshifts (due to such objects being fainter).

There is nothing to preclude a semi-empirical approach where observed properties can be used to inform theoretical models, and indeed such approaches can be advantageous particularly when some aspects of the underlying physics in galaxy formation are not well known or are poorly constrained. An important motivation for utilising such methods is the ability of the model, once tuned to be in agreement with the observable low-redshift Universe, can be used to predict the high-redshift Universe that is yet to be observed. Detailed reviews of semi-empirical modelling can be found in Baugh (2006); Benson (2010).

This process can be used to generate realistic catalogues of spectral data for galaxies, furthermore it can be easily modified to obtain photometric data by convolving the resultant spectra through sets of filters.

3.2.1 Modelling

The Initial Mass Function

The initial mass function (IMF) is the function that determines how many stars can be constructed out of the initial collapsing gas cloud that is the progenitor of the prospective galaxy, and the mass-distribution that they can have; thus defining a probability distribution function from which a histogram of (initial) stellar masses can be populated. Salpeter (1955) derived the first IMF, considering empirical observations of the Milky Way and the associated luminosity function (similar to equation (1.20) in section 1.3) which stipulates the number of stars that can be found within a luminosity interval solely for main sequence stars. Since then, with more refined observations, different IMFs have been proposed, including extensions to sub-stellar objects (e.g.: Kroupa 2001a).

Salpeter derived the following IMF, $\xi(m)$, from observations of faint stars in the vicinity of the solar neighbourhood,

$$\xi(m)\,\delta m = \xi_0\, m^{-\alpha}\,\delta m = \xi_0 \Big(\frac{m}{M_\odot}\Big)^{-2.35} \Big(\frac{\delta m}{M_\odot}\Big)\,, \tag{3.2}$$

where $\xi(m)\,\delta m$ represents the number of stars of masses in the interval m to $m+\delta m$. Salpeter found this to be a power law, and determined the exponent, α to be -2.35. For convenience m is often taken to be in terms of solar mass, $M_\odot$. This empirical derivation was obtained for stellar masses between 0.4 and 10 solar masses only, with Salpeter noting that the behaviour changed markedly beyond 10 solar masses though he was unable to discern if this was a real effect, or due to the poor luminosity data available at the time.

Further work done by others have shown the IMF to not be universally true for all stellar mass ranges, and the initial considerations of Salpeter to have been correct only within a limited mass range. Kroupa (2001a,b) have shown that the IMF may be better described by a piecewise power law, since the original work by Salpeter neglected to properly account for complicating factors such

as multiple star systems (which if unresolved introduce biases into the mass-luminosity relation); furthermore the lack of a significant population of massive stars ($> 10M_{\odot}$) also diverges from the observations made by Salpeter, however this mass region remains uncertain due to difficulties in direct measurements (Scalo et al. 1998; Kroupa 2001a). The IMF derived by Kroupa maintains the same form, i.e.: ($\xi(m) \propto m^{-\alpha_i}$), but with differing exponents for different mass ranges,

$$\begin{cases} \alpha_0 = 0.3 \pm 0.7, & 0.01 \leq m < 0.08, \\ \alpha_1 = 1.3 \pm 0.5, & 0.08 \leq m < 0.50, \\ \alpha_2 = 2.3 \pm 0.3, & 0.50 \leq m < 1.00, \\ \alpha_3 = 2.3 \pm 0.7, & 1.00 \leq m\,, \end{cases}$$

where m is in terms of the solar mass.

The IMF is *assumed* to be universally true both in space and time (and thus in redshift), however, since the work done in its derivation rests solely on observations of local stars within the Milky Way (and its satellite Magellanic Clouds), this may not be the case (van Dokkum 2008; Conroy et al. 2009). Conroy et al. assess the impact of variations in, and the propagation of any errors in, the IMF to synthesised stellar populations and note that the resultant impact has the potential to be significant.

Conroy et al. (2009) disfavour the piecewise power law IMF of Kroupa since this would imply that passively evolving systems do not then exhibit continuous luminosity evolution (as would otherwise be expected), and thus generate unphysical luminosity jumps over time. Conroy et al. instead opt for the IMF proposed in van Dokkum (2008).

The Star Formation Rate

Star formation in a galaxy is dependent upon local conditions and the amount of 'cold' stellar material (gas and dust) available for making those stars present in the dark matter halo. Evidently in gas-rich galaxies (generally spiral and merging galaxies), the star formation rate has the potential to be very high, however in gas-poor galaxies (generally ellipticals) the star formation rate will be comparatively much lower. Feedback processes (e.g.: supernovae driven winds) can slow down the rate of star formation since they can partially reheat a region of the medium (Heckman et al. 1990); with hotter gas and dust inhibiting star formation since their propensity to clump is hampered by the speed of the individual particles at collision. Additionally feedback processes can enhance the rate of star formation by returning material to the interstellar medium (ISM) via mass-loss associated with stellar transition to post-fusion, compact objects.

The star formation rate (SFR) is dependent upon the reservoir of gas and dust available for forming stars, and the merger history/local environment of the galaxy - with tidal interactions and mergers boosting star formation for short periods (on the timescale of galaxies).

There exists a characteristic time-scale for galaxies called the *dynamical time*, τ_{dyn}, which represents the fastest possible speed at which the available reservoir can collapse under self-gravity alone (neglecting pressure and angular momentum considerations). Events taking place on timescales larger than the dynamical time will allow for (at least partial) accretion of the reservoir of material into a disk as the cold gas and dust collapse under self-gravity, producing the characteristic spiral galaxy morphology. Events on shorter time-scales will precede the formation of such a disk, and have the potential to prevent its formation. It is thought that significant star formation occurring before the dynamical time has elapsed could lead to the formation of primordial elliptical galaxies directly without progressing through (mergers of) spiral stages (Dressler 1980). This would occur since stars

would continue falling towards the centre of mass of the forming galaxy, and enter into (disordered) orbits about that centre of mass, virialising the system and expelling and/or warming some of the gas through dynamical interactions; star formation after a dynamical time has passed, allows structure to form, and draws out the gas and dust into an ordered planar disk configuration, and importantly does not disrupt the infall, allowing further accretion onto the disk.

Star formation can be quiescent – occurring at a steadily decreasing rate as gas accumulates and is slowly used up to form stars (typical of older/isolated spirals); or it can be starburst – where a sudden and intense period of star formation is triggered from an external factor such as a tidal interaction, merger or cannibalisation of a satellite dwarf galaxy (typical of irregular galaxies and younger spiral galaxies). Star formation in ellipticals is generally very low, and for most practical purposes can be considered to have ceased at some point in the past.

The exact mechanism for star formation – whilst undoubtably due to the collapse of cool, dense, molecular clouds – is a matter of some debate. There exist two principal competing models for the primary mode of this collapse: a top-down approach (Krumholz et al. 2005) where the molecular cloud fragments into multiple regions as it collapses, with stars then forming in each sub-region; and a bottom-up approach where overdensities in the molecular clouds form low mass objects that steadily acquire mass from the cloud, competing against one another to draw gas from the cloud.

The rate of star formation is generally determined from observations of emission line or continuum intensities that are thought to characterise star formation (an in depth review is given in Kennicutt 1998), with different morphological types possessing different indicators. However, the determination of the SFR is dependent upon an assumption of the underlying IMF which is not currently well constrained. The present-day star formation rate of the Milky Way, for example, is between 0.68 and $1.45 M_{\odot}\, yr^{-1}$ (Robitaille and Whitney 2010), with this being poorly constrained to a large extent due to differing choices of IMF as underlying assumptions when computing the SFR from observational data (Kennicutt 1998).

Associated to the SFR is the star formation history (SFH) which defines the SFR over the extended lifetime of the galaxy. Technically the SFR refers to an instantaneous time in the SFH, though often the term SFR is used in place of SFH. Typically, different morphological types have different star formation histories. Galaxy evolution synthesis algorithms therefore have to account for variations in SFH between morphological types. The Pégase and GALEV algorithms (respectively, Fioc and Rocca-Volmerange 1999; Kotulla et al. 2009), both take these considerations into account. Each algorithm possesses the following 3 types of SFH, corresponding to elliptical (Hubble E-type), lenticulars and spirals (S0,Sa-Sc), and 'disturbed' spirals (Sd)[1],

$$\Gamma(t) = \frac{M_{total}}{\alpha}\, e^{-t/\tau}\ , \tag{3.3}$$

where M_{tot} is the total matter content of the galaxy (assumed 'closed-box'), α is a normalisation parameter of the SFH model, and τ is the 'e-folding' time. Evidently the star formation rate for elliptical galaxies decreases exponentially with lifetime. This quickly leads to the expected low SFR at later times.

The spiral SFH is the following,

$$\Gamma(t) = \frac{M_{gas}(t)}{\beta}\ , \tag{3.4}$$

[1]The Sd class of galaxies was added to Hubble's original classification scheme by De Vaucouleurs, and they are described as spiral galaxies possessing diffuse and/or broken arms consisting of individual stellar clusters and nebulae, typically with a very faint central bulge. As such they can be considered a type of 'slightly' irregular galaxy.

where $M_{gas}(t)$ is the matter content of the infalling gas, β is a normalisation parameter.

Lastly the Sd galaxy type SFH,

$$\Gamma(t) = \Gamma_0 \qquad \text{if} \quad t \leq \tau_{\text{cut}} \ , \tag{3.5}$$

where Γ_0 represents a constant star formation rate for all times prior to some cut-off time τ_{cut}, for all times beyond this the SFR is 0, however a cut-off time is not a necessary component of this model.

For the Pégase algorithm a further instantaneous starburst model SFH is included, where $\Gamma(t) = k\,\delta(t)$ intended to represent the burst of star formation triggered in strongly interacting galaxies. The duration of star formation is, on the timescale of galaxies, essentially transient and so is represented with a delta function, where k is a normalisation constant.

Metallicity

As stated previously, the *metallicity* of a star or gas cloud is the proportion of elements (by mass) within it that are heavier than Helium; even though strictly some of these elements are non-metals, they are collectively termed 'metals' for convenience since all of them would have been formed through previous stellar fusion processes (strictly speaking some primordial lithium was produced directly from Big Bang Nucleosynthesis, it is treated as negligible for metallicity considerations). The metallicity, Z, of an object of mass, M, is defined as,

$$Z = 1 - \frac{m_H}{M} - \frac{m_{He}}{M} \ , \tag{3.6}$$

where m_H and m_{He} are the masses of hydrogen and helium respectively contained within the object.

Often metallicity cannot be estimated directly and instead must be derived from proxy indicators such as relating the presence of a single metal, often the abundance of iron (which has many, characteristic lines that are identifiable in the visible spectrum) to the abundance of hydrogen, and is a logarithmic quantity measured in units of 'dex'. A measured metal (iron) content of 0 dex, implies that the object has the same iron content as the Sun, and it is assumed that *total* metallicity follows suit, thus that object would have a metallicity of Z=0.02 (i.e.: 2% metals by mass, as the Sun is measured to have).

Metallicity is an important factor for galaxy formation for a number of reasons: the halo of gas and dust that is available for infall and thus star formation, must be cold in order to collapse, and having a greater proportion of metals allows for a faster cooling time; enhanced metallicity impacts on the luminosities of newly formed stars (and in turn the mass-luminosity relation); lastly extinction is proportional to metallicity (Baugh 2006).

Given the SFR, the IMF, and the infall time and infalling mass, it is possible to trace the metallicity contribution over time from stellar remnants to the ISM.

3.2.2 Catalogue Generation

LePHARE

The LePhare program (Arnouts et al. 1999; Ilbert et al. 2006), is a program dedicated to photometric redshift estimation, based on a simple χ^2 fitting method between the theoretical and observed photometric catalogue; it can additionally be used to generate mock spectro-photometric catalogues. This is achieved by the use of a set of template galactic SEDs (for example CWW-Kinney) and filters. Through the use of luminosity functions for each spectral type, apparent magnitudes can be evolved

through redshift in any given filter band. From these apparent magnitudes and the redshift, SEDs can be associated to the object, via a χ^2 minimisation, in order to derive its other apparent magnitudes in different filter bands. From this ideal realisation we associate to each object a luminosity profile which is amalgamated with observational conditions to derive different kinds of realistic apparent magnitudes.

PEGASE & PEGASE-HR

Pégase and Pégase HR are a collection of algorithms that are designed to generate synthetic catalogues of high resolution ($R \sim 10{,}000$) galaxy spectra in the optical range (4,000 to 6,800 Å), and/or lower resolution of a more extended far UV - near IR wavelength range ($R \simeq 200$; from 91 Å to 160 μm), (Fioc and Rocca-Volmerange 1997, 1999; Le Borgne et al. 2004). These algorithms are again written in Fortran. The precise modelling used is described at length in Fioc and Rocca-Volmerange (1997); Le Borgne et al. (2004), and the user-guide and optional parameters are described in Fioc and Rocca-Volmerange (1999).

The Pégase code utilises stellar libraries of Lejeune et al. (1997, 1998), and the HR code uses the higher resolution ÉLODIE stellar library (Prugniel and Soubiran 2001).

The modelling used in Pégase allows the user to specify the stellar libraries to be used as stellar templates for the stellar evolution synthesis, the IMF, the SFR, the initial metallicity of the infalling gas, the infalling time (closely related to the dynamical time τ_{dyn}, the fraction (by mass) of sub-stellar (non-luminous) objects. The optional inclusion of galactic winds, nebular emission, and extinction (and associated inclinations) are at the discretion of the user.

3.2.3 The COSMOS Mock Catalogue

The COSMOS Mock Catalogue (CMC) (Jouvel et al. 2009) is a mock catalogue generated via a semi-empirical approach, and is based on the real spectra of the COSMOS survey (Ilbert et al. 2009; Capak 2009), where observed properties are deconstructed in order to create simulated galaxies from those properties. In this simulated catalogue, galaxies with redshifts and SEDs are taken as the best photo-z fits to multiwavelength observations of the COSMOS field, with galaxy size distributions originating from the COSMOS HST imaging and emission line strengths using the relations derived by Kennicutt (1998).

Photo-z estimates are obtained from best-fit templates (based on the CWW-Kinney, BC03 templates, and following the Calzetti extinction law Coleman et al. 1980; Kinney et al. 1996; Bruzual and Charlot 2003a; Calzetti et al. 2000). The best fit template of each galaxy is integrated through the instrument filter transmission curves to produce simulated magnitudes. The resultant mock catalogue being limited to 26.2 (in their i band).

Emission lines (Lyα, [O II], Hβ, [O III] and Hα) are then simulated. Firstly this is done by estimating the star formation rate from the UV luminosity (an indicator of star formation) as described in Kennicutt (1998), and using this information to fix the calibration of the [O II] emission line flux. The Lyα, Hβ, [O III] and Hα are then calibrated from the [O II] emission line flux via their standard ratios (without extinction) of 2, 0.28, 0.36 and 1.77 respectively.

The CMC has the advantage over a purely synthetic catalogue in that it better preserves the relations between galaxy size and colour (and likely also the morphological type and redshift) due to the inheritance of properties from the real COSMOS catalogue. The CMC is, however, limited (by construction) to the range of magnitude space where the COSMOS imaging is complete, whereas a purely synthetic catalogue would be free to extend to fainter magnitudes.

3.3 Photo-z Codes

There are many estimation codes available in both the 'template matching' and 'empirical' categories, for an extensive, publicly available (but by no means complete) list, see Hildebrandt et al. (2010).

In general, to gauge the effectiveness of photo-z codes, they are compared with known values of redshift. This usually takes the form of running the photo-z code on a set of data that includes at least some z estimates already (from spectroscopy); or, alternatively, on a mock catalogue of data, generated from computer models. Sometimes mock catalogues may contain galactic populations that are less complex and less varied than the real Universe, and, as such, if a photo-z code fails to produce good redshift estimates from mock catalogues, it will also fail to produce good results for real data. The converse is not necessarily true; being able to replicate mock catalogue photometric redshifts well, does not guarantee that the code will be successful when applied to real data (indeed, the codes generally perform more poorly with real data due to the richer content of real data).

The accuracy of photo-z codes are usually characterised by 3 parameters; their scatter, bias and outlier rates. A perfect photo-z code would produce zero values for all these parameters. All the codes discussed by Hildebrandt et al. have these three parameters tested and compared with mock data, PHAT0; and with real data from the GOODS survey (PHAT1) — where PHAT stands for PHoto-z Accuracy Testing, and was conceived as an open 'competition' for the photo-z community. The objective of PHAT was as 'an international initiative to test and compare different methods of photo-z estimation'.

3.3.1 LePhare — a Template Matching Method

Template matching methods are the most common methods used in photo-z estimation codes. As previously mentioned they rely on a library of 'ready-made' SEDs which are predominantly based on *local* galaxies, often with some interpolation performed by modelling; however, templates are, on occasion, based solely on models. The appropriate selection of these SED templates is crucial for obtaining good results for the photo-z. The best results are obtained when the galaxies in the library of template SEDs are chosen to bear as close a resemblance as possible to the target galaxies in the photo-z survey.

In general Template Matching Methods (hereafter TMMs) do best when attempting to account as closely as possible for the Physics involved between the galaxy producing its light, and that same light reaching us; this will usually entail taking into account effects from galaxy evolution, total stellar output, internal extinction, IGM opacity, reddening, etc. Unfortunately not all the Physics is known, and what is known is not necessarily well modelled. This leads to problems with TMMs in estimating the redshifts of some candidates, in particular those at higher redshifts (and thus further into the history of the universe) where our knowledge of, for example, evolution of all the components of galaxies, begins to become less confident.

Commonly used templates include those by Coleman et al. (1980), consisting of (quiet) spirals and ellipticals; and Kinney et al. (1996), consisting of starbursts, which are often combined to form an SED template library consisting of just 10 templates, yet encompassing the majority of the general morphological classes of galaxies, figure 3.1.

MacDonald and Bernstein (2010), highlight the danger of biased estimates for photo-z if those estimates are of faint galaxies which have been based on templates of bright local galaxies. Indicating that relatively small shifts in metallicity and AGN flux can cause small, but significant, bias to the

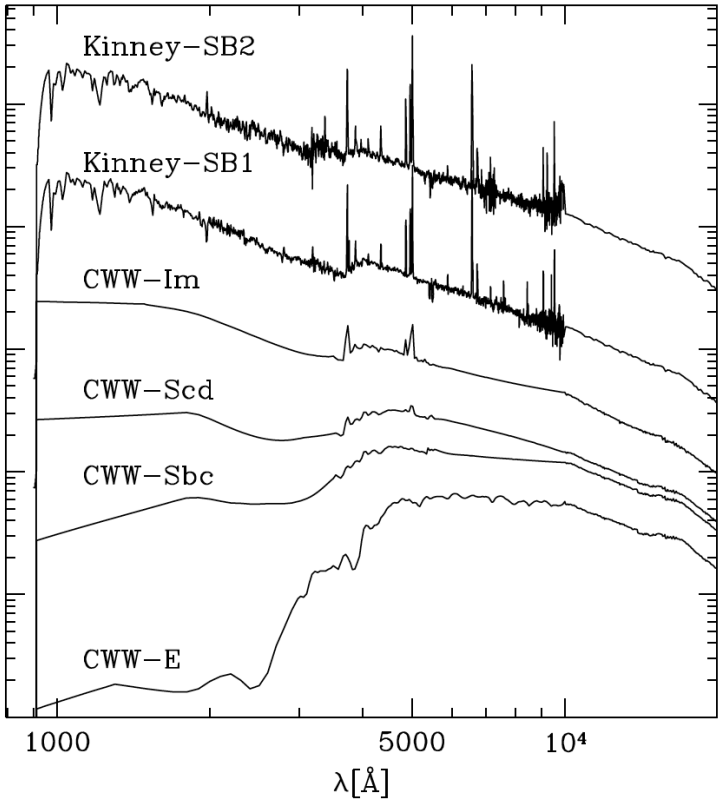


Figure 3.1: PHAT0 Template Set — The SEDs that were used in the PHAT0 simulation of Hildebrandt et al. were those of Coleman et al. and two of those of Kinney et al. The fluxes have been normalised to allow for direct comparison of SED profiles.

resulting photo-z estimate. This situation would only deteriorate when measuring at increasingly high redshifts, since the observed galaxies will become progressively much more faint. Together, this results in a pressing need to make certain that the template library is as representative of the target set of galaxies as possible; or alternatively, to reduce the dependency of codes on local templates.

From an imaged target galaxy, the measured magnitude in the several colour bands of the detector will be used by the code to produce an approximate SED (or even just part of one). The target SED would then be matched, by shifting all the templates in the SED library in redshift space (along the redshift axis), and selecting the one that was most similar to the target SED. This selection is most commonly done via a cross-correlation mechanism that minimises the χ^2 value between the template (model) and the target (constructed from measurement) SEDs. This manual shifting of local SEDs into a more distant redshift space ignores any temporal evolution of galaxies; as such, for better redshift estimates, models must be employed to approximate evolution.

The LePhare package[2], deveolped by Arnouts et al. (1999); Ilbert et al. (2006). The package consists of: model templates, with many tuneable parameters and calibrations, such as the priors used, as well as interchangeable SED libraries and filter sets; the χ^2 fitting algorithm itself, that attempts to match the observed to the theoretical catalogues; and finally it also contains a program to generate simulated catalogues with the intention that they are realistic multi-colour catalogues that also account for observational effects.

The simulated catalogues that can be generated by LePhare were used by Hildebrandt et al. (2010)

[2]Available for download at `http://www.cfht.hawaii.edu/~arnouts/LEPHARE/cfht_lephare/lephare.html` (includes an extensive user-guide).

(with some minor modifications) in the PHAT0 tests of the competing codes.

3.3.2 ANNz — an Empirical Method

In contrast to TMMs, empirical methods (EMs hereafter) care very little, if at all, about the complex Physics involved when estimating redshifts from photometric data. There are multiple types of EMs, including: artificial neural networks (hereafter ANNs) that mimic biological neural functions; decision tree based methods; and methods based around polynomial fitting.

Ultimately all EMs operate by attempting to deduce the redshift from input data. This deduction means that EMs bypass considerable difficulties associated with modelling the Physics involved and learn a relationship between photometric magnitudes and the resultant redshift. Often this entails that EMs are 'coached' in how to do this with the help of a training set of data, obtained from spectroscopy (and assumed correct), with which the EMs practice and derive the relationship between photometry and redshift.

This however can lead to those EMs that require a training set, inheriting the same difficulty of TMMs, namely that of the *representativity* of the known redshifts from spectroscopy. Again, the target galaxies in the survey will in general be more varied and dimmer than the ones in the training catalogue consisting of sample spec-z catalogues.

ANNz is an ANN that is used specifically for obtaining redshifts, it is written in C++ and is a 'black box' method. The user simply specifies the 'architecture' (see figure 3.2) and inputs the data; and at no point do they have to worry about how the program works. These features make ANNz a simple and easy to use program, whilst still obtaining powerful results. A full introduction[3] to ANNz can be found in Collister and Lahav (2004).

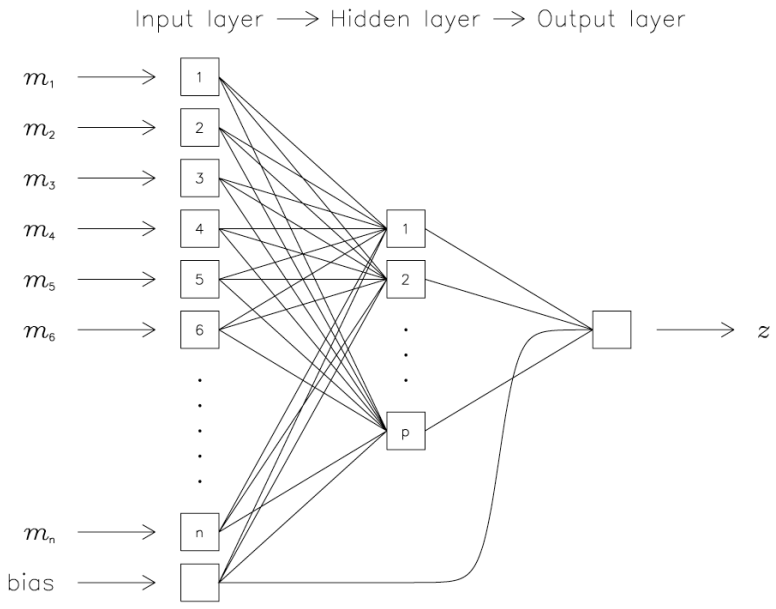


Figure 3.2: ANNz — The figure depicts the structure, or *architecture* of ANNz, which consists of layers. The inputs m_1 to m_n form the first layer, representing the various magnitudes in each filter. There exists one hidden layer of p nodes, and one output node, the redshift. The bias node allows for an additive constant in the calculations.

[3]Available for download at http://www.homepages.ucl.ac.uk/~ucapola/annz.html (includes a short user-guide).

The architecture is defined primarily by the complexity of the situation, with the number of nodes (the order) in the input layer being fixed by the problem (one input node is assigned to a magnitude in one filter band), and the output layer (just a single node) is the resultant redshift. The intervening layers – both the number of layers, and their order – are left to the discretion of the user. Each node in a particular layer is connected to every node in the preceding and subsequent layers.

The ANN architecture is essentially identified with the definition of a weighting function at each node, with the connectors representing the individual weightings. The user specifies and initial seeding (random) and the ANN subsequently optimises each weighting by minimising a cost function when applied to a training set. To ascertain whether or not the cost function is at a minimum, it is applied to a 'validation set', and if not, the process it iterated until a minimum is reached. This is to avoid over-fitting of the ANN to the training set, especially if it is small.

The overall architecture design can impact performance, as demonstrated by Firth et al. (2003). Increasing both the number and the order of intervening layers only provide minor improvements in efficiency. The increase of either of these parameters allows for a freer parameterisation and thus a closer fit to the data, however, a closer fit to the data is not always desirable since real data is fundamentally limited to an rms fit due to the presence of random noise, and over-fitting becomes a possibility. The reason behind the validation set is precisely to counteract such over-fitting. Additionally, a more expansive network will include a significantly larger number of weights to train, impacting the speed at which computation can proceed.

Generally, a network architecture needs to be chosen such that it is as simple as possible whilst still allowing sufficient 'room' to parameterise the data; an intervening layer of 1 or 2 nodes, for example, would leave insufficient room to parameterise a 5 node input.

The weighted networks once minimised, often correspond to *local* minima of the cost function, and thus will generally not represent a global minimum for the network architecture. To overcome this, multiple ANNs are recruited to work together in a 'committee' (Bishop 1995). A committee is essentially the mean of the individual outputs from a group of ANNs, and is usually a better overall estimate for the true value of the redshift than any one ANN would be on its own. The standard practice is to use at least 3 ANNs in a committee. This act of using ANNs in a committee allows the user to side-step any issues that may arise from variance in the network.

In principle, committees can be constructed from ANN members that have different architectures and that may have been used on different training and validation sets. The only requirement is that all the ANNs in the committee have to have been used on the same test data.

A comparison of LePhare and ANNz on the same SDSS dataset of 6,000 galaxies is shown in figure 3.3. The test was performed with default parameters in LePhare (with allowances for SDSS filters). ANNz, however, had the advantage of a training set of a further 5,000 galaxies, and a having a validation on 1,000 galaxies, furthermore 4 ANNs were placed in committee to obtain the results. ANNz clearly performs better in this example, however, the parameters for LePhare may have been sub-optimal when chosen as default. Important to note is the presence of catastrophic failures where the estimated redshifts are significantly different from the true redshift. Additionally, the errors of the estimates that lie closer to the diagonal line (where the spectral and photometric redshifts agree) are large – much larger than would be expected for a similar spectral redshift estimation since photometry, by construction, has less discernible information in the data from which to determine redshift accurately.

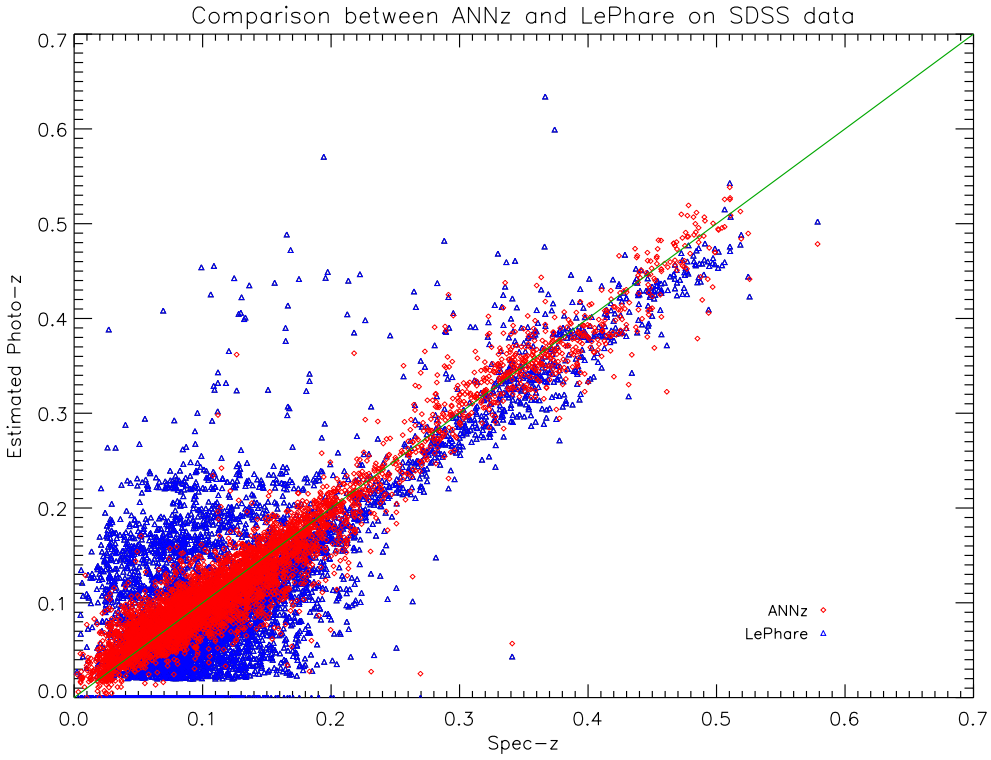


Figure 3.3: The results of the redshift estimation of a test SDSS photometric catalogue as obtained by LePhare (blue) and ANNz (red). The green line depicts the model solution where the photometric redshift is found to agree with the spectroscopic (assumed correct) redshift. The results from ANNz represent a committee of 4 neural networks, and LePhare was run with default parameters and SDSS filters. The results are clearly better in this example for ANNz.

3.4 Redshift Estimation by Cross-Correlation - PCAz

To estimate galaxy redshifts, we employ a cross-correlation method similar to that described by Glazebrook et al. (1998). This method involves a cross-correlation of test galaxy spectra at unknown redshift values with template spectra that are progressively shifted in redshift.

We assume that any test spectrum S'_λ may be represented as a linear combination of template spectra $T_{i\lambda}$,

$$S'_\lambda = \sum_i a_i \, T_{i\lambda} \, , \tag{3.7}$$

where each template spectrum is normalised according to,

$$\sum_\lambda T_\lambda^2 = 1 \, . \tag{3.8}$$

If we choose to bin our spectra on a logarithmic wavelength axis, redshifting becomes proportional to a translation,

$$\begin{aligned} \Delta &= \log(1+z) \\ &= \log\left(\lambda_{\,observed}\right) - \log\left(\lambda_{\,rest\ frame}\right) . \end{aligned} \tag{3.9}$$

The estimate of the goodness-of-fit between the template, now allowed to shift along the wavelength axis, and the test spectrum, at an unknown redshift, can be found by computing the minimum distance via a standard χ^2, where the previous coefficients a_i are now dependent upon redshift through Δ,

$$\chi^2(\Delta) = \sum_{\lambda} \frac{w_\lambda^2}{\sigma_\lambda^2} \left[S_\lambda - \sum_i a_i(\Delta)\, T_{i(\lambda+\Delta)} \right]^2 . \tag{3.10}$$

We can obtain the values of the expansion coefficients, a_i, by maximising equation (3.10) with respect to a_i. Following the prescription in Glazebrook et al. (1998), we take the weighting function, w_λ, and the normally distributed errors, σ_λ, to be wavelength independent and constant, which gives,

$$a_i(\Delta) = \frac{\sum_\lambda S_\lambda\, T_{i(\lambda+\Delta)}}{\sum_\lambda T^2_{i(\lambda+\Delta)}} . \tag{3.11}$$

The numerator in equation (3.11) is simply the cross-correlation of the galaxy spectrum with the i^{th} template spectrum. Substituting back into equation (3.10), we obtain,

$$\chi^2(\Delta) \propto \sum_{\lambda} \left[S^{\,2}_{\lambda} - \sum_i a_i^2(\Delta)\, T^2_{i(\lambda+\Delta)} \right] . \tag{3.12}$$

For a large test catalogue that includes a variety of morphological types, a large number of templates is needed to ensure the best match-up between template and test spectra. To use all of them in the cross-correlation would be excessively time-consuming. If it were possible to reduce the number of templates whilst still retaining most of the information content of these templates then we can render the method more practical.

Principal Component Analysis (PCA) is a simple tool that allows us to do just that: to reduce the dimensionality of this problem by extracting the most important features from our set of template spectra, the principal components. The general procedure involves the construction and subsequent diagonalisation of a correlation matrix to find eigenvectors and eigenvalues. It is possible to construct a correlation matrix either between the templates, or between the wavelength bins; the result is equivalent. We have chosen to do the correlation between the templates since in our case the number of templates is less than the number of wavelength bins, resulting in a smaller matrix that is simpler to manipulate,

$$C_{ij} = \sum_{\lambda} T_{i\lambda}\, T^T_{j\lambda} . \tag{3.13}$$

This correlation matrix is always real & square-symmetric, hence it can be diagonalised,

$$\mathbf{C} = \mathbf{R}\mathbf{\Lambda}\mathbf{R}^{\mathbf{T}} , \tag{3.14}$$

where $\mathbf{\Lambda}$ represents the matrix of ordered eigenvalues (largest to smallest) and $\mathbf{R}$, the matrix of correspondingly ordered eigenvectors. The eigentemplates, $\mathbf{E}$, can then be obtained,

$$E_{\,j\lambda} = \frac{\sum_i R^T_{ij}\, T_{i\lambda}}{\sqrt{\Lambda_j}} , \tag{3.15}$$

with the resulting eigentemplates having the same dimensions as the original dataset, and satisfying the orthonormality condition,

$$\sum_{\lambda} E_{\,i\lambda}\, E^T_{j\lambda} = \delta_{\,ij} . \tag{3.16}$$

The effect of PCA is that it re-orientates the dataset to lie along the orthogonal eigenvectors (axes) sorted by descending variance. It effectively creates an 'importance order', such that the eigenvector with the greatest variance (largest eigenvalue) will tend to correspond to the strongest signal features of the untransformed dataset, with subsequent eigenvectors representing less significant signal features, and the final eigenvectors, with the smallest variances, representing noise. For example, if H_α is a very prominent feature in most of the template spectra, it will be present in one or more of the first few eigentemplates.

With this in mind we can now re-cast equation (3.7) in terms of an approximation of the sum of the first N eigentemplates that are now allowed to be shifted along the wavelength axis,

$$S_\lambda \simeq \sum_{i=1}^{N} b_i(\Delta)\, E_{i(\lambda+\Delta)}\,, \tag{3.17}$$

where $b_i(\Delta)$ are new expansion coefficients for the new basis.

Using the orthogonality condition from equation (3.16), equations (3.11) and (3.12) then become,

$$b(\Delta) = \sum_{\lambda} S_\lambda\, E_{i(\lambda+\Delta)}\,, \tag{3.18}$$

$$\chi^2(\Delta) \propto \sum_{\lambda} S_\lambda^{\,2} - \sum_{i=1}^{N} b_i^2(\Delta)\,. \tag{3.19}$$

We then observe that the first term in equation (3.19) is a constant in the χ^2 function, and can be disregarded; therefore minimising the χ^2 function in equation (3.19) is equivalent to *maximising* the related function, $\widetilde{\chi}^{\,2}$, defined as,

$$\chi^2(\Delta) \sim \widetilde{\chi}^{\,2}(\Delta) = \sum_{i=1}^{N} b_i^2(\Delta)\,. \tag{3.20}$$

Hence, $\widetilde{\chi}^{\,2}(\Delta)$ is computed by first computing the cross-correlation of each of the N retained eigentemplates E_i with the galaxy spectrum (equation (3.18)), and then summing $b_i^2(\Delta)$ over these eigentemplates. We can further simplify the problem by noting that a convolution between two real signals transforms into a multiplication in Fourier space between the individual Fourier transforms of the galaxy and non-redshifted template spectra, with the advantage that Δ becomes a free parameter.

Hence we obtain,

$$b_i(\Delta) = \mathcal{F}^{-1}\big(\hat{S}_k\, \hat{E}_{ik}\big) = \frac{1}{M} \sum_{k=0}^{M-1} \hat{S}_k\, \hat{E}_{ik}\, e^{\frac{2\pi \mathbb{i} k \Delta}{M}}\,, \tag{3.21}$$

and,

$$\widetilde{\chi}^{\,2}(\Delta) = \sum_{i=1}^{N} \Big[\mathcal{F}^{-1}\big(\hat{S}_k\, \hat{E}_{ik}\big)\Big]^2\,, \tag{3.22}$$

where $\hat{S}_k$, $\hat{E}_{ik}$, represent the Discrete Fourier Transforms (DFTs) of S_λ, $E_{i\lambda}$; and $\mathbb{i}$, $\mathcal{F}^{-1}$ represent $\sqrt{-1}$ and the inverse DFT respectively.

Now that we have obtained equation (3.22) it is an easy task to extract the estimate for the redshift, z. The $\widetilde{\chi}^2$ function reaches a maximum when the shift of the templates along the log-wavelength axis corresponds to the true shift of the galaxy spectrum, so that the redshift is estimated to be where $\Delta = \Delta_{\widetilde{\chi}} \, (= \Delta|_{\widetilde{\chi}=\widetilde{\chi}_{max}})$, giving,

$$z_{est} = 10^{\,\delta_s \, \Delta_{\widetilde{\chi}}} - 1 \,, \tag{3.23}$$

where δ_s is the grid spacing on the $\log_{10}$-wavelength axis.

Note that, for this PCA/cross-correlation redshift estimation method, both the template and galaxy spectra must be free of continuum. This is important to ensure that it is the spectral features from each spectrum that are being matched to one another, rather than to any continuum features, which may lead to spurious correlations and hence confusion in the determination of the galaxy redshift.

3.5 Conclusion

In this chapter we briefly described the concept and purpose behind large-sky surveys, with a particular focus on the Sloan Digital Sky Survey (SDSS) and its forthcoming descendant DESI. We additionally described in detail, the process of the construction of realistic mock catalogues that emulate large-sky surveys (section 3.2), with a particular focus on the construction of synthetic spectra from the modelling of its base components: the initial mass function, star formation rate, metallicity, (accumulated) stellar spectra, and reddening from dust and gas.

Different publicly available algorithms for catalogue generation are investigated in section 3.2.2, as well as the publicly available catalogue, the COSMOS Mock Catalogue (CMC), which we adapt for use in subsequent algorithmic development and analysis (chapter 5).

Two main different, but complementary methods for measuring redshifts of galaxies in surveys were illustrated; spectroscopy and photometry. Two broad categories of obtaining photo-z estimates were explored; namely template matching methods (TMMs) and empirical methods (EMs), with the better of the codes (according to Hildebrandt et al. (2010)) in each of these two categories being explored further: LePhare and ANNz, and a comparison between these two was made.

A detailed description of the PCAz algorithm (section 3.4) as developed by Glazebrook et al. (1998), was given, with this later being incorporated into the Darth Fader (**d**enoised and **a**utomatic **r**edshifts **th**resholded with a **fa**lse **de**tection **r**ate) algorithm (Machado et al. 2013) in chapter 4. The importance of continuum-free spectra for this procedure is taken into account by the Darth Fader algorithm, where we use an entirely empirical method for subtracting the continuum that is based on the wavelet decomposition of the spectrum; this method will be described in detail in chapter 4.

Chapter 4

Darth Fader Algorithm

Summary

4.1 Darth Fader Algorithm

The Darth Fader (**d**enoised and **a**utomatic **r**edshifts **th**resholded with a **fa**lse **de**tection **r**ate) algorithm (Machado et al. 2013) is a computational algorithm to determine the redshift values of a catalogue of galaxy spectra in a blind and automated fashion. This is achieved by means of standard cross-correlation techniques and PCA coupled with an application of wavelet-based techniques, in order to remove continua and perform denoising, and self-select the the resultant redshifts into two classes: likely correct, and unreliable. The inputs to the programme are a test catalogue of galaxy spectra containing both spectral lines and continuum features (as well as noise, which if non-stationary, the error-curve detailing the noise response per pixel is also required), and a training catalogue in order to generate a series of eigentemplates. A schematic of the full algorithm is shown in figure 4.1.

In order to estimate the redshift of galaxies by cross-correlation, both the templates and the test galaxy spectra must be continuum-free. The reason for this is that emission/absorption lines are the best indicator of redshift, and retaining the continuum can lead to erroneous cross-correlation between line features and continuum features (such as breaks); in addition, spectra with different lines (and thus a different redshift) but superficially similar continua could result in a stronger correlation where

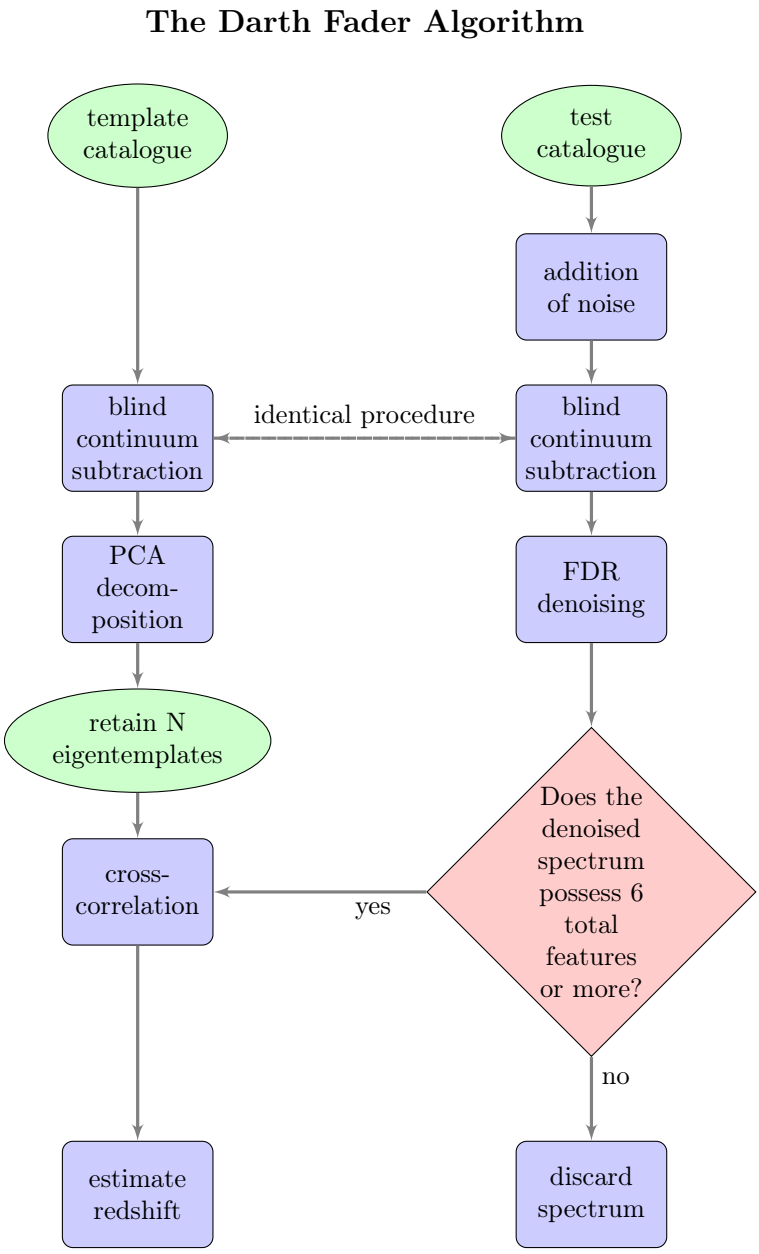


Figure 4.1: This figure illustrates the operation of the Darth Fader algorithm. The number of eigentemplates to be retained is at the discretion of the user, and may depend on the distribution of spectral types in the data. For example, a subset of eigentemplates can be selected such that their combined eigenvalues represent at least 99% of the total eigenvalue weight, and in general such a choice would result in significantly fewer eigentemplates than the number of spectra in the original template catalogue that was used to generate them.

The FDR denoising procedure denoises the positive and negative halves of the spectrum independently, with positivity and negativity constraints respectively. The requirement of six features or more in the denoised spectrum (this criterion is one that was empirically determined in chapter 5) effectively cleans the catalogue of galaxies likely to yield catastrophic failures in their redshift estimates. It should be noted that a 'no' decision represents the termination of that spectrum from the test catalogue and our analysis. An alternative feature-counting criterion could be used that is not focused entirely on the quantity of features, instead focussing on which features are present (and indeed we do this when applying the Darth Fader algorithm to data from the WiggleZ survey in chapter 6.

the continua best overlap, rather that picking out where the lines best overlap – with this being particularly true if the spectra possess a large continuum flux, and comparatively little line flux – and thus incorrectly determining the redshift. Furthermore, the associated implicit periodicity in the discrete Fourier transforms would suffer from an induced discontinuity due to a mismatch at each end of the spectrum should the continuum be retained (Tonry and Davis 1979; Glazebrook et al. 1998). A separate treatment with the continua can be used to determine an estimate or as a cross-check for the redshift, however, line features are much more constraining, and generally sufficient to determine the redshift alone.

Current methods for continuum subtraction rely on a handful of principal techniques: careful modelling of the physics of galaxies to estimate the continuum, a matching of continua between featureless galaxy spectra (typically sourced from elliptical galaxies or galactic bulges) and the spectra from which we wish to remove the continuum, or a polynomial fitting. (Koski and Osterbrock 1976; Costero and Osterbrock 1977; Panuzzo et al. 2007).

The first two of these methods have the disadvantage of requiring some knowledge of galaxy physics (which may not be precisely known), and being somewhat restricted to lower redshift/higher SNR galaxies. Careful modelling is computationally intensive and liable to result in failure if unusual galaxy types are found, or if the physics involved is not fully understood or modelled well enough. Continuum-matching methods require *a priori* knowledge of the galaxy type of a set of galaxies, and/or are reliant on the correspondence between similar looking continua (with one relatively featureless, in order to remove one from the other) which may not exist for all spectra. All matching methods have a problem at high noise levels where different but superficially similar spectra are mistakenly associated. Polynomial fitting is a further alternative that is limited to high signal-to-noise spectra, or spectra that have been denoised beforehand (as applied to the SDSS data release, Stoughton et al. (2002)).

By contrast the new method of continuum subtraction presented here is completely empirical, requires no knowledge of the physics or type of the galaxy involved and can be used even with very noisy spectra. This method relies on a **multiscale** modelling of the spectra, as described below (modelled on the DWTs as described in sections 2.3.2 and 2.3.3).

4.2 Spectra modelling

We can model the galaxy spectrum as a sum of three components – continuum, noise and spectral lines:

$$S = L + N + C\,, \tag{4.1}$$

where L contains the spectral line information, N is the noise and C is the continuum. L itself can also be decomposed into two parts, emission lines L_e, and absorption lines L_a, such that: $L = L_e + L_a$, and where, provided the continuum has been removed, $L_e > 0$ and $L_a < 0$.

The problem is then to estimate these components, L and C, from a unique data set. This is possible assuming that the features (L) are important only on small and intermediate scales, while the continuum (C) contains no small scale features at all, and is dominant on large scales only. Such an assumption is valid when one considers how a spectrum is generated from the separate constituents within a galaxy (as in figure 1.5), primarily continuum emission is aggregate thermal emission, whereas lines are the result of specific atomic transitions of varying ionisation states, dependent upon the respective abundances of such chemical species.

However, to effectively separate these components from one another, some important problems remain:

- Strong emission/absorption lines impact significantly at all frequencies, so a low pass filtering of the input is not sufficient to properly estimate the continuum, C.
- Both emission and absorption lines can be difficult to detect because of the presence of noise.

The method presented here is achieved in the following four steps; two for continuum estimation and two for absorption and emission line estimation.

1. First detect strong emission and absorption lines, which could be seen as outlier values for continuum emission.
2. Subtract from the data these specific strong features and estimate the continuum from this strong-line deficient spectrum. The estimated continuum can then be subtracted from the original spectrum.
3. Re-estimate the emission lines from the original data, now continuum-subtracted, via steps 1 and 2.
4. Re-estimate the absorption lines in a similar way.

4.3 Continuum removal

The proposed solution for continuum removal is a two-step procedure: firstly remove the strongest and therefore potentially problematic lines (either emission or absorption), and then estimate the continuum.

4.3.1 Strong line removal using the pyramidal median transform

In order to detect strong emission and absorption lines that could be seen as outliers for the continuum, we need a tool that is highly robust to these outliers. The choice of median filtering is generally the correct one for such a task, however this would require exact knowledge of the size of features for which to fix the median filtering window size. In our case, a better choice therefore is the multiscale median transform that was proposed for cosmic ray removal in infrared data (Starck et al. 1996a; Starck and Murtagh 2006). Furthermore its pyramidal nature allows us to significantly speed up computation time (Starck et al. 1996b). In this framework, strong features of different width can be efficiently analysed.

In a general multiscale transform[1], a spectrum of n bins, $S_\lambda = S\,[1, \ldots, n]$ can be decomposed into a coefficient set, $W = \{w_1, \ldots, w_J, c_J\}$, as a superposition of the form

$$S_\lambda = c_J(\lambda) + \sum_{j=1}^{J} w_j(\lambda)\ , \tag{4.2}$$

where c_J is a smoothed version of the original spectrum S_λ, and the w_j coefficients represent the details of S_λ at scale 2^{-j}; thus, the algorithm outputs $J+1$ sub-band arrays each of size n. The present indexing is such that $j = 1$ corresponds to the finest scale or highest frequencies.

[1]IDL routines to compute this and other wavelet transforms are included in the **iSAP** package available at: http://www.cosmostat.org/software.html

We use a similar multiscale transform in the following algorithm for strong line detection:

- Take the pyramidal median transform (PMT) of the input spectrum S (a median window of size 5 was used in all our experiments), we get a set of bands w_j and c_J at different scales j, $\mathcal{P}(S) = \{w_1, \ldots, w_J, c_J\}$. Where w_j corresponds to multiscale median coefficients, and can be interpreted as the information lost between two resolutions when downgrading is performed using median filtering, followed by a downsampling of factor 2. The c_J term corresponds to a very coarse resolution of the input signal. Further details can be found in Starck et al. (2010).
- For each band w_j, threshold all coefficients with an amplitude smaller than four times the noise level (this is a denoising procedure).
- Set the coarse resolution, c_J, to zero (this is approximate to removing the continuum).
- Reconstruct the denoised spectrum S_1.

S_1 represents a crude estimation of the lines L, mainly because the noise behaviour in the pyramidal decomposition cannot be as well calculated as in a linear transform such as the Fourier or wavelet transforms. However the process is much more robust than with a linear transform since strong lines with small width will not contaminate the largest scales as would be the case for instance with wavelets, resulting in artefacts termed 'ringing' when removing the continuum. Ringing produces features either side of a strong line (after removing the coarsest scale), such that positive features (emission lines) will have accompanying negative features (absorption lines) either side of the positive one; similarly for a strongly negative feature. These induced artefacts possess counterparts in the largest scale (such that they disappear upon summation/reconstruction), and content from highest scales can be viewed as having 'leaked' into the coarsest scale. In the interests of feature counting, ringing artefacts can be highly problematic since they do not represent true lines, and can produce significant feature contamination in the final spectrum.

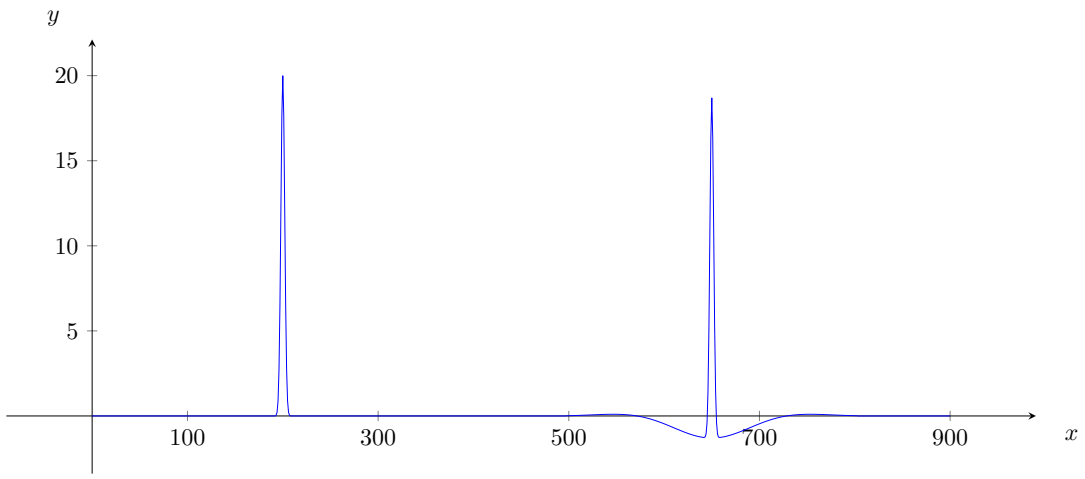


Figure 4.2: A simple example of ringing - The Gaussian on the left is what we would like to obtain from continuum subtraction, however the right hand Gaussian shows what is obtained – artefacts either side of the feature of interest termed ringing. These artefacts can vary in both width and height depending on the progenitor feature, with narrower and taller progenitor features producing more pronounced ringing.

A simple example of ringing is shown in figure 4.2, a Gaussian function (representing an idealised emission line) with magnitude of 20 added to a flat continuum of unit magnitude is to have its continuum removed: the left hand Gaussian is the idealised solution we wish to obtain with continuum subtraction (the original Gaussian function without the additional unit continuum), however the right hand Gaussian, with the ringed artefacts either side, is what we obtain with a naïve continuum subtraction. The nature of these ringed artefacts vary, with both their height/depth and width being affected by the progenitor feature (which can include both emission and absorption lines, and strong break-type features in the continuum). Such artefacts, particularly if they become large and narrow, strongly resemble the features we are attempting to distinguish but remain indistinguishable from them and as such have the potential to pose a significant problem in feature detection.

Since S_1 contains strong lines and the signal, $S_2(= S - S_1)$ will be free of any strong features, and a more robust continuum estimation can easily be derived from it, since the impact of ringing will now be significantly reduced.

4.3.2 Continuum extraction

The second step is therefore to estimate the continuum from S_2. The largest scale of S_2 should contain the continuum information (see first term in equation (4.2)), whilst the noise and undetected lines are expected to be dominant on smaller scales. So now the coarsest scale in a wavelet decomposition, or any low pass filtering, would give us a good estimation for the continuum. The advantage of wavelets for this task, as compared to a simple low pass filtering performed in Fourier space for example, is to allow a greater flexibility for handling the border (i.e.: minimising edge effects), and there being no requirement to assume periodicity of the signal. We do this using the starlet transform, also called the isotropic undecimated wavelet transform (IUWT, equation (4.2), see also section 2.3.2).

This transformation is simply a new representation of the original signal, which can be recovered through a simple summation. For a detailed description of the starlet transform see Starck et al. (2010), which has further been shown to be well-adapted to astronomical data where, to a good approximation, objects are commonly isotropic (Starck and Murtagh 1994, 2006).

We therefore estimate the continuum by first taking the wavelet transform of S_2, i.e.: $W^{[S_2]} = \{w_1^{[S_2]}, \ldots, w_J^{[S_2]}, c_J^{[S_2]}\}$, and then retaining only the largest scale: $c_J^{[S_2]} = C$. This continuum can now be subtracted from the original noisy spectrum to yield a noisy, but now continuum-free, spectrum.

4.3.3 Example

We show in figure 4.3 an example noiseless spectrum from our simulated catalogue (as described in chapter 5), containing both strong line features and continuum. In figures 4.4a and 4.4b, we show this spectrum with added noise with SNR (r-band, SDSS) values of 5 & 1; galaxy surveys typically make SNR cuts with a lower bound at 5-10 on the continuum. Over-plotted in the latter two figures is the continuum as estimated by the method described above, for an SNR of 5, the continuum fit can be seen to be quite good. At lower SNR, however, the fit is quite poor due to the inability to determine a rough estimate of the lines, S_1 in the first denoising step (for this particular noise realisation). The more intense noise effectively conceals the continuum, however, this continuum estimate is still well within the noise limits, and the correct order of magnitude. For reference we include the SNR on the H_α line in each case, with these being 8.9, and 1.7 respectively for the r-band SNRs 5 and 1; SNR values for specific lines are generally not correlated to SNR values in filter bands, however.

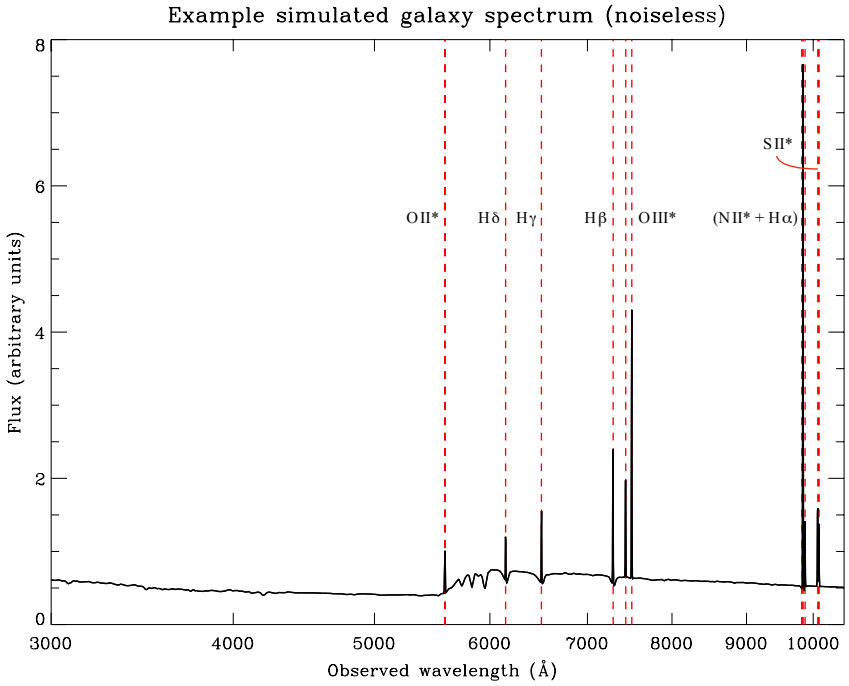
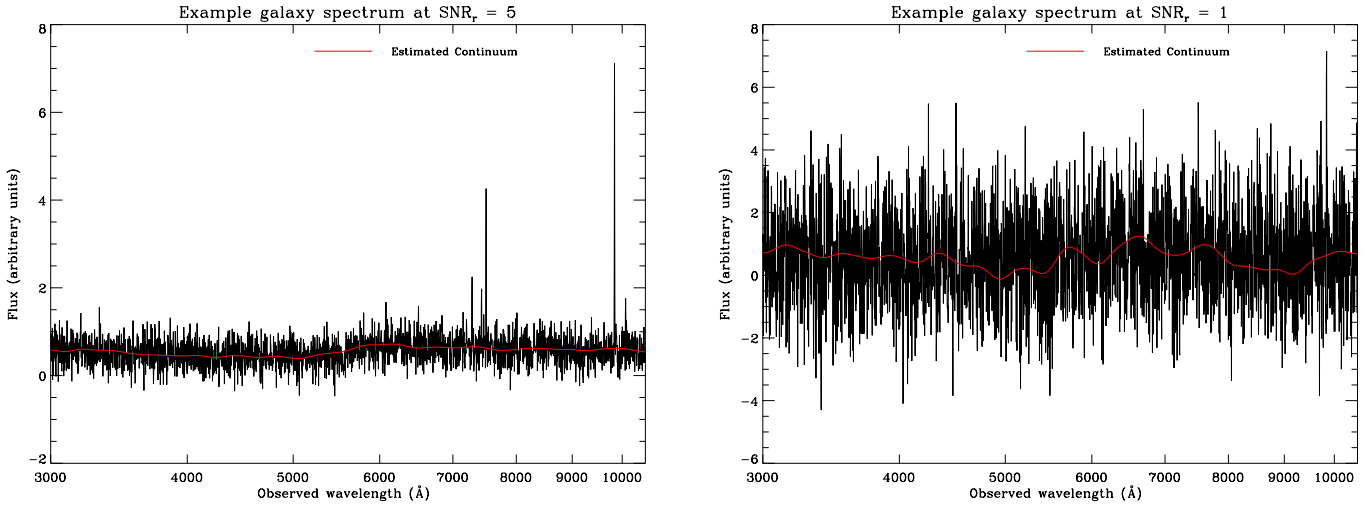


Figure 4.3: This figure shows an example spectrum from the test catalogue ($z = 1.4992$), prior to the addition of noise. The main emission lines are labeled; with an asterisk denoting a doublet feature. The [O II] doublet is fully blended in this spectrum.



(a) (b)

Figure 4.4: This figure shows a same spectrum as that in figure 4.3 but with manually added white-Gaussian noise at a signal-to-noise level in the r-band of 5 in figure 4.4a, and of 1 in figure 4.4b. The red lines indicate the empirically-determined continua in each case. Many of the prominent lines are easily visible by eye at the higher SNR of 5, whereas at the lower SNR of 1 most of the lines are obscured, with only H_α being sufficiently prominent so as to be detectable. The continuum estimate is good at the SNR of 5, and comparatively poor, but of the correct order of magnitude, at the lower SNR due to the dominating influence of noise. As an indication of *line-SNR*, we quote the values for the SNR on H_α for these particular spectra as 8.9 and 1.7 respectively for figures 4.4a and 4.4b.

4.4 Absorption/emission line estimation using sparsity

The wavelet representation of a signal is useful because it enables one to extract features at a range of different scales. In many cases, a wavelet basis is seen to yield a sparse representation of an astrophysical signal (such as a spectrum or a galaxy image), and this sparsity property can be used for many signal processing applications, such as denoising, deconvolution, and inpainting to recover missing data (e.g.: Fadili and Starck 2009; Starck et al. 2010). This work focuses on one such application: that of denoising spectra using wavelet filtering.

The basic idea underlying sparse wavelet denoising is that the signal we are aiming to recover is sparsely represented in our chosen wavelet dictionary. This means that the signal is completely represented by a small number of coefficients in wavelet space.[2] This sparsity property means that if we are able to identify the important coefficients, it is straightforward to extract the signal from the noise (since Gaussian noise is *not* sparse in the wavelet domain). Wavelet denoising has been previously applied successfully to both stellar (Fligge and Solanki 1997; Lutz et al. 2008) and galactic spectra (Stoughton et al. 2002, for the SDSS early data release).

There are various methods to do this; one simple method would be $K\sigma$ denoising, where a threshold is set relative to an estimate of the noise, and all pixels with an SNR less than K are set to zero (see section 2.1.1). This method, however, cannot give us any guarantees on feature contamination, since, with each pixel having an associated probability of being incorrectly identified, the number of contaminants can become unacceptably large with increasing numbers of pixels in a spectrum (much larger than the initial threshold might suggest). A more sophisticated method involves the use of a False Discovery Rate (FDR) threshold, which allows us to control contamination from false positive lines arising from noise features. This method was introduced in section 2.1.3 and is described further below.

4.4.1 Sparse Wavelet Modelling of Spectra

As the continuum C is now estimated, we can use the continuum free spectrum $S_c = S - C$ to tackle the remaining problem, which is to properly estimate the lines, assuming $S_c = L + N$ and $L = L_e + L_a$. We can exploit the wavelet framework in order to decompose it into two components: line features, and noise. This is done using a modified version of a denoising algorithm based on the Hybrid Steepest Descent (HSD) minimisation algorithm developed by Yamada (2001).

Hence, we can reconstruct L by solving the following optimisation problem,

$$\min_{L} \left\| \hat{\mathcal{W}} L \right\|_1, \qquad \text{s.t.} \qquad S \in \mathcal{C}, \tag{4.3}$$

where $\hat{\mathcal{W}}$ is the wavelet transform operator, $\left\| . \right\|_1$ is the ℓ_1 norm, which promotes sparsity in the wavelet domain, and $\mathcal{C}$ is a convex set of constraints, the most important of which is a linear data fidelity constraint,

$$\left| w_j^{[S_c]}(\lambda) - w_j^{[L]}(\lambda) \right| \leq \varepsilon_j, \ \forall \ (j, \lambda) \in \mathcal{M} \ . \tag{4.4}$$

Here $w_j^{[S_c]}$ and $w_j^{[L]}$ are respectively the wavelet coefficients of S_c and L, and ε_j is an arbitrarily small parameter that controls how closely the solution L matches the input data. The constraint set $\mathcal{C}$ may also include further constraints, such as positivity for emission line-only spectra, etc. Note that the large scale coefficients c_J are not considered in this minimisation, as we do not expect the largest

[2] This is analogous to the representation of periodic signals in Fourier space, where they may be represented by only a few frequencies in this domain.

scales to contain any useful information since the continuum has already been subtracted. $\mathcal{M}$ is the *multiresolution support* (Starck et al. 1995), which is determined by the set of detected significant coefficients at each scale j, and wavelength λ, as,

$$\mathcal{M} := \{(j, \lambda) \mid \text{if } w_j(\lambda) \text{ is declared significant}\} \ . \tag{4.5}$$

The multiresolution support is obtained from the noisy data S_c by computing the forward transform coefficients, $W = \{w_1, \ldots, w_J, c_J\}$, and recording the coordinates of the coefficients w_j with an absolute value larger than a detection level threshold τ_j, often chosen as $\tau_j = K\sigma_{j,\lambda}$, where K is specified by the user (typically between 3 and 5) and $\sigma_{j,\lambda}$ is the noise standard deviation at scale j and at wavelength λ. When the noise is white and Gaussian, we have $\sigma_{j,\lambda} = \sigma_j$, and σ_j can be derived directly from the noise standard deviation in the input data. When the noise is Gaussian, but not stationary, which is generally the case for spectral data, we can often get the noise standard deviation per pixel σ_λ from the calibration of the instrument used to make the observation, and $\sigma_{j,\lambda}$ can be easily derived from σ_λ (Starck and Murtagh 2006).

An interesting and more efficient alternative to this standard $K\sigma$ detection approach is the procedure to control the False Detection Rate (FDR). The FDR method (Benjamini and Hochberg 1995)[3] allows us to control the average fraction of false detections made over the total number of detections. It also offers an effective way to select an adaptive threshold, α.

In the most general context, we wish to identify which pixels of our galaxy spectrum contain (predominantly) signal, and are therefore 'active', and those which contain noise and are therefore 'inactive'. The measured flux in each pixel, however, may be attributed to either signal or noise, with each having an associated probability distribution. When deciding between these two competing hypotheses, the null hypothesis is that the pixel contains only noise, and the alternative hypothesis is that the pixel contains signal.

The FDR is given by the ratio:

$$FDR = \frac{V_f}{V_a} \ , \tag{4.6}$$

where V_f is the number of pixels that are truly inactive (i.e.: are part of the background/noise) but are declared to be active (falsely considered to be signal), and V_a is the total number of pixels declared active.

The procedure controlling the FDR specifies a fractional threshold, α, between 0 and 1 and ensures that, *on average*, the FDR is no bigger than α:

$$\langle FDR \rangle \leq \frac{V_i}{V_T}.\alpha \leq \alpha \ . \tag{4.7}$$

The unknown factor V_i/V_T is the proportion of truly inactive pixels; where V_i is the number of inactive pixels, and V_T the total number of pixels.

FDR works because it ranks all pixels, based on a function of their p-values, and then excludes all failing pixels with a single statistical test based on the threshold, α (equation (2.3)). This means that for a fixed choice of alpha, a spectrum of an arbitrary length will not incur a compounding of errors based on the number of pixels it possesses, as would be the case for K-σ thresholding. Hence we can maintain confidence that, on average, the features we identify as signal are likely to indeed be signal $(1 - \alpha) \times 100\%$ of the time, which is in direct contrast to the K-σ method.

[3] Benjamini and Hochberg term FDR as false *discovery* rate in their paper; it is exactly analogous to what we term false detection rate in this work.

A complete description of the FDR method can be found in Starck and Murtagh (2006) and, from an astrophysical perspective, in Miller et al. (2001). FDR has been shown to outperform standard methods for source detection (Hopkins et al. 2002), and Pires et al. (2006) have shown that FDR is very efficient for detecting significant wavelet coefficients for denoising of weak lensing convergence maps. In this work, the FDR method is applied at each wavelet scale, and hence gives a detection threshold τ_j per wavelet scale.

The minimisation in equation (4.3) can be achieved using a version of the Hybrid Steepest Descent (HSD) algorithm adapted to non-smooth functionals, full details of which can be found in Starck et al. (2010).

In practice, we separately estimate emission & absorption lines by applying the algorithm twice, first with a positivity constraint to get L_e, then with a negativity constraint to estimate L_a. This approach was found to be more efficient than a single denoising without constraint, allowing for better constraining of ringing around detected lines. Our final estimate of L is then obtained by $L = L_e + L_a$.

4.5 Example

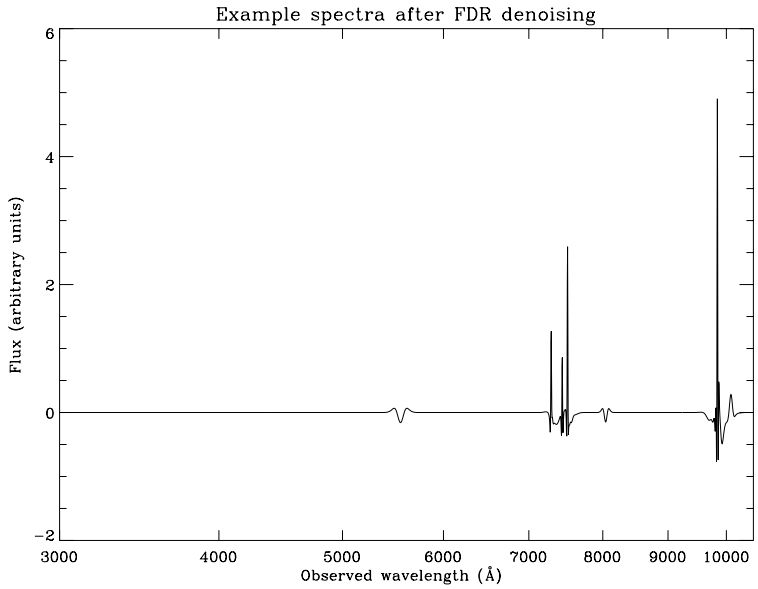
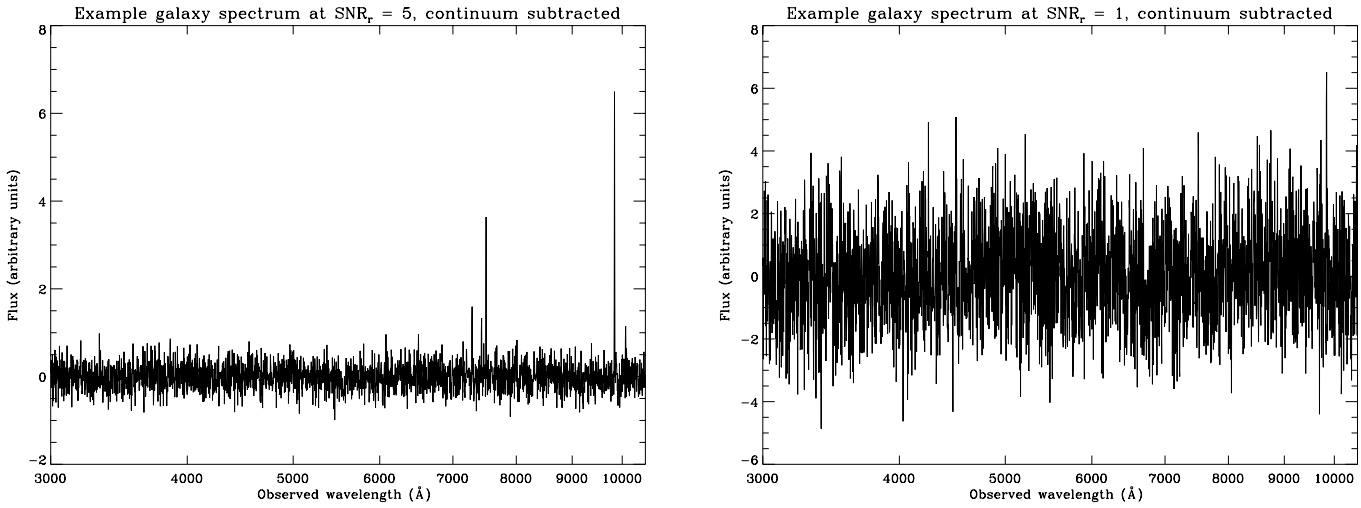


Figure 4.5: This figure is the result of an unrestricted denoising of the spectrum in figure 4.4a with an FDR threshold corresponding to an allowed rate of false detections of $\alpha = 4.55\%$. The [O III] doublet, H_α and H_β are all cleanly identified. There are small features corresponding to [O II] and [S II], and a spurious feature at just over 8,000 Å. The FDR denoising of figure 4.4b fails to detect any features for this particular spectrum, noise-realisation and choice of FDR threshold, and thus returns a null spectrum (not shown).

As an example, shown in figure 4.5 is the first attempt at the reconstruction of the lines, L, from figure 4.4a, using an FDR threshold of $\alpha = 4.55\%$. Here, the positive and negative halves of the spectrum have *not* received independent treatment and the denoising is unrestricted since it is for the purpose of continuum-subtraction. It is the FDR denoising with the aim of feature-counting (figure 4.7) that requires a separate treatment of positive & negative halves of the spectrum; this procedure is not required for continuum subtraction. The denoising of figure 4.4b fails to detect any features, and thus returns a null spectrum.



(a) (b)

Figure 4.6: Figures 4.6a and 4.6b are the spectra as shown in figures 4.4a and 4.4b with their empirically determined continua subtracted.

The secondary step – denoising to determine the number of features – is shown in figure 4.7, for the continuum subtracted spectrum shown in figure 4.6a. Note how the denoising artefacts (ringing) in figure 4.5 are no longer present, and as such are not mis-counted as features. In the noisier example (figure 4.6b) the denoising once again fails to detect any features and returns a null spectrum (for this particular noise realisation and FDR threshold of 4.55% allowed false detections), and this would lead to the spectrum being discarded from our catalogue as unlikely to yield an accurate redshift estimate.

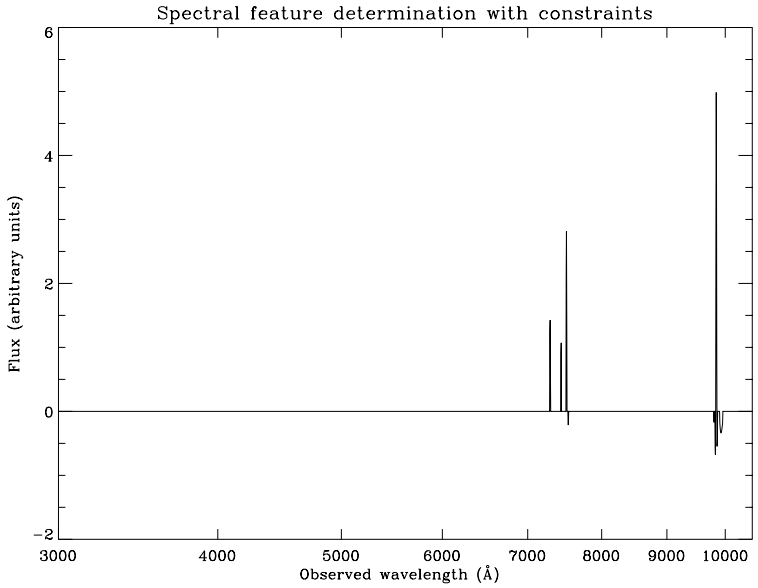


Figure 4.7: This figure shows the result of denoising the positive and negative sections (shown together) of the spectrum shown in figure 4.6a with positivity and negativity constraints respectively. Note the reduced ringing, which leads to a more representative result with respect to the number of true features. Once again the FDR denoising of our noisier example (figure 4.6b) yields a null spectrum (not shown), and would thus result in the discarding of this spectrum from the redshift analysis.

4.6 Redshift Estimation

For our method, we follow the PCA procedure as described in section 3.4 on a set of noise-free template spectra to obtain eigentemplates, of which we keep only the first N principal components such that they comprise at least 99% of the total eigenvalue weight, which in our case resulted in the retention of 20 eigentemplates (comprising 99.93%). We continuum-subtract our spectra as described in section 4.4.1. Since the white-Gaussian noise on a spectrum will in principle be uncorrelated with the eigentemplates, we choose to use the noisy galaxy spectra in the cross-correlation. This ensures that we preserve all the line information in the spectrum, rather than discarding some of the signal through denoising, and hence we are not probing potential systematics of the denoising simultaneously with the redshift estimation.

However, when dealing with the pixel-dependent noise, it is the denoised spectra that must be used in the cross-correlation, since very noisy pixels in proximity to less noisy pixels will produce features that strongly resemble lines, and would thus be highly correlated with features in the eigentemplates (independent of the redshift of the spectra involved) if not denoised. For example, an error-curve peaking strongly at 7,600 Å, may frequently produce features at this wavelength in the noisy spectra that strongly resemble lines. The effect of this false feature is to bias the cross-correlations such that large features in the templates (for example H_α) consistently match up to this false line, independent of the true redshift of the spectrum, resulting in redshift estimates that consistently favour an incorrect redshift. In this example many spectra would be biased to have an estimated redshift of 0.158, irrespective of their true redshift values. As such we must use the denoised versions of the spectra with non-stationary noise for the cross-correlations. However, the redshift estimation will thus incur any potential systematics of the denoising procedure itself, this is explored further in section 5.3.

Clearly, at low SNR, some of these cross-correlations will produce inaccurate results due to many features becoming lost in the noise. Higher SNR is not a guarantee of a successful redshift estimate; it is possible that line confusion, a lack of features, or poor representation in the basis of eigentemplates will result in a catastrophic failure.

A simple, but effective, criterion for the selection of galaxy spectra that will be likely to yield an accurate redshift estimate can be developed by considering the number of significant line features (either absorption or emission) present in the spectrum. For a spectrum containing many prominent features, it should be very easy to determine the redshift via cross-correlation with a representative set of eigentemplates. In cases where only one prominent feature is present, for example, we expect that the cross-correlation function will have several local maxima, each occurring when the location of the line feature in the test spectrum aligns with any of the features present in the (shifted) template spectrum, there would be no reason however to expect the global maximum to equate to the true redshift. A similar effect would be expected for a spectrum with many – but not particularly prominent – features, obscured by noise. In such cases, it will not generally be possible to obtain a correct redshift unless we have more information about that feature or the spectrum (for example identifying the continuum shape/using photometric data which would help in identifying the colour of the galaxy; redder being indicative - but not definitively - of higher redshift), and/or we make an assumption that the most prominent feature is a specific standard line (for example, H_α). There is also the possibility that the dominant feature in the spectrum is a noise feature (this could be the case for multiple features if the spectrum is very noisy), in which case it will be impossible to estimate the redshift correctly. As such, feature deficient spectra ought to be classified as 'unreliable' for cross-correlation redshift determination.

With an increasing number of detected features, and a high degree of certainty that they are not

the result of noise contamination, it should become clear that the redshift estimate obtained for such a test spectrum becomes progressively more reliable. Evidently, to estimate redshift, the greater the abundance of true features, the more likely it will become that accurate redshift determination is possible. This is the reasoning behind the usage of FDR denoising, since it, unlike K-σ thresholding for example, can reliably tell us on average, what percentage of pixels will be correctly classified.

A question arises as to quite how many features are sufficient to distinguish reliable redshifts from those which are not reliable and we wish to discard. Through empirical tests, we have chosen 6 features in total as the criterion by which we decide the reliability of the redshift estimate of a test spectrum in our catalogue (as described in chapter 5). This criterion however, will be dependent upon the specificities of the galaxy catalogue: resolution, wavelength span, EM regions (optical, IR etc). Generally however, it should be more than 2 since there is no guarantee that one of these lines will not be spurious, and there can exist line confusion between pairs of lines that are separated by a similar gap on a log-λ axis.

With this in mind, we use the denoising procedure described in section 4.4.1 on the continuum-subtracted spectrum and identify the number of features present in the denoised spectrum via a simple feature-counting algorithm[4]. We then partition the catalogue into two classes: a cleaned catalogue comprised of noisy spectra for which denoising presents 6 or more features, where we keep the redshift determination as likely to be accurate; and a discarded catalogue with spectra only possessing 5 features or fewer upon denoising, where the redshift estimates are deemed to be unreliable.

Features are considered to be 'peaks' anywhere where the derivative of the spectrum changes from positive to negative (maxima), but only in the spectrum's positive domain; this means that, for example, a Gaussian-like function with two maxima (e.g.: a partially blended line-doublet), would count as *two* features. Employing this method alone would ignore absorption features; to overcome this we denoise and feature-count the positive and negative halves of the spectrum separately, independently detecting both emission and absorption features.

At low SNR there is a trade-off between relaxing the FDR threshold to pick up more features – or indeed any features – and imposing a stricter threshold to prevent the detection of spurious lines. Recall that the FDR parameter constrains the average ratio of false detections to total detections. Therefore, for an FDR parameter of $\alpha = 0.05$, for example, we allow on average one false feature for every 20 features detected; i.e.: an average ratio of 19 true features to 1 false feature. In very noisy data, it might not be possible to achieve this statistical accuracy, and therefore no features will be identified by the algorithm.

It follows that even if 6 features can be obtained from the denoising of the spectrum, some of them may still be spurious, and this could lead to an erroneous redshift estimate from cross-correlation (particularly if the spurious line is dominant, with this strongly biasing the cross-correlation) and false-positive contamination of our retained data. However, as noted, a maximum for this false line contamination is set by the FDR threshold, α. In addition, the spectra that possess fewer than 6 features may provide redshift estimates that would otherwise be reliable; the criterion chosen leads them to be discarded. There exists this necessary trade-off between the fraction of catastrophic failures in the resulting redshift catalogue and the amount of data that is discarded.

[4]Algorithm adapted from 'peaks.pro', available from: http://astro.berkeley.edu/~johnjohn/idlprocs/peaks.pro

4.7 Conclusion

In this chapter we have combined the methods detailed in chapter 2 and chapter 3 into a unified algorithm for estimating redshift in the low SNR regime. The algorithm schematic was presented in section 4.1. Additionally we presented algorithmic methods for empirical continuum extraction (section 4.3), and empirical strong-line isolation and FDR denoising in order to isolate likely true features (section 4.4).

We presented worked examples in sections 4.3.3 and 4.5 using a single example spectrum taken from the CMC, with different (white-Gaussian) noise levels added.

Lastly in section 4.6 the considerations for selecting a spectrum to which we can assign a likely correct redshift were explored. The reasoning behind FDR denoising was detailed, with the importance of detected features being likely true being highlighted. The rationale for a feature-counting criterion was also explained, with an empirical determination of 6 features being found to be optimal for flagging spectra as likely correct for the analysis performed in chapter 5 (however this is revised for real data from the WiggleZ survey in chapter 6).

Chapter 5

Simulations & Results on Simulations

Summary

5.1 Sub-catalogue Generation from CMC Master Catalogue

The redshift estimation method described in section 3.4 requires two separate spectral catalogues: a set of galaxy spectra with noise and covering a range of redshifts that we aim to estimate (the test catalogue), and a set of noise-free, zero-redshift template spectra. We use the CMC set of simulations as provided by Jouvel et al. (2009); Zoubian and Kneib (2013), which are based on the observed COSMOS SEDs of Ilbert et al. (2009); Capak (2009), as described in section 3.2.3, to generate both the test catalogue and the template spectra. We then rebin a randomly selected sub-sample of the CMC master catalogue onto an evenly spaced $\log_{10}\,\lambda$ wavelength grid, spanning the range between 3,000 Å to 10,500 Å for the test catalogue, with a wider range for the template spectra of 3,000 Å to 20,900 Å.

This choice of binning, and a similar pixelisation scheme as in Smee et al. (2012), gives a constant resolution across the spectrum of $R\left(\propto \lambda/\Delta\lambda\right) \sim 850$ for all the catalogues, and a grid spacing of $\delta_s = 2.17 \times 10^{-4}\ \log_{10}$ Å; as compared to SDSS where the resolution and grid spacing are $R \sim 1{,}845$, and $\delta_s = 1.0 \times 10^{-4}\ \log_{10}$ Å respectively. The simulated spectra from the CMC mock catalogue are therefore less 'detailed' than real SDSS spectra (about half as detailed).

The CMC mock catalogue establishes its spectra on a linear wavelength grid, hence to convert the CMC mock catalogue to a log-wavelength format, the data must be rebinned. This is obtained by a least-squares quadratic interpolation from the linear wavelength grid onto the logarithmic one – with the resolution chosen so as not to excessively oversample the higher wavelength region.

The template set consisted of 277 simulated spectra, blueshifted to be at zero redshift; the test catalogue of 2,860 simulated spectra with redshifts in the range $0.005 < z < 1.7$. The template catalogue is chosen such that it is approximately 10% of the size of the test catalogue, and disjoint to the test catalogue (there are no spectra that are in both sets).

The resolution of the spectra will influence number of features we can detect: if the resolution

is poor, there is more uncertainty in the precise wavelength of the spectral lines, making it easier to confuse lines because they are slightly smeared out. Another, more important concern, is the potential blending of doublets or other lines in close proximity in wavelength, such as [N II] with H_α. These localised groupings of lines provide powerful constraints on the galaxy redshift since the wavelength gap between the two lines in a doublet or close pair is often sufficient to conclusively isolate which emission lines are present, and hence deduce the redshift. Poorer resolution will often result in blending of such features, limiting the number of detectable features as well as broadening the possible location of the feature. Poor resolution therefore impacts both the number of features through blending, and the detected locations of the features in wavelength due to coarser pixelisation of the data.

A trade-off with maintaining the feature-counting criterion can be obtained with poorer resolution spectra by increasing the wavelength range they cover, whereby – provided features exist in this extended range – more features can be found to counteract the loss of feature detections and precision as a consequence of the poorer resolution. It is for this reason that our simulated spectra cover a larger wavelength range than SDSS currently does (however DESI (Levi et al. 2013), itself a merger of the BigBOSS (Schlegel et al. 2011) & DESpec (Abdalla et al. 2012) spectroscopic surveys, is expected to have a similar wavelength range). In reality this trade-off is a minor consideration, however, since the practicalities of instrument design and the 'redshift desert' (see section 1.2.2) are the limiting factors for the wavelength range of spectra in real surveys, and since typical instruments have a much better resolution than our simulations.

The template spectra are required to have a larger wavelength span than the test spectra since they must be able to accommodate a large enough overlap to identify the correct (global) cross-correlation minima with these test spectra at all redshifts under consideration. A restricted wavelength span on the set of template spectra will necessarily reduce the maximum redshift at which you can cross-correlate. This frequently will result in the cross-correlation picking a local minimum that exists in the overlap – since the global minimum lies outside this overlap – often resulting in a confusion between one principal feature for another, thus systematically mis-identifying the redshift. A further necessary requirement is that the test catalogue and the template set be binned onto identical (intersecting) grids for the cross-correlation method described previously to work, though they need not possess identical starting or terminating wavelengths.

Wavelength-independent (white) Gaussian noise was then added to the test catalogue to generate several catalogues in which all the galaxies within the test catalogue have the same SNR. We define our SNR in the same manner as in the SDSS pipeline (Bolton et al. 2012, for BOSS/SDSS III)[1], relative to the median SNR in the SDSS r-band filter (Fukugita et al. 1996):

$$\mathrm{SNR}_r = median\left[\frac{flux}{\sigma}\right]_{5{,}600\,\text{Å}}^{6{,}760\,\text{Å}}, \tag{5.1}$$

where σ is the standard deviation of our added white-Gaussian noise, and the subscript r denotes that the median is calculated between the bounds of the SDSS r-band filter (5,600 Å to 6,760 Å).

We choose this particular definition of SNR so as to be consistent with a realistic survey such as SDSS. The specific choice of SDSS band for which to base the SNR definition on does not affect the method presented in this work. The motivation for choosing the r-band over any of the other SDSS bands is fully explained in Strauss et al. (2002), and is reliant upon an expected higher photometric SNR in this band. Being a redder band, K-corrections (a correction to a band's magnitude due to different, partial, sections of the spectrum being present in the band at different redshifts) are smaller in this band; the dominant component to this band will be the population of elder, redder, stars;

[1]The idlspec2d pipeline software package is available at: http://www.sdss3.org/dr8/software/products.php

and the reddening effects are less significant. Furthermore, the r-band has a less pronounced and less variable sky background than the redder i-band.

Whilst this definition of SNR is a good proxy for SNR on the continuum, and as such allows a simple comparison between different spectra, it should be cautioned that it is not necessarily a good proxy for the SNR on specific features.

Real spectrographs have a sensitivity that varies – sometimes quite strongly – with wavelength or per pixel, primarily as a result of the sky brightness and instrumental efficiency. We simulate a realistic error-curve that spans the typical optical survey wavelength range, and in figure 5.1 we present a 1σ error-curve per pixel that we use to mimic the expected noise behaviour of a realistic instrument. This is similar to what could be expected for an existing survey such as SDSS, or the forthcoming DESI spectroscopic survey, as well as other projects involving multi-object spectrographs[2]. The SNR definition for the spectra possessing pixel-dependent noise such as this remains unchanged from the one in equation (5.1).

The Darth Fader algorithm can then use this information present in the error-curve to assist in the denoising step, better accounting for the complex noise properties of the observed spectrum, thus enhancing the ability to discriminate between true features and those arising due to noise.

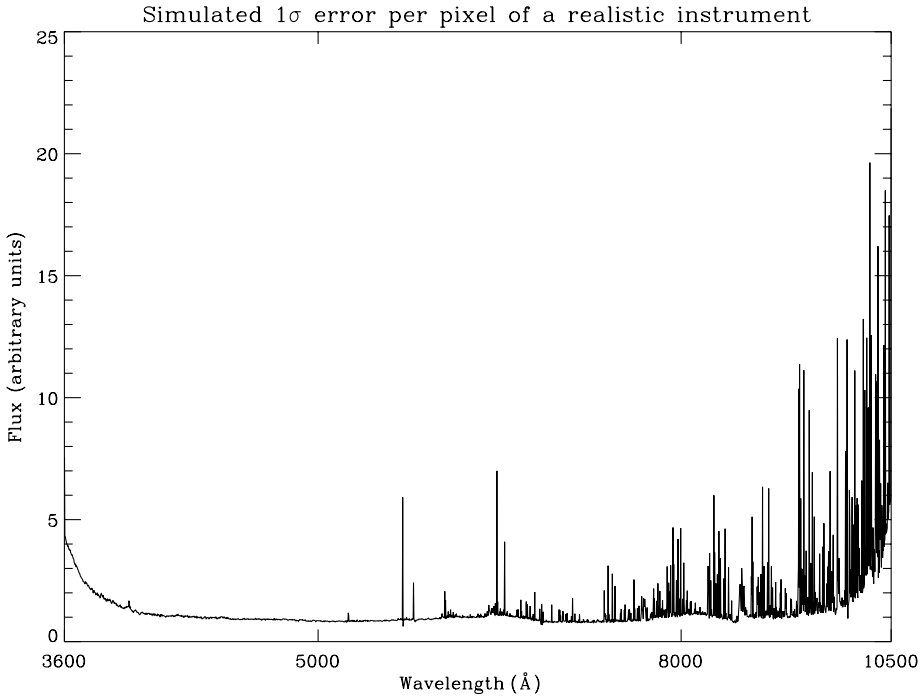


Figure 5.1: A realistic error-curve, where the resolution and binning are the same as for our mock catalogue, but with the wavelength range being slightly shorter, in order to be more proximal to the wavelength range of a realistic instrument. Gaussian noise is added to each pixel in our simulated data, with a standard deviation given by the value of the error-curve at that same pixel.

In addition to the single noise-level catalogues, a further mixed SNR catalogue was generated by adding pixel-dependent Gaussian noise such that the spectra in the catalogue had a uniform distribution in SNR in the range $1.0 < \text{SNR} < 20$ as described above. For convenience we will collectively call all the single SNR valued catalogues with flat white Gaussian noise catalogues 'SNWG' (for single noise, white Gaussian), and when specifying a particular SNR value, for example 2, this

[2]These surveys are, however, expected to be at much higher resolution than our simulations.

will be referred to as SNWG-2; the mixed SNR catalogue with pixel-dependent noise will be called 'VNPD' (varying noise, pixel dependent), where the noise in each pixel is Gaussian, but neighbouring pixels will generally not have the same level of noise for a constant signal (as in figure 5.1).

In order to test our algorithm, we investigate the effect of the choice of the False Detection Rate (FDR) parameter (as described in section 2.1.3) on the rate of catastrophic failures and the fraction of retained data in the cleaned SNWG in order to understand the interaction between noise content and the value chosen for the FDR threshold. We use the simulated data described previously, and apply the Darth Fader algorithm over multiple FDR thresholds, keeping the signal-to-noise constant; and again over catalogues with different SNRs, keeping the FDR threshold constant. Lastly we apply Darth Fader to the VNPD, utilising a range of values for the FDR threshold. The inclusion of the VNPD allows us to investigate a catalogue that is a step closer to a real survey, (though such a uniform distribution in signal to noise would not be expected in a real survey) in particular the effect of including pixel-dependent (Gaussian) noise.

We define the retention $\mathcal{R}$; catastrophic failure rates before cleaning, $\mathcal{F}$, and after cleaning, $\mathcal{F}_c$; and capture rate $\mathcal{C}$ of the sample to be:

$$\mathcal{R} = \frac{\mathcal{T}_c}{\mathcal{T}} \times 100\% \,, \tag{5.2}$$

$$\mathcal{F}_{(c)} = \left(1 - \frac{\mathcal{U}_{(c)}}{\mathcal{T}_{(c)}}\right) \times 100\% \,, \tag{5.3}$$

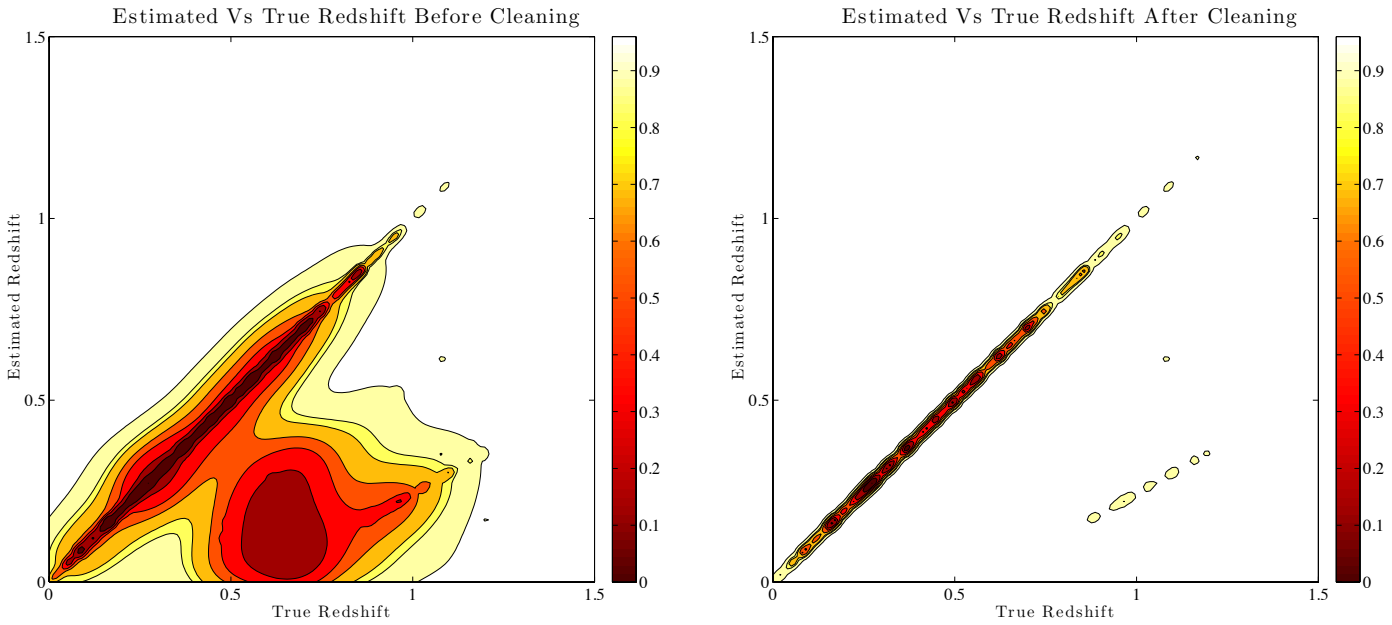
$$\mathcal{C} = \frac{\mathcal{U}_c}{\mathcal{U}} \times 100\% \,, \tag{5.4}$$

where $\mathcal{T}$ and $\mathcal{T}_c$ respectively denote the total number of galaxies in the sample (before cleaning) and the retained number of galaxies in the sample after cleaning (the number that satisfy the feature-counting criterion). Similarly, $\mathcal{U}$ and $\mathcal{U}_c$, respectively denote the number of successful redshift estimates in the sample before and after cleaning. In equation (5.3), the bracketed subscripts denote the option of calculating the catastrophic failure rate before cleaning (ignoring the subscripts) or the catastrophic failure rate after cleaning (inserting the subscript c everywhere shown). The number of successes after cleaning, $\mathcal{U}_c$, cannot be greater than $\mathcal{U}$, hence the capture rate represents the proportion of correct estimates available before cleaning that are retained post-cleaning.

5.2 Results for SNWG

We present the result of cleaning the SNWG-2 using an FDR threshold of $\alpha = 4.55\%$ in figure 5.2. The two panels compare the distribution of redshift estimates before and after cleaning of the catalogue using the feature-counting criterion. A clear improvement is seen when cleaning is applied: the fraction of catastrophic failures in the catalogue is reduced from 34.5% before cleaning to 5.1% after cleaning. In addition, we have retained 76.2% of the galaxies which yielded a correct redshift estimate before cleaning (the capture rate), with the retained catalogue comprising 52.6% of the total number of galaxies in the test catalogue. Clearly outliers still exist after cleaning of the catalogue (off-diagonal elements), where the redshift estimation has failed, but these are very few and the result remains reliable (with a 94.9% certainty).

Prior to cleaning there clearly exist two components to the redshift estimates, a strong square diagonal (the x=y line where the redshift estimates are likely correct) and a cloud of misidentifications (with a small, non-square, diagonal component) where the estimated redshifts are generally underes-



(a) (b)

Figure 5.2: A contour plot to show the effect on redshift estimation before and after cleaning of the SNWG-2, cleaned with an FDR threshold of 4.55%. Contours indicate the fraction of points enclosed within them. figure 5.2a depicts the results before cleaning, and figure 5.2b, after. Just under two thirds of all the estimated redshifts lie on the diagonal (and are thus correct) before cleaning being applied. The result has a high certainty, with 94.9% of the estimates being correct. The capture rate for this catalogue and at this FDR threshold is 76.2%.

timates of the true redshift. It is important to note that failures at this point are due to the standard cross-correlation, with non-square but diagonal components often being indicative of line confusion (for example between H_α and [O III]).

This represents a snapshot of how the Darth Fader algorithm works: we can *blindly* isolate a correct subset of galaxies, ensuring a good coverage of the correct data available in the pre-cleaned catalogue, and we can – by choosing an appropriate FDR threshold – guarantee that the resultant catalogue contains a very low catastrophic failure rate. Though not implemented in Darth Fader, the data rejected could be subject to further analysis, using additional information (e.g. photometry) and alternative methodology to determine the redshifts of the galaxies.

Figure 5.3 shows the catastrophic failure rate of Darth Fader before and after catalogue cleaning for a fixed FDR threshold of $\alpha = 4.55\%$, as a function of median SNR in the r-band. At high SNR ($\sim$ 20), the standard cross-correlation method yields a low catastrophic failure rate, and cleaning yields little improvement. At an SNR of 10, however, the standard cross-correlation method experiences a progressive increase in the catastrophic failure rate, approaching 50% at an SNR of 1; in contrast our method can maintain a low catastrophic failure rate ($\lesssim 5\%$) for SNR levels of ≥ 1.

An important point to note is that the catastrophic failure rate before cleaning ($\mathcal{F}$) represents a theoretical minimum amount of data that *must* be discarded with a perfect catalogue cleaning method (where $\mathcal{F}_c$ and $\mathcal{C}$ would be 0 and 100% respectively); thus the theoretical maximum retention is given by 100% - $\mathcal{F}$. In practice the amount discarded is usually greater (since we inevitably discard galaxies that would otherwise yield correct redshifts), but it can also be less than this if a more relaxed

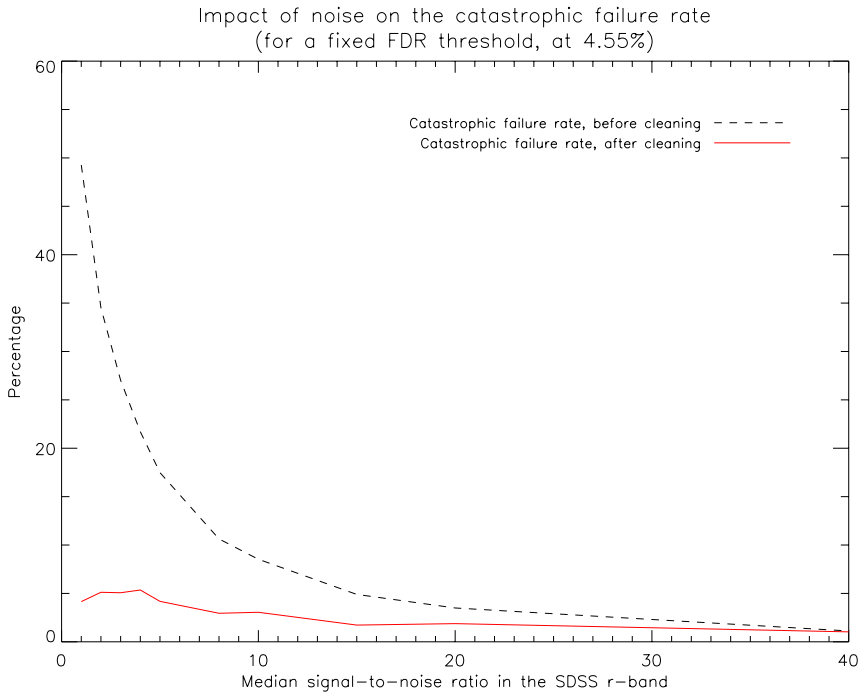


Figure 5.3: This figure illustrates how Darth Fader improves the catastrophic failure rate of the redshift estimates of the SNWG for a fixed FDR threshold of 4.55% allowed false detections. Note the marked improvement in the SNR range 1.0 - 10.0 where catastrophic failure rates are reduced by up to 40%. For this choice of α, the catastrophic failure rate is always found to be $\lesssim 5\%$ after cleaning, for SNR values ≥ 1. Our catastrophic failure rate after cleaning at an SNR of 1 is similar to the rate for an SNR value of 15 without cleaning. The catastrophic failure rate before cleaning (dashed line) represents the theoretical minimum amount of data that must be discarded for perfect catalogue cleaning.

threshold is used, necessarily incurring false positive contamination in the retained data set.

Using an FDR threshold of 4.55% allowed false detections, the SNWG-2 has a catastrophic failure rate of 34.5% before cleaning, thus our maximum expected retention in a perfect catalogue cleaning should only be 65.5%, with our actual retention at that FDR threshold being 52.6%. It should therefore be unsurprising that at the lower end of signal-to-noise the expected retention values for a cleaned catalogue are (necessarily) low; however this can still represent a large proportion of the *correct* data available. The recovery of this data still represents a net gain when compared to a total loss of this data, particularly when this recovered data can be known to be highly accurate.

The impact of cleaning is reduced at higher SNR due to denoising not revealing significantly more useful diagnostic information at lower noise levels; the number of features present in the noisy spectrum will more frequently already meet the selection criterion before cleaning, and thus cleaning the catalogue removes fewer spectra. To ensure a similarly low catastrophic failure rate in low SNR data would require a stricter FDR threshold to be used, and therefore would result in more data being discarded.

To demonstrate the effect of the choice of FDR threshold on the catastrophic failure rate after cleaning, the retention and the capture rate in the very low SNR regime, we test Darth Fader on two catalogues, SNWG-2 and SNWG-1, additionally we test on the VNPD for comparison (see section 5.3 for further details).

Figure 5.4 clearly demonstrates the tradeoff that exists between the catastrophic failure rate after

cleaning and the capture rate. Relaxing the threshold (i.e.: increasing α) improves both the retention and the capture rate by detecting more features in the spectra, more of which are likely to be false features rather than true ones, and thereby increasing the number of spectra accepted under the feature-counting criterion, but at a cost to the catastrophic failure rate since more erroneous spectra will also be accepted. A more conservative approach leads the FDR denoising to remove more real features, with the guarantee that very few of the remaining features will be false detections. This leads to a general decrease in both the retention and the capture rate since fewer spectra will exhibit the required number of features after denoising, with the benefit of this being a decrease in the catastrophic failure rate.

Notice also in figure 5.4 that beyond a certain point the catastrophic failure rate saturates for the SNWG (and shows little improvement for the VNPD), and stricter FDR thresholding (resulting in a smaller fraction of false detections) does not yield significant reductions in the rate of catastrophic failures; indeed this only serves to penalise both retention and the capture rate.

The results of the VNPD, being uniformly mixed in SNR and pixel-dependent, represent a step toward a more realistic view of what a real galaxy survey could look like. A real survey would not, however, have such a uniform distribution of SNR values, and would be skewed toward a greater population at lower SNR, with the actual distribution in signal-to-noise being specific to the type of survey and the instrument used.

5.3 Denoising for the VNPD

Pixel-dependent noise is potentially a more complex challenge to overcome in denoising, and in order to test the validity of the denoising steps for the VNPD, we show below further considerations for the denoising of spectra with pixel-dependent noise.

Figure 5.5 shows a continuum-subtracted spectrum from our catalogue, truncated to match the wavelength range of the error-curve (and hence our simulated instrument), before and after the addition of the wavelength-dependent noise of figure 5.1. We also plot the spectrum after denoising (again with an FDR threshold of 4.55% allowed false detections) with the Darth Fader algorithm when supplied with the error-curve in figure 5.1. This spectrum has a median SNR of 5 in the r-band at this particular redshift, ($z = 1.5$). However, for the same noise level, this SNR would vary between ~ 3 and 5, with redshift, as a result of differing continuum flux being located within the boundaries of the r-band.

To test the effectiveness and robustness of the denoising, we use the same test spectrum as in figure 5.5, and apply $10,000$ random (wrap-around) shifts in order to randomise the location of the principal features. For each shifted spectrum, pixel-dependent Gaussian noise is added as before, and at the same level. We then perform a denoising on each spectrum, and compute the residual with the input noisy spectrum. The RMS residual gives an estimate of the noise with its statistical distribution – if the denoising has been effective – matching the input error-curve. In figure 5.6, we show the ratio of the noise standard deviation as a function of wavelength, computed from the $10,000$ residuals, to the input error-curve for both the pixel-dependent noise in figure 5.1, at FDR parameters of $\alpha = 4.55\%$ & $\alpha = 0.27\%$, and for the flat white-Gaussian noise ($\alpha = 4.55\%$).

The randomised shifting of the spectrum allows us to determine the effectiveness of the denoising independently of the locations of the true features, and removes any cumulative biasing arising from features being undetected after denoising, or denoising artefacts such as ringing. We do however expect

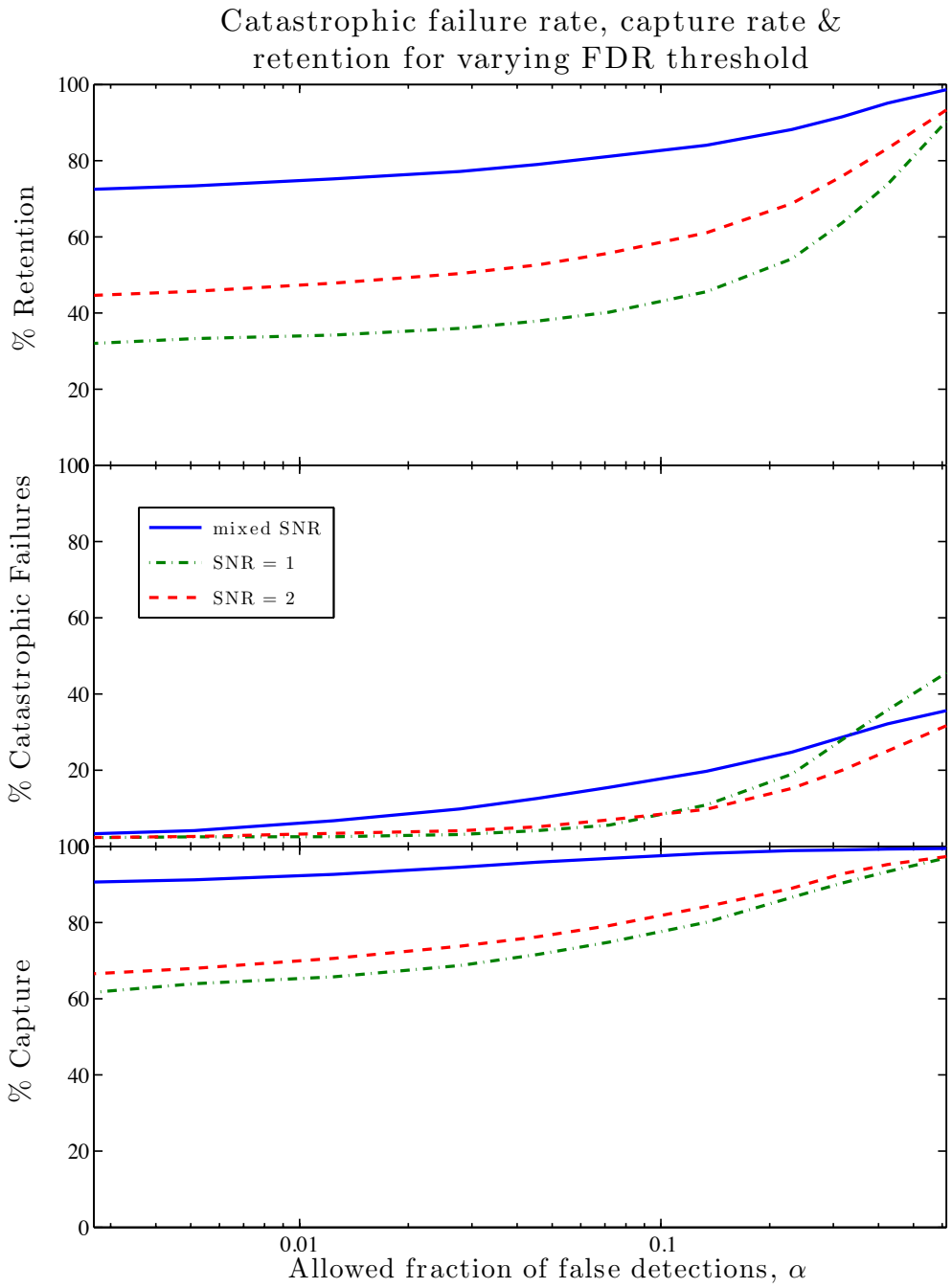


Figure 5.4: This figure illustrates the effect of the choice of FDR threshold on the catastrophic failure rate after cleaning, the retention and the capture rate on SNWG-2, SNWG-1, and VNPD. Note the greater sacrifices required both in retention and capture rate in order to obtain the same catastrophic failure rate at an SNR of 1.0 compared to 2.0. Note also that we are able to obtain a 1.8% failure rate in our redshift estimates for the cleaned catalogue, a retention of 15.0%, and a capture rate of 52.2% with the catalogue at an SNR of 2 at an FDR threshold of 4.55%.

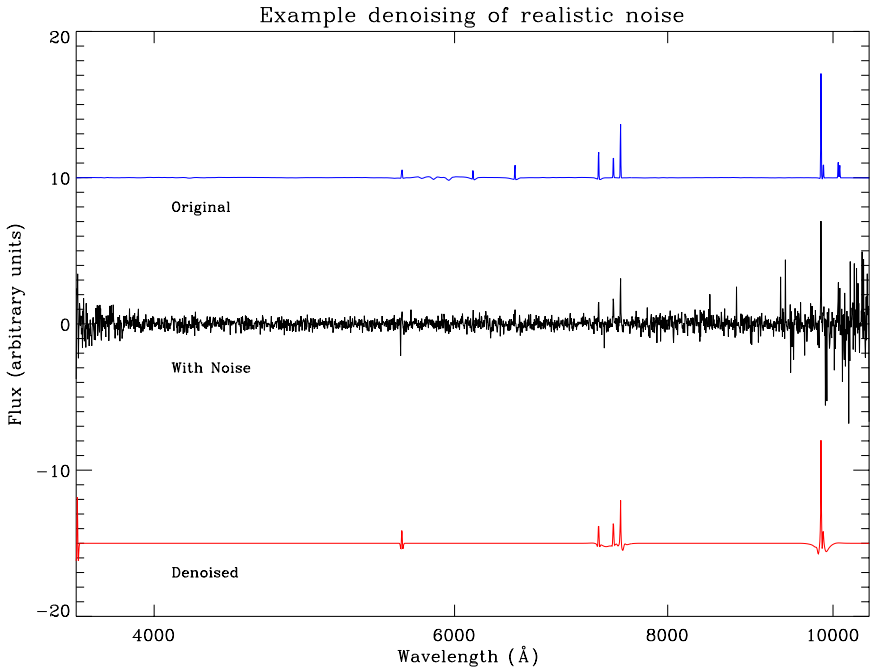


Figure 5.5: Denoising of test spectrum (c.f. figure 4.3, continuum-subtracted) with pixel-dependent noise. Note how most of the main features are detected and how, for this particular noise realisation, no false detections are found in the complicated & noisy long-wavelength region. We do incur an edge-effect false detection at the very short-wavelength end of the spectrum.

a biasing at the edges of the spectrum at both the long and short-wavelength ends, this arises due to a lack of information 'beyond' the edge limiting the ability to correctly characterise local features at the edge as either signal or noise; problems at edge regions are common to a substantial number of data analysis procedures and do not represent a flaw specific to the Darth Fader algorithm.

As can be seen in figure 5.6, the addition and subsequent denoising of flat noise behaves as one would expect; small deviations about a flat line situated at y=1 (shifted down in the figure for clarity). Minor artefacts are present at the edges due border effects specific to the wavelet transform, and features occasionally straddling the edges of the spectrum. The more complicated noise proves a considerably more difficult task than the flat noise, and clearly has some persistent residual features after denoising, particularly in the longer wavelength range where the error-curve is most complex. This discrepancy is due to the denoising not fully accounting for the rapidly changing noise from one pixel to the next.

Clearly this will impact on feature detection, resulting in a greater number spurious detections particularly at longer wavelengths. Increasing the FDR parameter, α, does provide significant improvement in the efficacy of the denoising (as shown by the middle curve). It may be possible to further ameliorate these systematic effects with a more optimal wavelet choice (should one exist), or by assigning a weight to each pixel to counterbalance the effect of the denoising not fully accounting for the complex noise properties. Additionally, figure 5.4 (solid blue line) shows that a stricter FDR thresholding is already effective in mitigating these systematic effects of the denoising.

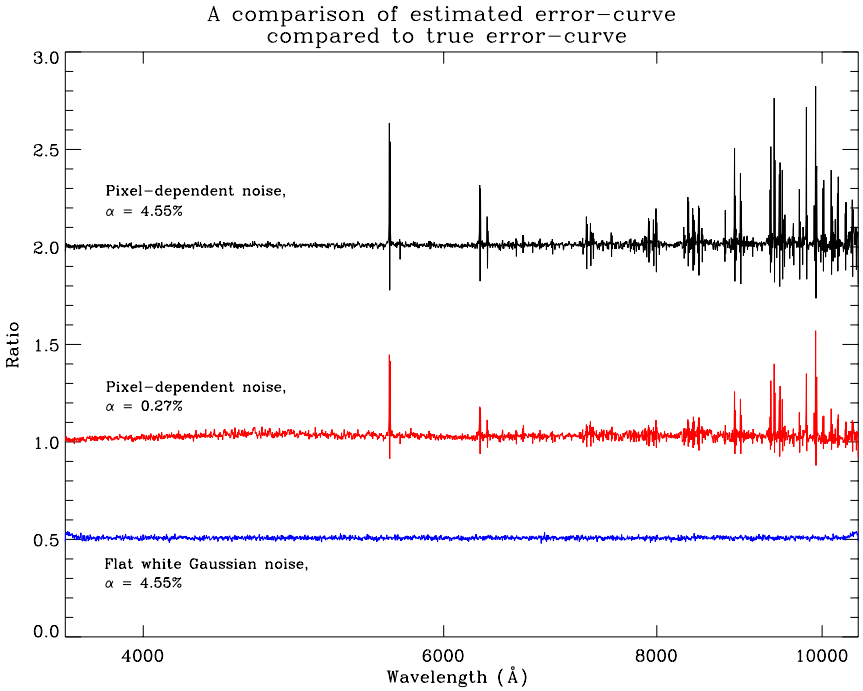


Figure 5.6: In this figure we plot the ratio of the true error-curve with respect to the derived error-curve from the rms error per pixel on the difference between the original input spectrum and the denoised spectrum for both flat noise and pixel-dependent noise. The lower curve (blue) has been shifted down (by 0.5) for clarity, and the upper curve (black), has also been shifted up (by 1.0) for clarity. Note the minor systematic edge effects on the denoising of white-Gaussian (flat) noise. Clearly the complex noise region has an marked systematic effect on the denoising, with rapidly changing noise regions experiencing both over and under estimates in the noise strength. This systematic effect is dependent upon the FDR threshold chosen, with thresholding that is less strict (upper curve) being more prone than stricter thresholding (middle curve).

5.4 Conclusions

In this chapter, we have applied the Darth Fader algorithm to idealised catalogues based on simulated spectra from the COSMOS Mock Catalogue (CMC). Details of these catalogues, their generation, and their added SNRs values based on an SDSS definition of SNR can be found in section 5.1.

Results for the SNR catalogues with single values of SNR and with white-Gaussian additive noise are given in section 5.2. A single snapshot of the operation of the Darth Fader algorithm, the impact of increasing noise, and the impact of changing the FDR parameter for denoising are investigated.

The simulations we use are of considerably lower resolution ($R \sim 850$ compared to $R \sim 1{,}845$) than would be expected for a modern day spectral survey, with SDSS resolution being over twice as high, and the forthcoming DESI survey (Levi et al. 2013) – a merger of the BigBOSS and DESpec surveys (Schlegel et al. 2011; Abdalla et al. 2012, respectively) – expected to be higher still. The wavelength range of our simulated spectra (3,000 Å to 10,500 Å) are slightly longer than would be expected for a realistic instrument ($3{,}500 \lesssim \lambda \lesssim 10{,}000$), but given the poorer resolution in our simulations, it is justifiable to extend the range. These factors do not, however, prevent these catalogues from being realistic.

In an effort to approach a more realistic catalogue, in section 5.3, we include a realistic instrumental response curve (1σ error) in order to generate noise that varies in intensity per pixel (but remains Gaussian in each pixel), and run the Darth Fader algorithm on a mixed SNR catalogue. In a

real (ground-based) survey however, the fixed instrumental response and the variable sky brightness produce a composite, variable, 1σ error curve per measured spectrum; with this composite error-curve being recorded together with said measurement. We have chosen to ignore this minor detail in this analysis, since it would not practically affect the results.

From these results we can conclude that the Darth Fader algorithm provides effective and robust denoising, regardless of whether the noise is stationary or wavelength dependent, provided that a choice of FDR parameter is made such that it is appropriate to the type of noise present.

Chapter 6

Real Data & Results on Real Data

Summary

6.1 Real Data & Results on Real Data

In the previous chapter, we demonstrated the robustness of the Darth Fader algorithm on simulations. Real data differ from our simulations in a number of important ways: rare galaxy types/properties may exist within real data catalogues, and these may not necessarily be well encompassed by our simulations; the spectral resolution in our simulations is poorer than would be expected for real instruments; and real data can often have more complex noise properties. It is therefore important to test whether our denoising methods, and feature-counting criterion, can be applied to real data.

Here we expand on the work in chapter 5 to show that our feature detection methods work well on real spectra from the SDSS archive. A full test on real SDSS data, however, would not be practical since, for any given spectrum, no value of redshift is known *a priori* (though this would be true for any real survey), and there is no guarantee that the redshifts given by SDSS are both accurate and unbiased. Performing a test on SDSS data could not therefore guarantee that deviations from the quoted redshifts are solely the result of the Darth Fader algorithm, or indeed solely the result of potential systematics in the SDSS catalogue itself. Furthermore, the low SNR spectra measured by the SDSS instrument are automatically discarded by magnitude cuts in the SDSS data processing pipeline and are not generally available in their data products. Spectroscopic cuts for the main galaxy sample are taken at magnitudes in $r < 17.77$ (Petrosian) and the resulting spectra have a median

SNR (per pixel) > 4 (Strauss et al. 2002). Additionally, since the data products available are of high SNR, and the Darth Fader algorithm will not be applicable to this regime and is not expected to outperform other methods.

A more appropriate test for the Darth Fader algorithm would be to apply it to a noisy data set, where the redshifts have been pre-determined by an alternative method. To that end we select the WiggleZ survey (Drinkwater et al. 2010; Parkinson et al. 2012) as an ambitious test for the algorithm, where the SNR is generally significantly poorer than the SDSS, and the majority of the redshifts have been determined by eye rather than (solely) algorithmically. This survey is described further in section 6.4. Below we first demonstrate that the Darth Fader algorithm can identify features in real data from the SDSS archive.

6.2 Example Spectra from the SDSS Catalogue

In order to demonstrate the broader applicability of Darth Fader to real spectra at higher resolution and covering a narrower wavelength range, we take three SDSS galaxy spectra and their respective 1σ error-curves – those of an emission line galaxy (ELG), a luminous red galaxy (LRG) and a 'typical' galaxy – as fiducial type galaxies that well represent the SDSS galaxy catalogue.[1]

These spectra have a resolution $R\left(\propto \lambda/\Delta\lambda\right)$ significantly higher than that of our simulations, namely $R \sim 1{,}845$ compared to $R \sim 850$, and as such features are better separated. The r-band SNR for these galaxies is quoted to be 9.2 for the ELG, 9.3 for the LRG, and 15.0 for the typical galaxy, respectively.

In denoising these spectra we use an FDR threshold of $\alpha = 0.27\%$ as motivated by the discussion in § 5.3 and the results in figure 5.4. We apply a positivity (and 'negativity') constraint, as before, to denoise the positive and negative sections of each spectrum independently, and recombine them to form the final denoised spectrum. The procedure uses the same positivity constraint, once on denoising the spectrum, and once on denoising the reverse-signed spectrum – this is entirely equivalent to denoising once with a positivity constraint, and again with a 'negativity' constraint.

6.3 Results for Real SDSS Spectra

In figure 6.1, we show the continuum-subtracted spectrum, the FDR denoised spectrum, and the line features we detect for the emission-line galaxy. We also plot the 1σ error-curve, which we assume as Gaussian.

This ELG spectrum has many strong features, so it is not surprising that the FDR denoising detects most of them. We do however miss one very weak emission feature that is comparable to the noise, at $\sim 7{,}800$ Å. It should also be noted that the potential line-like features arising from the noise, namely at $\sim 5{,}600$ Å and again at $\sim 8{,}950$ Å are completely ignored by the FDR denoising since, by supplying the error-curve, these features are correctly identified as likely arising from noise rather than signal.

Feature detection in the LRG spectrum (figure 6.2) presents a more difficult challenge. Despite the signal-to-noise ratio in the r-band being of similar value to the ELG, the sizes of the features compared to the noise (i.e.: the signal-to-noise values on the lines) are much smaller. We successfully detect five absorption features, despite them not being at all prominent. There are a further three

[1] These three example galaxies can be found at: http://www.sdss.org/gallery/gal_spectra.html with Plate ID: 312, MJD: 51689 and Fibre IDs: 220 (LRG), 255 (Typical), & 529 (ELG).

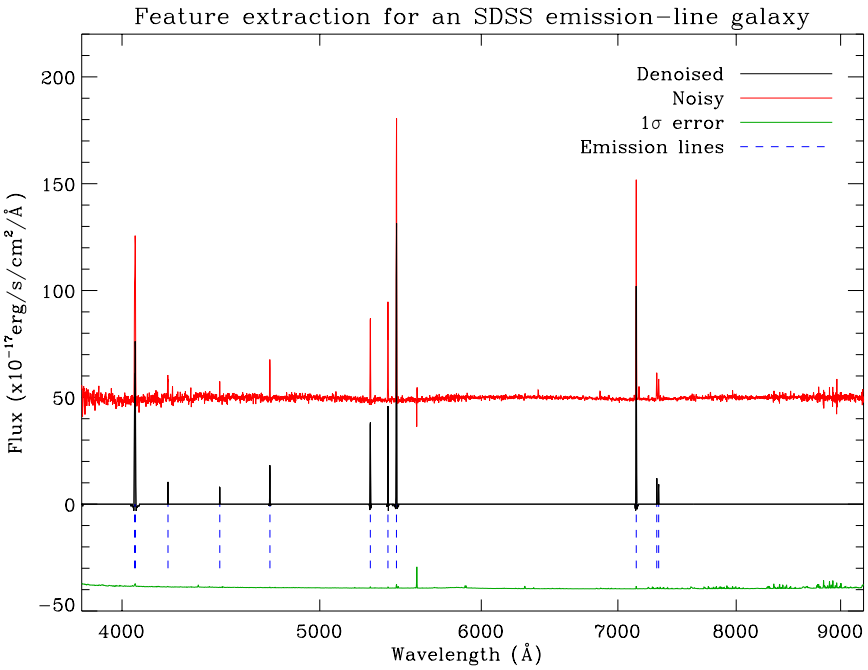


Figure 6.1: Denoising and feature extraction for an SDSS ELG. The noisy spectrum (red) has been shifted up, and the error-curve (green) shifted down, for clarity. The vertical dashed lines (blue) indicate the locations of detected features that correspond to true emission features. The FDR denoising and feature extraction clearly pinpoints all of the major features without any difficulty. The three largest lines are, from left to right, the [O II] doublet, [O III] and H_α.

detections, one at $\sim$ 6,400 Å and two between 7,000 Å & 8,000 Å that may or may not be spurious, as we cannot easily associate them to any common lines. The one just before 9,000 Å is almost certainly spurious due to the high noise in that region.

The results for the typical galaxy are similar to those of the LRG (figure 6.3). In this case, we again detect all five of the absorption features, in addition we obtain some unidentifiable features that may or may not be spurious, and do not appear to be associable to common lines.

For each of the galaxy types shown here we can detect at least six features, though it is possible that not all of them are true detections, and we do not require them to necessarily be true detections. Though we only consider Gaussian noise here, the tools used in Darth Fader are in principle not limited to purely Gaussian errors, and can be utilised with different types of errors (in particular Poisson, Gaussian + Poisson, multiplicative and correlated errors), and provided that denoising can be done appropriately the impact on the Darth Fader method will be minimal.

We have demonstrated that it is possible to detect features in relatively good quality data, thought we wish to determine if this can be properly extrapolated to lower quality data, with an overall poorer signal-to-noise which is the goal of the Darth Fader algorithm. We proceed with an analysis of data from the WiggleZ survey below.

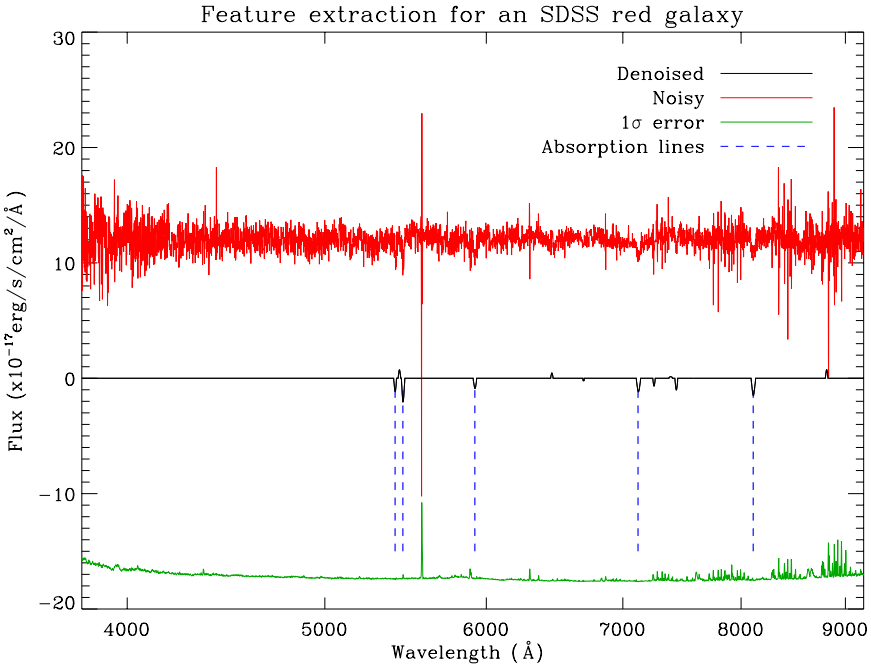


Figure 6.2: Denoising and feature extraction for an SDSS LRG. The absorption lines from left to right are CaII (H and K), G-band, MgI and NaI. (Note: the G-band is not strictly an absorption *line*, but rather an aggregate absorption feature due to the presence of multiple lines arising from metals (mainly iron) in the numerous G-type stars present in the galaxy population. Also not to be confused with the SDSS photometric filter g-band).

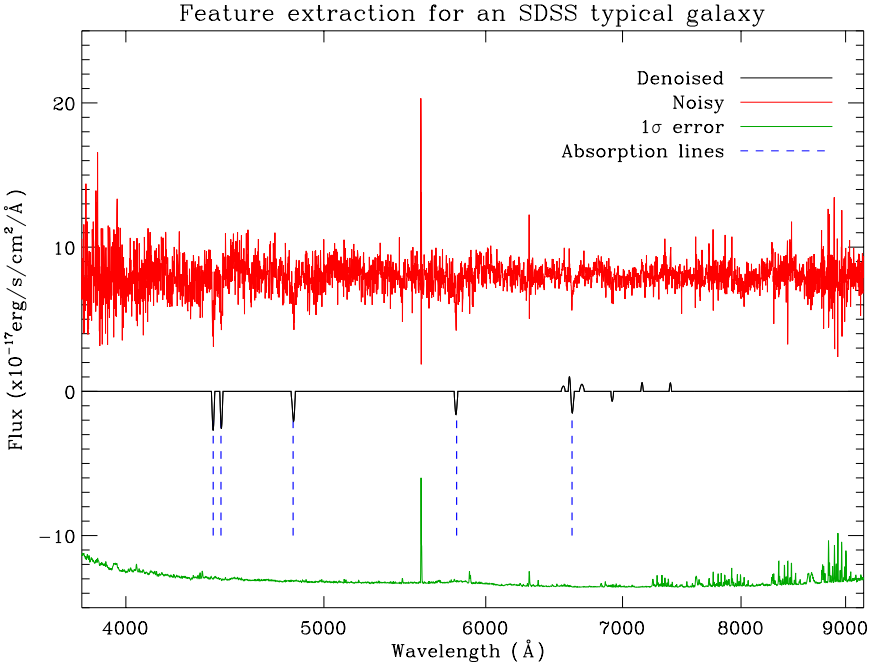


Figure 6.3: Denoising and feature extraction for an SDSS typical galaxy. This spectrum is similar to that of the LRG, the highlighted absorption lines being the same as previously.

6.4 WiggleZ Survey

6.4.1 Survey Design

As mentioned previously in section 6.1, an ambitious test of the Darth Fader algorithm on real data would be an application to spectra from the WiggleZ survey (Drinkwater et al. 2010; Parkinson et al. 2012). We select this survey in particular due to the majority of the redshifts being determined by visual inspection (after an initial algorithmically assigned redshift), and due to the generally very low signal-to-noise of the measured spectra.

The WiggleZ survey itself is a ground based survey operating from the Anglo-Australian Telescope (AAT) in New South Wales, Australia. The survey targets at emission line galaxies (ELGs) in the UV range, covering a volume of approximately 1 Gpc^3 and an area on the sky of 1,000 deg^2. The redshift range of the catalogue primarily spans the region $0.2 < z < 1.0$, and contains on the order of 240,000 spectra, spanning a wavelength range of 3,700-8,500 Å in the first phase, later changing to 4,700-9,500 Å in the second phase (after an upgrade of the spectrograph). Targets are selected based on both NUV and r-band magnitude limits, with $NUV < 22.8$ & $20 < r < 22.5$. The resolution of the spectra is quoted as $R \approx 1300$.

The scope of the survey is to determine the characteristic BAO scale in large scale structure at higher redshifts than previously and verify the important cosmological parameters of the $\mathbf{\Lambda}$CDM model.

The reasoning for targeting these ranges in redshift and wavelength is primarily for the following: the peak in star formation rate beyond $z > 0.5$ implies the existence of a significant population of galaxies with strong emission lines in this range; this in turn allows reliable redshift estimates to be made from relatively short exposures since the redshift can be deduced from the lines; furthermore there exists pre-existing imaging data from the MIS survey (Martin and GALEX Science Team 2005) for GALEX (Galaxy Evolution Explorer) (Bianchi et al. 1997) in the UV range; additionally SDSS supporting data (in the visible) for the targets already exists and can be used to probe any systematics in the data. An additional aspect of this survey is the secondary science possible with the data with respect to the study of star forming galaxies at high redshift.

A disadvantage of targeting these galaxies is that the clustering signal is less pronounced than for red galaxies, and thus more susceptible to noise, however, the BAO signal is also more prominent in this regime since the non-linear growth of structure is reduced (which would tend to obscure the BAO signal).

6.4.2 Data: products, quality and analysis

Data is reduced algorithmically via a PCA method in order to remove sky lines and cosmic ray features from the spectrum. The data is further processed through a redshift estimation pipeline, near-exclusively relying on emission-line matching, to obtain a preliminary redshift estimate. Drinkwater et al. concede that due to the presence of high noise and spurious features, many of the spectra possess an automated redshift estimation that is unreliable. As a consequence of this unreliability, the same pipeline issues a quality value for each spectrum: ranging from $Q = 1$ (poor quality) to $Q = 5$ (excellent quality); the spectra are then visually inspected in order to reassess the redshift estimate, and the quality.

Quality is defined as follows: 1 represents a very noisy spectrum, such that no redshift assignment is possible; 2, a tentative redshift can be assigned, though it is highly uncertain, and some confusion between [O II] and sky line residuals are present in this group; 3, redshift can be assigned with

confidence, more than one emission line, or solely the [O II] doublet (at least partially resolved), is present; 4, represents spectra with many emission lines in mutual agreement with the inferred redshift, and are thus very good candidates for a positive estimation; and lastly a quality value of 5 denotes excellent spectra with high SNR, with the potential for using these as templates. Using spectra of a quality of 2 or below for analysis is cautioned against.

The algorithmic part of the analysis operates to find prominent peaks and derives the corresponding redshift from them, checking to see if this fits with other expected lines for this initial redshift determination. A preliminary quality determination is given, to then be reassessed by visual inspection.

Drinkwater et al. caution that the spectra still possess artefacts such as cosmic-ray and skyline removal residuals and that they also suffer from fibre-fringing and cross-talk. These factors will necessarily imply that many spectra will have additional features that have an origin other than the galaxy spectrum of interest, with the possibility of producing spurious features that will consequently not be removed during denoising.

The signal-to-noise ratio of the spectra involved in the survey are on the order of about 1 per pixel (on the continuum), with the additional requirement that they possess an SNR > 3 in the NUV filter band. Redshifts are identified almost exclusively on the emission lines present, with WiggleZ predominantly targeting the [O II] (3,727 Å), H_β, and [O III] (4,959,5,007 Å) emission lines; at lower redshifts the [O II] line is beyond the wavelength range of the spectrum, however this is often compensated by the coming into range of the H_α line. Additionally, at higher redshift ($z > 0.95$), the emission lines present move toward the improved resolution longer wavelength range and hence the [O II] doublet can reach the stage where it becomes partially or even completely resolved, allowing for a redshift determination with a single line.

The WiggleZ survey claims a success rate of ∼60% for $Q \geq 3$ all observations, with this rising to 70% on the inclusion of repeated observations on some objects. Evidently if we can perform near this level or better, only by algorithmic means, this would represent a significant achievement of the Darth Fader algorithm - even if the subset retained by the algorithm is reduced in size compared to the entire dataset, this could still significantly reduce the need for such extensive visual inspection, and the subset could be used to inform further investigation of the discarded dataset. Additionally, there may be scope for re-assessing some $Q = 2$ galaxies with the algorithm as there is potential that this subset of the data could be cleaned and the resultant cleaned data be useful for analysis.

6.5 Darth Fader test on real WiggleZ spectra

6.5.1 Refinements to the Darth Fader algorithm

Between publication of Machado et al. (2013), and the tests performed below, the initial version of the Darth Fader algorithm has undergone some modifications for the purposes of public release. Significant work has been made for computational efficiency and for the aim of making the algorithm more user-friendly. Furthermore two important changes have been made in algorithmic design: a modification of the implementation of the starlet wavelet transform (which is now 2^{nd} generation) for ringing reduction; and the improvement of flagging by including the option of line-matching to standard emission and absorption lines, (the set of lines to match is adaptable to the needs of the user and the targets involved).

In the following analysis we select a subset of 3,000 galaxies from the WiggleZ catalogue as our test catalogue, and construct two sets of eigentemplates: one from the CMC mock catalogue, and another

as a subset of the high quality ($Q = 4, 5$) and low redshift error ($z_{\rm err} < 0.0005$) spectra present in the WiggleZ catalogue, blueshifted to be at the same redshift ($z = 0$). We rebin these onto a logarithmic wavelength grid using a Lanczos resampling, at a resolution of $R \sim 850$.

Generation 2 starlet

The implementation of the 2$^{\rm nd}$ generation starlet wavelet transform is chosen for its ability in reducing ringing artefacts when compared to the 1$^{\rm st}$ generation. The mechanism for this transform is very similar to the one in section 2.3.2, and for the same scaling function (based on the B_3-spline as before, equation (2.16)) the wavelet function is slightly different to the one in equation (2.17),

$$\frac{1}{2}\psi\Big(\frac{x}{2}\Big) = \phi(x) \; . \tag{6.1}$$

With the 1$^{\rm st}$ generation starlet wavelet being a difference of scaling functions it is not everywhere positive; the important difference with the 2$^{\rm nd}$ generation is that the wavelet function is everywhere positive (strictly speaking, everywhere non-negative), this is due to the scaling function being positive everywhere and the wavelet function being directly related to it by a simple scaling. The effect of this is to reduce ringing by maintaining each wavelet coefficient as positive.

Line matching

Since redshift estimation in the WiggleZ survey is almost exclusively emission-line derived, and the data suffers from strong contamination, it would be prudent to include the possibility of using line information in the flagging process. The principal lines used for redshift estimation in the survey have been mentioned previously in section 6.4.2, and of these, the most ubiquitous lines in the survey are those of [O II], $\rm H_\beta$, and the stronger line of the [O III] doublet (5,007 Å).

Hence we modify the flagging mechanism to accommodate the detection of important lines whereby a feature found to be near to the expected location of a prominent line - given the cross-correlation redshift (with the pixel distance being determined by a function of $(1 + z)$). The spectrum is kept under this new flagging mechanism if it possesses these prominent lines - irrespective of the number of features detected.

We include the option of adding any line information that the user may wish to add as an input to the algorithm, making the algorithm adaptable to survey setups other than WiggleZ. Additionally we include the option of both 'OR' & 'AND' logical operations with such lines; the retention of the spectrum is predicated upon the detection of features at all (emission) line locations specified (AND) or at least one of the line locations specified (OR) given the estimated redshift, and the expected location of those lines.

As should be evident, the requirement of the detection of all lines under consideration will be highly constraining (dependent upon the number of different lines tested) and many spectra could be unnecessarily rejected. Additionally, requiring two lines that are distant in wavelength places fundamental constraints on the redshift range at which successful estimates can be made - for example requiring both [O II] and $\rm [O\,III]_b$ (at 3,727 and 5,007 Å respectively) to be detected necessarily imposes theoretical bounds on the redshift of $0.26 < z < 0.90$, which is narrower than the bounds of the WiggleZ survey.

6.5.2 Darth Fader Results on WiggleZ data

We perform an analogous analysis on WiggleZ data to the one performed on the simulated data as presented in chapter 5, utilising the same definitions as previously for the retention and catastrophic failure rates (equations (5.2) and (5.3) in section 5.1). We also include the capture rate (as defined previously, equation (5.4)), however it should be cautioned that the concept of capture rate will be less meaningful in this analysis since we have no a priori knowledge of which redshifts provided by the WiggleZ survey are truly correct, particularly as these themselves have an estimated catastrophic failure rate of $\sim$30% for $Q \geq 3$ spectra. An implicit assumption in all subsequent analysis is that the WiggleZ estimates for all the redshifts of the spectra in the survey are correct.[2]

It remains an open question as to whether better results for redshift estimation can be obtained from data or from synthetic templates. Hence we perform an analysis utilising a selection of templates from the CMC mock catalogue (analogous to as was previously done in chapter 5), and we additionally perform an analysis taking a subset of simultaneously high quality and high SNR spectra from the WiggleZ catalogue, blueshifting them to use as a template set in the generation of eigentemplates. It should be noted that the CMC mock catalogue was not constructed to be representative of the WiggleZ survey and hence will likely not contain the appropriate proportions of galaxy types; notably the proportion of elliptical/low star forming galaxies will likely be much higher than the WiggleZ survey which is specifically designed to target emission line galaxies and can make little use of spectra that contain predominantly continua in the brief 1-hour exposures available.

In this comparison between the effectiveness of CMC or data derived templates, they are compared under an 'OR' line-matching option for [O II] & H_β, i.e.: if spectra possess either of these features after the FDR denoising, then the redshift is kept.

As can be immediately seen in figure 6.4 the like-for-like comparison between utilising eigentemplates generated from the CMC mock catalogue does not perform as well as eigentemplates generated from a subset of the WiggleZ data (at least under this OR line-matching criterion). Small improvements of a few percent in capture rate can be seen for the use of WiggleZ data over CMC mock catalogue derived templates, larger differences can be see in the retention (a consistent 8% improvement), but the largest, and most relevant difference to this analysis, is the catastrophic failure rate which is improved by about 15% for the majority of FDR parameter value choices. The outperforming of the CMC mock catalogue by an eigentemplate set generated from the data itself should not be surprising since the CMC mock catalogue is unlikely to be representative of the WiggleZ survey since it does not contain the expected proportions of the different morphological spectral types. Hence for future analysis of the WiggleZ data that follows, we no longer utilise the CMC mock catalogue.

Additionally observable in figure 6.4 is the reduced impact of the FDR denoising when compared to the results of the mixed SNR realistic noise catalogue (VNPD) in figure 5.4. This is explained by the heavy contamination of the spectra and the greater skew of low SNR spectra (on the order of 1 on the continuum). The heavy contamination with line-like features (from fringing, cosmic rays, and improper sky-line removal for example) will be treated by an FDR feature detection procedure as true features, and this is not reduced by the inclusion of the RMS error curve. These contaminating features can also frequently be of a large amplitude (appear as high fluxes in selected pixels) and are therefore not often contained within the noise, and are thus likely to be preserved after a denoising; therefore although they are indeed 'false' features, they will not be counted as false detections.

[2]Given that these spectra have been segregated on quality, visually inspected, and have a modest catastrophic failure rate, this is a reasonable assumption for most of the galaxies in the survey.

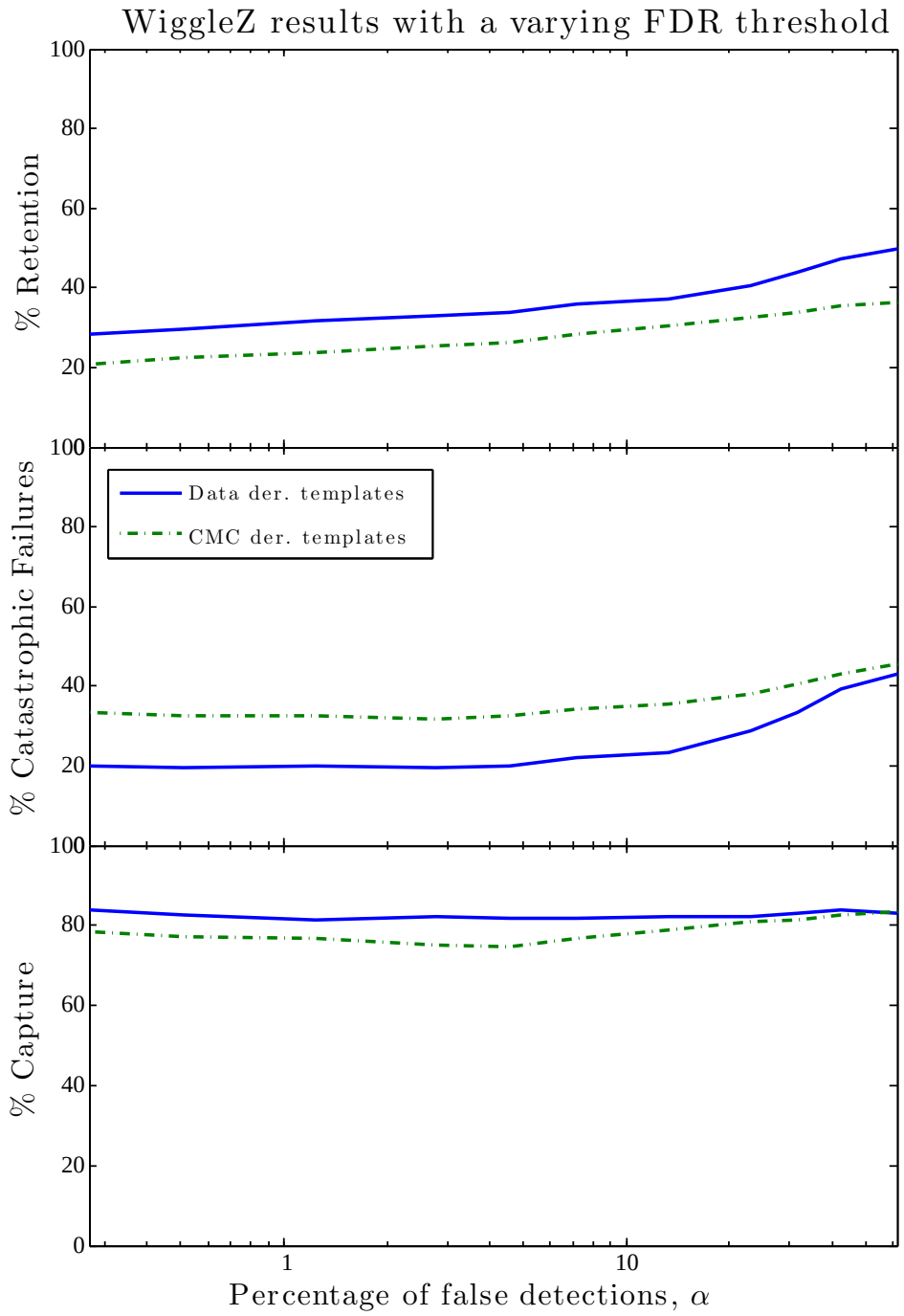


Figure 6.4: This figure shows the effect of varying the FDR parameter, α, on the catastrophic failure, retention and capture rates of a test set of 3,000 WiggleZ galaxies for two separate eigentemplate sets: one having been derived from the CMC mock catalogue (dot-dashed, green) and the other derived from the WiggleZ survey itself (solid, blue). Both sets of results are based on an 'OR' selection criterion.

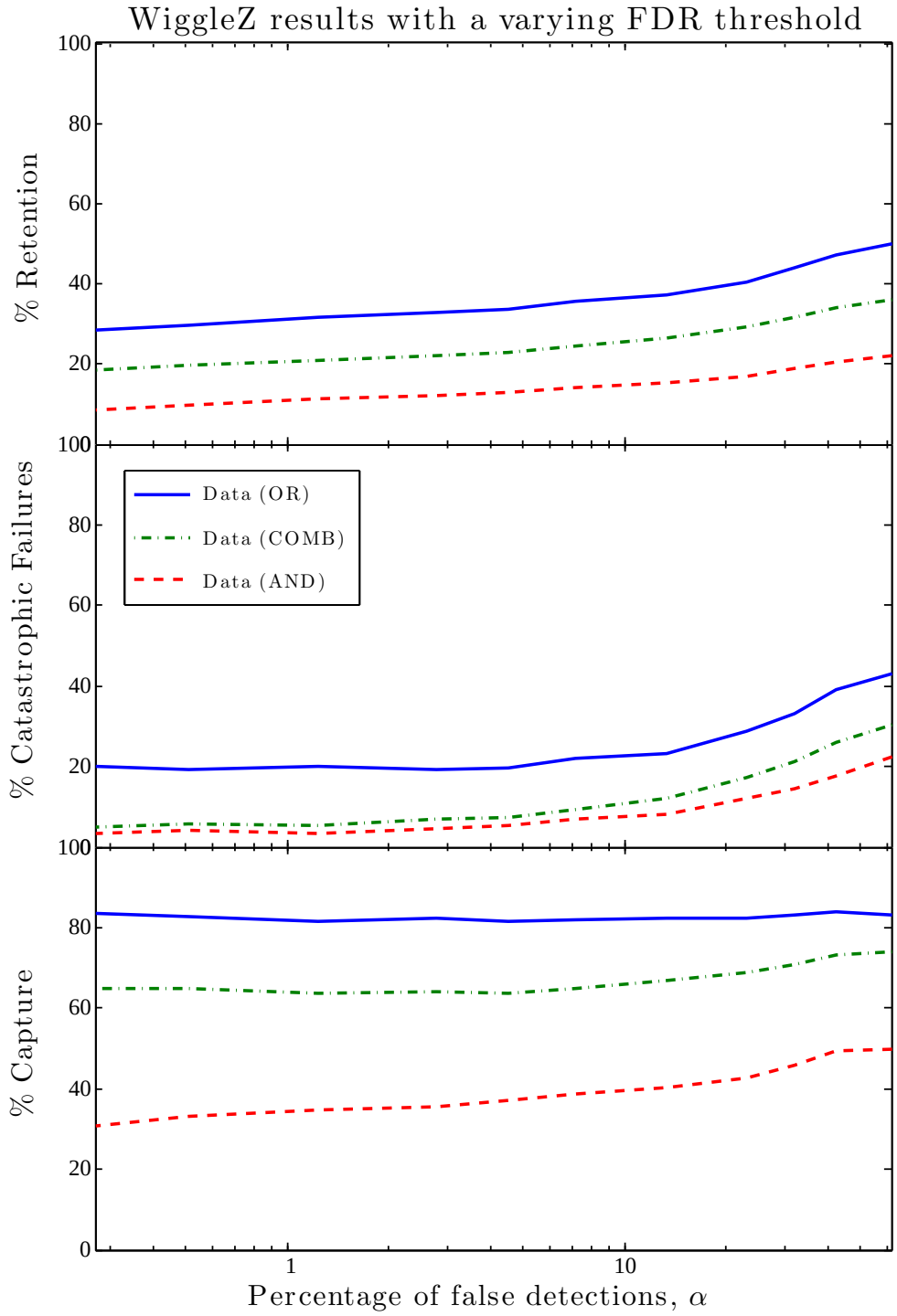


Figure 6.5: This figure shows the effect of varying the FDR parameter, α, on the catastrophic failure, retention and capture rates of a test set of 3,000 WiggleZ galaxies, utilising a template set derived from a high quality subset of the WiggleZ data. The results shown are for the different selection options, a comparatively unrestrictive 'OR' ([O II], H_β), a restrictive 'AND' [O II] & H_β, and a semi-restrictive 'COMB' combination where any two of the set {[O II], H_β, $[O\,III]_b$} is minimally required. (The data for the 'OR' selection is identical to the data in figure 6.4 and is shown again for convenience).

We perform a further test comparing the effects of a combined ('AND') selection of both [O II] & H_β, and (with the prospect of this potentially being too restrictive) an additional test composed of both 'OR' and 'AND' combinations of features as follows: if a spectrum possesses any two, or more, of the set of features $\{[O\,II], H_\beta, [O\,III]_b\}$ then it is kept, otherwise it is discarded (hereafter, 'COMB'). These 3 particular features are chosen based on empirical tests of their prevalence in the spectra from the WiggleZ survey, whose selection composition can be seen as 'OR' combinations of 4 pairs of 'AND' choices.[3] These results are depicted in figure 6.5.

As can be seen from figure 6.5 selecting spectra based on a simultaneous detection of both [O II] & H_β ('AND') is very effective in reducing the overall rate of catastrophic failures, however, it is also very restrictive: both retention and capture rates are impacted strongly as a result. We obtain a very low catastrophic failure rate of 5.4% but only retain 13% of the galaxies at an FDR parameter value of 4.55% allowed false detections. This would equate to a subset of ∼31,000 galaxies from the total WiggleZ catalogue.

The 'COMB' selection, by admitting only very marginal increases in the rate of catastrophic failures allows us to gain consistent increases in both retention and capture rates. Thus the performance of the 'COMB' selection can be seen as intermediate in terms of both the retention and capture rates, however the rate of catastrophic failures closely mirrors that of the more restrictive 'AND' selection, and hence we can conclude that this combined feature selection option is superior to the more basic 'AND'. For the same choice of FDR parameter (4.55% allowed false detections), a 'COMB' feature selection option yields a marginally larger catastrophic failure rate than 'AND' at 7.3%, however the retention and capture rates of 22.8%, and 63.7% respectively, are comparatively enhanced under the 'COMB' option. Such a performance of 'COMB' extrapolated to the entire WiggleZ catalogue would represent ∼55,000 galaxies with a redshift known to a high precision and a potential near-quartering of the man-hours involved in analysing the data.

Alternatively, if we were just seeking to perform as well as the ∼30% catastrophic failure rate of the WiggleZ survey, this would be best obtained under the 'OR' option at the rather relaxed FDR parameter value of α =23.0% allowed false detections, yielding a catastrophic failure rate of 28.8% (similar to the overall WiggleZ catastrophic failure rate for all $Q \geq 3$ galaxies), with a retention of 40.6% of the catalogue - this representing an overall reduction of 40% of the man-hours involved whilst still yielding a similar result.

[3] In full this is a combination of: ([O II] AND H_β) OR (H_β AND $[O\,III]_b$) OR ([O II] AND $[O\,III]_b$) OR ([O II] AND H_β AND $[O\,III]_b$), with any other features, or combinations of features, considered as not relevant.

6.6 Conclusions

In chapter 5 our analysis focused on catalogues which were simulated, and as such may have represented catalogues that were overly simplistic compared to real data; they did however fully represent the expected various galactic morphological types, distribution and redshift properties. The noise properties we used in our simulations may also have been less complicated than those found in real data, however, non-stationary Gaussian noise (varying per pixel) is a good enough approximation to real data. To overcome these possible difficulties, it was necessary to see if the Darth Fader algorithm could tackle real data.

We have shown in section 6.2 that the competent denoising of the SDSS example spectra (figures 6.1 to 6.3) is possible with the algorithm, and that feature detection can be made in real data effectively.

We have expanded our analysis to include tests on real data, from the WiggleZ survey section 6.4, which to date has been analysed primarily by visual inspection (with some prior cataloguing by algorithmic means).

We have shown, with some minor adaptations to the Darth Fader algorithm - a change in the type of starlet transform, and a modification to the method of feature selection - that the low signal-to-noise regime can be tackled effectively, demonstrating its applicability to data from the WiggleZ survey, and producing both comparable results, and a significantly enhanced result for a reduced subset of galaxies, entirely algorithmically.

The Darth Fader algorithm could therefore have been used to reduce the number of man-hours needed for the visual inspection of the spectral redshifts for the WiggleZ survey, and/or additionally it has the potential to be used for a full secondary analysis of the data. There may even be scope - after some signifiant work on the proper removal of contamination - to tackle the $Q = 2$ data and thus 'rescue' some useful data from it, though this is entirely speculative.

Conclusion

The goal of this work has been to improve redshift estimation in low signal-to-noise regimes. This is important because current surveys usually bluntly discard data that does not meet specific magnitude or SNR limits, which could very well be discarding still-relevant data. Additionally, low SNR catalogues can suffer from unduly high catastrophic failure rates, which can have important impact on secondary studies that need accurate spectral redshift data, for example weak lensing and photometric calibration. Future work and discovery lies on the frontier of these magnitude and SNR limits, and attempting to stretch surveys to include as much data as possible, without incurring unduly poor data is an important goal and we have developed some key tools with our Darth Fader algorithm that can do just this.

⋆ Continuum removal - We have shown that the continuum of a spectrum can be removed empirically with a wavelet decomposition. Continuum removal in this way, when compared against elaborate physical modelling, may be seen as a comparatively naïve method. However, there is no loss of generality in its usage in cross-correlation based redshift estimation methods, and it benefits from being a blind method requiring no prior knowledge of how galactic spectra arise.

The wavelet-based continuum subtraction procedure used in Darth Fader is in principle not limited to galactic spectra, and preliminary tests suggest that it will prove useful for the continuum-modelling of the more structurally rich spectra of stars. Indeed, for any spectra whose components - continuum and features (and noise) - are distinct, provided an appropriate choice of wavelet is made, we expect our empirical continuum subtraction method to work as demonstrated.

⋆ True Feature Detection with FDR Denoising - We have shown that noise can be effectively removed whilst still preserving features that are (majority) real features rather than spurious ones. This is done by the application of a False Detection Rate denoising in wavelet space. A useful aspect is the tuneable FDR parameter α, which allows us to control the threshold for the proportion of false detections we admit in the denoising with a more relaxed choice allowing for more true detections, but also admitting the possibility of more false ones, a stricter choice leaves little possibility of detected features being false, but is liable to rejecting some true features as noise.

We only considered the numbers of features in the simulated data, however the ability of the Darth Fader algorithm to detect likely true features was readily adapted to deal with feature *identification* in the real data from the WiggleZ survey. In particular for spectra where noise

levels are very high, our method provides an advantage over $K\sigma$ denoising since this would have a stronger tendency to remove signal (including the features we seek to identify) as well as noise.

As we have shown, Darth Fader is a powerful tool for the improvement of redshift estimation without any *a priori* knowledge of galactic composition, type or morphology; its predominantly empirical nature enahnces the adaptability of the algorithm for use with different surveys, and indeed different applications.

We can successfully make an estimate of the continuum without needing to know the physical model behind the spectra, and we can confidently make use of data at signal-to-noise levels that were previously beyond the reach of other techniques. This is achieved by denoising the data with an appropriately chosen false detection rate threshold and implementing a simple feature-counting criterion, resulting in very low catastrophic failure rates for redshift estimation for a subset of galaxies in the catalogue.

This is the most useful aspect of Darth Fader - it can be used as a flagging mechanism to extract what is likely to be good data for redshift estimation from what is likely to yield an inaccurate redshift estimate, with a good level of confidence. Even at signal-to-noise levels as low as 1.0 on the continuum (typical of the WiggleZ survey data), the algorithm can identify a subset of this catalogue - approaching a quarter of its total size - with a catastrophic failure rate of just under 8%, and this is for a catalogue that previously has required human intervention in the form of visual inspection in order to achieve a full cataloguing of the redshifts of the spectra in the survey.

This cleaning, and the low catastrophic failure rates that we can reach, therefore have applications in the calibration of photometric redshift estimation for other large surveys, such as the upcoming Euclid survey (Refregier et al. 2010), which will require such a spectroscopic redshift catalogue with very few catastrophic failures. Photometric redshift errors impact greatly on weak lensing and large-scale structure studies, which Euclid and other large surveys aim to probe, and accurate calibration is therefore an important ingredient for the success of such missions.

The Darth Fader algorithm represents a potential greater reach of spectroscopic surveys in terms of depth, since the faintest (and thus noisiest) galaxies in a survey - those at the detection limit of the instrument - will tend to be those at higher redshifts. Current methods of spectroscopic survey redshift estimation usually employ signal-to-noise cuts or magnitude/flux limits, resulting in low signal-to-noise data being treated as unreliable and thus unusable. Our algorithm demonstrates that these current methods, with a blunt cut-off in the signal-to-noise/flux for what is considered to be informative data, can be significantly improved upon, and that improvements are available all the way down to very low signal-to-noise levels, offer a more sophisticated mechanism for selecting spectra in the 'grey-area' where redshifts could be correct, or incorrect.

The levels of retention of catalogues presented in this work may seem moderate, however, for such low signal-to-noise data they can only be expected to be so, as redshift estimation necessarily fails for a large fraction of spectra at these high noise levels; where our algorithm excels is in the reduction of catastrophic failures. Hence – even when overall retention values appear moderate – the low rate of catastrophic failures together with the empirical and automated nature of the algorithm offers a

substantial gain when the alternative is to throw away the entire dataset with a blunt SNR cut, or to employ the efforts of people to undertake visual inspection of the entire catalogue (and whose man-hours could be put to better use), which will necessarily become increasingly impractical with larger datasets.

Darth Fader is clearly useful for both redshift estimation and empirical continuum estimation and will be made publicly available as part of the **iSAP**[4] suite of codes. The empirical nature of our algorithm, together with the ability to handle realistic noise, and tests on difficult to tackle data, show promise for its inclusion in future spectral survey pipelines and data analyses.

[4]**iSAP** package available at: http://www.cosmostat.org/software.html

Bibliography

R. Aaij, C. Abellan Beteta, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, and et al. First Observation of CP Violation in the Decays of B_s^0 Mesons. *Physical Review Letters*, 110(22):221601, May 2013. doi: 10.1103/PhysRevLett.110.221601. 14

F. Abdalla, J. Annis, D. Bacon, S. Bridle, F. Castander, M. Colless, D. DePoy, H. T. Diehl, M. Eriksen, B. Flaugher, J. Frieman, E. Gaztanaga, C. Hogan, S. Jouvel, S. Kent, D. Kirk, R. Kron, S. Kuhlmann, O. Lahav, J. Lawrence, H. Lin, J. Marriner, J. Marshall, J. Mohr, R. C. Nichol, M. Sako, W. Saunders, M. Soares-Santos, D. Thomas, R. Wechsler, A. West, and H. Wu. The Dark Energy Spectrometer (DESpec): A Multi-Fiber Spectroscopic Upgrade of the Dark Energy Camera and Survey for the Blanco Telescope. *ArXiv e-prints*, arXiv:1209.2451, September 2012. 47, 78, 86

C. P. Ahn, R. Alexandroff, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, É. Aubourg, S. Bailey, F. A. Bastien, and et al. The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment. *ArXiv e-prints*, July 2013. 47

R. A. Alpher and R. C. Herman. On the Relative Abundance of the Elements. *Physical Review*, 74: 1737–1742, December 1948. doi: 10.1103/PhysRev.74.1737. 12

R. Antonucci. Unified models for active galactic nuclei and quasars. *ARA&A*, 31:473–521, 1993. doi: 10.1146/annurev.aa.31.090193.002353. 17

S. Arnouts, S. Cristiani, L. Moscardini, S. Matarrese, F. Lucchin, A. Fontana, and E. Giallongo. Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North. *Monthly Notices of the Royal Astronomical Society*, 310:540–556, December 1999. doi: 10.1046/j.1365-Sem inaireBourbaki.8711.1999.02978.x. 53, 56

I. K. Baldry, K. Glazebrook, J. Brinkmann, Ž. Ivezić, R. H. Lupton, R. C. Nichol, and A. S. Szalay. Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *ApJ*, 600:681–694, January 2004. doi: 10.1086/380092. 18

J. E. Barnes. Transformations of galaxies. I - Mergers of equal-mass stellar disks. *ApJ*, 393:484–507, July 1992. doi: 10.1086/171522. 17

J. E. Barnes and L. Hernquist. Dynamics of interacting galaxies. *ARA&A*, 30:705–742, 1992. doi: 10.1146/annurev.aa.30.090192.003421. 17

C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69:3101–3156, December 2006. doi: 10.1088/0034-SeminaireBourbaki.4885/69/12/R02. 50, 53

W. A. Baum. Photoelectric Magnitudes and Red Shifts. In *Problems of Extra Galactic Research*, volume 15 of *IAU Symposium*, page 390, 1962. 20

E. F. Bell, C. Wolf, K. Meisenheimer, H.-W. Rix, A. Borch, S. Dye, M. Kleinheinrich, L. Wisotzki, and D. H. McIntosh. Nearly 5000 Distant Early-Type Galaxies in COMBO-17: A Red Sequence and Its Evolution since z˜1. *ApJ*, 608:752–767, June 2004. doi: 10.1086/420778. 18

Y. Benjamini and Y Hochberg. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, 57:289, 1995. 32, 33, 71

Y Benjamini and D Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165–1188, 2001. 33

A. J. Benson. Galaxy formation theory. *Phys. Rep.*, 495:33–86, October 2010. doi: 10.1016/j.physrep.2010.06.001. 50

H. A. Bethe and R. E. Marshak. The physics of stellar interiors and stellar evolution. *Reports on Progress in Physics*, 6:1–15, January 1939. doi: 10.1088/0034-SeminaireBourbaki.4885/6/1/301. 23

L. Bianchi, A. Szalay, C. Martin, P. Friedman, B. Madore, B. Milliard, and R. Malina. The Galaxy Evolution Explorer. In *American Astronomical Society Meeting Abstracts #190*, volume 29 of *Bulletin of the American Astronomical Society*, page 790, May 1997. 93

S. Bianchi, R. Maiolino, and G. Risaliti. AGN Obscuration and the Unified Model. *Advances in Astronomy*, 2012:782030, 2012. doi: 10.1155/2012/782030. 17

Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995. ISBN 0198538642. 58

A. S. Bolton, D. J. Schlegel, É. Aubourg, S. Bailey, V. Bhardwaj, J. R. Brownstein, S. Burles, Y.-M. Chen, K. Dawson, D. J. Eisenstein, J. E. Gunn, G. R. Knapp, C. P. Loomis, R. H. Lupton, C. Maraston, D. Muna, A. D. Myers, M. D. Olmstead, N. Padmanabhan, I. Pâris, W. J. Percival, P. Petitjean, C. M. Rockosi, N. P. Ross, D. P. Schneider, Y. Shu, M. A. Strauss, D. Thomas, C. A. Tremonti, D. A. Wake, B. A. Weaver, and W. M. Wood-Vasey. Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astronomical Journal*, 144: 144, November 2012. doi: 10.1088/0004-SeminaireBourbaki.6256/144/5/144. 78

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344: 1000–1028, October 2003a. doi: 10.1046/j.1365-SeminaireBourbaki.8711.2003.06897.x. 54

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344:1000–1028, October 2003b. doi: 10.1046/j.1365-SeminaireBourbaki.8711.2003.06897.x. 28

D. Calzetti. Reddening and Star Formation in Starburst Galaxies. *AJ*, 113:162–184, January 1997. doi: 10.1086/118242. 27

D. Calzetti, A. L. Kinney, and T. Storchi-Bergmann. Dust extinction of the stellar continua in starburst galaxies: The ultraviolet and optical extinction law. *ApJ*, 429:582–601, July 1994. doi: 10.1086/174346. 28

D. Calzetti, A. L. Kinney, and T. Storchi-Bergmann. Dust Obscuration in Starburst Galaxies from Near-Infrared Spectroscopy. *ApJ*, 458:132, February 1996. doi: 10.1086/176797. 27

D. Calzetti, L. Armus, R. C. Bohlin, A. L. Kinney, J. Koornneef, and T. Storchi-Bergmann. The Dust Content and Opacity of Actively Star-forming Galaxies. *ApJ*, 533:682–695, April 2000. doi: 10.1086/308692. 54

P. L. Capak. High Redshift COSMOS (z > 4). In *American Astronomical Society Meeting Abstracts #214*, volume 214 of *American Astronomical Society Meeting Abstracts*, page #200.06, May 2009. 54, 77

S. M. Carroll. The Cosmological Constant. *Living Reviews in Relativity*, 4:1, February 2001. doi: 10.12942/lrr-SeminaireBourbaki.2001-SeminaireBourbaki.1. 15

S. Charlot and A. G. Bruzual. Stellar population synthesis revisited. *ApJ*, 367:126–140, January 1991. doi: 10.1086/169608. 49

D. Clowe, A. Gonzalez, and M. Markevitch. Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter. *ApJ*, 604:596–603, April 2004. doi: 10.1086/381970. 13

G. D. Coleman, C.-C. Wu, and D. W. Weedman. Colors and magnitudes predicted for high redshift galaxies. *Astrophysical Journal Supplement Series*, 43:393–416, July 1980. doi: 10.1086/190674. XI, 28, 54, 55, 56

A. A. Collister and O. Lahav. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *Publications of the Astronomical Society of the Pacific*, 116:345–351, April 2004. doi: 10.1086/383254. 57

A. J. Connolly, I. Csabai, A. S. Szalay, D. C. Koo, R. G. Kron, and J. A. Munn. Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *Astronomical Journal*, 110:2655, December 1995. doi: 10.1086/117720. 21, 28

A. J. Connolly, A. S. Szalay, M. Dickinson, M. U. Subbarao, and R. J. Brunner. The Evolution of the Global Star Formation History as Measured from the Hubble Deep Field. *ApJ*, 486:L11, September 1997. doi: 10.1086/310829. 20

A. J. Connolly, T. Budavári, A. S. Szalay, I. Csabai, and R. J. Brunner. An Orthogonal Approach to Photometric Redshifts. In R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, editor, *Photometric Redshifts and the Detection of High Redshift Galaxies*, volume 191 of *Astronomical Society of the Pacific Conference Series*, pages 13–18, 1999. 21

C. Conroy, J. E. Gunn, and M. White. The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *ApJ*, 699:486–506, July 2009. doi: 10.1088/0004-SeminaireBourbaki.637X/699/1/486. 51

R. Costero and D. E. Osterbrock. The Optical Spectra of Narrow-Line Radio Galaxies. *Astrophysical Journal*, 211:675–683, February 1977. doi: 10.1086/154977. 65

C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: A Confirmation. *Science*, 124:103–104, July 1956. doi: 10.1126/science.124.3212.103. 12

C. de Jager, S. Duhau, and B. van Geel. Quantifying and specifying the solar influence on terrestrial surface temperature. *Journal of Atmospheric and Solar-Terrestrial Physics*, 72:926–937, August 2010. doi: 10.1016/j.jastp.2010.04.011. X, 37

V. de Lapparent, M. J. Geller, and J. P. Huchra. A slice of the universe. *ApJ*, 302:L1–L5, March 1986. doi: 10.1086/184625. 5

E. D'Onghia, M. Vogelsberger, and L. Hernquist. Self-perpetuating Spiral Arms in Disk Galaxies. *ApJ*, 766:34, March 2013. doi: 10.1088/0004-SeminaireBourbaki.637X/766/1/34. 17

C. Doppler. *Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels*. 1842. 4

A. Dressler. Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, 236:351–365, March 1980. doi: 10.1086/157753. 17, 51

A. Dressler, A. Oemler, Jr., H. R. Butcher, and J. E. Gunn. The morphology of distant cluster galaxies. 1: HST observations of CL 0939+4713. *ApJ*, 430:107–120, July 1994. doi: 10.1086/174386. 17

M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, K. Glazebrook, R. Sharp, M. B. Pracy, S. Brough, M. Colless, W. J. Couch, S. M. Croom, T. M. Davis, D. Forbes, K. Forster, D. G. Gilbank, M. Gladders, B. Jelliffe, N. Jones, I.-H. Li, B. Madore, D. C. Martin, G. B. Poole, T. Small, E. Wisnioski, T. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: survey design and first data release. *MNRAS*, 401:1429–1452, January 2010. doi: 10.1111/j.1365-Seminai reBourbaki.2966.2009.15754.x. 90, 93, 94

A. S. Eddington. On the relation between the masses and luminosities of the stars. *MNRAS*, 84: 308–332, March 1924. 25

A. Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322:891–921, 1905. doi: 10.1002/andp.19053221004. 5, 23

A. Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*, pages 844–847, 1915. 5

D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, November 2005. doi: 10.1086/466512. 48

M. J. Fadili and J.-L. Starck. Monotone operator splitting for optimization problems in sparse recovery. In *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt*, pages 1461–1464. IEEE, 2009. 70

M. Fioc and B. Rocca-Volmerange. PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts. *A&A*, 326:950–962, October 1997. 28, 49, 54

M. Fioc and B. Rocca-Volmerange. PEGASE.2, a metallicity-consistent spectral evolution model of galaxies: the documentation and the code. *ArXiv 9912179*, December 1999. 28, 52, 54

A. E. Firth, O. Lahav, and R. S. Somerville. Estimating photometric redshifts with artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, 339:1195–1202, March 2003. doi: 10.1046/j.1365-SeminaireBourbaki.8711.2003.06271.x. 58

D. J. Fixsen. The Temperature of the Cosmic Microwave Background. *ApJ*, 707:916–920, December 2009. doi: 10.1088/0004-SeminaireBourbaki.637X/707/2/916. 12

M. Fligge and S. K. Solanki. Noise reduction in astronomical spectra using wavelet packets. *Astronomy & Astrophysics Supplement Series*, 124:579–587, September 1997. doi: 10.1051/aas:1997208. 70

A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, 1922. doi: 10.1007/BF01332580. 8

A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, 21:326–332, December 1924. doi: 10.1007/BF01328280. 8

M. Fukugita, T. Ichikawa, J.Ẽ. Gunn, M. Doi, K. Shimasaku, and D.P̃. Schneider. The Sloan Digital Sky Survey Photometric System. *Astronomical Journal*, 111:1748, April 1996. doi: 10.1086/117915. 21, 48, 78

K. Glazebrook, A. R. Offer, and K. Deeley. Automatic Redshift Determination by Use of Principal Component Analysis. I. Fundamentals. *Astrophysical Journal*, 492:98, January 1998. doi: 10.1086/305039. 28, 59, 60, 62, 65

A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, 23:347–356, January 1981. doi: 10.1103/PhysRevD.23.347. 12

Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910. ISSN 0025-5831. doi: 10.1007/BF01456326. URL http://dx.doi.org/10.1007/BF01456326. 41

E. Hatziminaoglou, G. Mathez, and R. Pelló. Quasar candidate multicolor selection technique: a different approach. *Astronomy & Astrophysics*, 359:9–17, July 2000. 20

T. M. Heckman, L. Armus, and G. K. Miley. On the nature and implications of starburst-driven galactic superwinds. *ApJS*, 74:833–868, December 1990. doi: 10.1086/191522. 26, 51

W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift fur Physik*, 43:172–198, March 1927. doi: 10.1007/BF01397280. 36

H. Hildebrandt, S. Arnouts, P. Capak, L. A. Moustakas, C. Wolf, F. B. Abdalla, R. J. Assef, M. Banerji, N. Benítez, G. B. Brammer, T. Budavári, S. Carliles, D. Coe, T. Dahlen, R. Feldmann, D. Gerdes, B. Gillis, O. Ilbert, R. Kotulla, O. Lahav, I. H. Li, J.-M. Miralles, N. Purger, S. Schmidt, and J. Singal. PHAT: Photo-z Accuracy Testing. *Astronomy & Astrophysics*, 523:A31, November 2010. doi: 10.1051/0004-SeminaireBourbaki.6361/201014885. XI, 55, 56, 62

M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. Wavelets: time-frequency methods and phase space - A real-time algorithm for signal analysis with the help of the wavelet transform. In *Proceedings of the international conference, Marseille, France, December 14-18, 1987*, 1989. 44, 45

A. M. Hopkins, C. J. Miller, A. J. Connolly, C. Genovese, R. C. Nichol, and L. Wasserman. A New Source Detection Algorithm Using the False-Discovery Rate. *Astrophysical Journal*, 123:1086–1094, February 2002. 72

F. Hoyle. The synthesis of the elements from hydrogen. *MNRAS*, 106:343, 1946. 23

G. Huang, H. Jiang, K. Matthews, and P. Wilford. Lensless Imaging by Compressive Sensing. *ArXiv 1305.7181*, May 2013. 44

E. Hubble. No. 304. N.G.C. 6822, a remote stellar system. *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 304:1–25, 1925. 4, 11

E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929. doi: 10.1073/pnas.15.3.168. 4, 7, 9, 11

E. P. Hubble. Extragalactic nebulae. *ApJ*, 64:321–369, December 1926. doi: 10.1086/143018. 16

O. Ilbert, S. Arnouts, H. J. McCracken, M. Bolzonella, E. Bertin, O. Le Fèvre, Y. Mellier, G. Zamorani, R. Pellò, A. Iovino, L. Tresse, V. Le Brun, D. Bottini, B. Garilli, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, G. Vettolani, A. Zanichelli, C. Adami, S. Bardelli, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, O. Cucciati, S. Foucaud, P. Franzetti, I. Gavignaud, L. Guzzo, B. Marano, C. Marinoni, A. Mazure, B. Meneux, R. Merighi, S. Paltani, A. Pollo, L. Pozzetti, M. Radovich, E. Zucca, M. Bondi, A. Bongiorno, G. Busarello, S. de La Torre, L. Gregorini, F. Lamareille, G. Mathez, P. Merluzzi, V. Ripepi, D. Rizzo, and D. Vergani. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *Astronomy & Astrophysics*, 457:841–856, October 2006. doi: 10.1051/0004-SeminaireBourbaki.6361:20065138. 53, 56

O. Ilbert, P. Capak, M. Salvato, H. Aussel, H. J. McCracken, D. B. Sanders, N. Scoville, J. Kartaltepe, S. Arnouts, E. Le Floc'h, B. Mobasher, Y. Taniguchi, F. Lamareille, A. Leauthaud, S. Sasaki, D. Thompson, M. Zamojski, G. Zamorani, S. Bardelli, M. Bolzonella, A. Bongiorno, M. Brusa, K. I. Caputi, C. M. Carollo, T. Contini, R. Cook, G. Coppa, O. Cucciati, S. de la Torre, L. de Ravel, P. Franzetti, B. Garilli, G. Hasinger, A. Iovino, P. Kampczyk, J.-P. Kneib, C. Knobel, K. Kovac, J. F. Le Borgne, V. Le Brun, O. L. Fèvre, S. Lilly, D. Looper, C. Maier, V. Mainieri, Y. Mellier, M. Mignoli, T. Murayama, R. Pellò, Y. Peng, E. Pérez-Montero, A. Renzini, E. Ricciardelli, D. Schiminovich, M. Scodeggio, Y. Shioya, J. Silverman, J. Surace, M. Tanaka, L. Tasca, L. Tresse, D. Vergani, and E. Zucca. Cosmos Photometric Redshifts with 30-Bands for 2-deg^2. *Astrophysical Journal*, 690:1236–1249, January 2009. doi: 10.1088/0004-SeminaireBourbaki.637X/690/2/1236. 54, 77

S. Jouvel, J.-P. Kneib, O. Ilbert, G. Bernstein, S. Arnouts, T. Dahlen, A. Ealet, B. Milliard, H. Aussel, P. Capak, A. Koekemoer, V. Le Brun, H. McCracken, M. Salvato, and N. Scoville. Designing future dark energy space missions. I. Building realistic galaxy spectro-photometric catalogs and their first applications. *Astronomy and Astrophysics*, 504:359–371, September 2009. doi: 10.1051/0004-Sem inaireBourbaki.6361/200911798. 54, 77

G. Kauffmann, S. D. M. White, and B. Guiderdoni. The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. *MNRAS*, 264:201, September 1993. 17

D. Kazanas, K. Fukumura, E. Behar, and I. Contopoulos. Towards a unified AGN structure of accretion powered sources. In *Proceedings of "An INTEGRAL view of the high-energy sky (the*

first 10 years)" - 9th INTEGRAL Workshop and celebration of the 10th anniversary of the launch (INTEGRAL 2012). 15-19 October 2012. Bibliotheque Nationale de France, Paris, France. Published online at <A href="http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=176">http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=176</A>, id.60, 2012. 17

R. C. Kennicutt, Jr. The integrated spectra of nearby galaxies - General properties and emission-line spectra. *ApJ*, 388:310–327, April 1992. doi: 10.1086/171154. 27

R. C. Kennicutt, Jr. Star Formation in Galaxies Along the Hubble Sequence. *ARA&A*, 36:189–232, 1998. doi: 10.1146/annurev.astro.36.1.189. 17, 52, 54

A. L. Kinney, D. Calzetti, R. C. Bohlin, K. McQuade, T. Storchi-Bergmann, and H. R. Schmitt. Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections. *Astrophysical Journal*, 467:38, August 1996. doi: 10.1086/177583. XI, 28, 54, 55, 56

A. T. Koski and D. E. Osterbrock. Electron Temperature in the Elliptical Galaxy NGC 1052. *Astrophysical Journal, Letters*, 203:L49, January 1976. doi: 10.1086/182017. 65

R. Kotulla, U. Fritze, P. Weilbacher, and P. Anders. GALEV evolutionary synthesis models - I. Code, input physics and web interface. *MNRAS*, 396:462–484, June 2009. doi: 10.1111/j.1365-Seminaire Bourbaki.2966.2009.14717.x. 28, 52

P. Kroupa. On the variation of the initial mass function. *MNRAS*, 322:231–246, April 2001a. doi: 10.1046/j.1365-SeminaireBourbaki.8711.2001.04022.x. 26, 50, 51

P. Kroupa. The Local Stellar Initial Mass Function. In S. Deiters, B. Fuchs, A. Just, R. Spurzem, and R. Wielen, editors, *Dynamics of Star Clusters and the Milky Way*, volume 228 of *Astronomical Society of the Pacific Conference Series*, page 187, 2001b. 50

M. R. Krumholz, R. I. Klein, and C. F. McKee. Radiation pressure in massive star formation. In R. Cesaroni, M. Felli, E. Churchwell, and M. Walmsley, editors, *Massive Star Birth: A Crossroads of Astrophysics*, volume 227 of *IAU Symposium*, pages 231–236, 2005. doi: 10.1017/S1743921305004588. 52

G. Lake and R. G. Carlberg. The collapse and formation of galaxies. II - A control parameter for the Hubble sequence. III - The origin of the Hubble sequence. *AJ*, 96:1581–1592, November 1988a. doi: 10.1086/114908. 17

G. Lake and R. G. Carlberg. The Collapse and Formation of Galaxies. III. The Origin of the Hubble Sequence. *AJ*, 96:1587, November 1988b. doi: 10.1086/114909. 17

S. K. Lamoreaux. Demonstration of the Casimir Force in the 0.6 to 6 μm Range. *Physical Review Letters*, 78:5–8, January 1997. doi: 10.1103/PhysRevLett.78.5. 15

D. Le Borgne, B. Rocca-Volmerange, P. Prugniel, A. Lançon, M. Fioc, and C. Soubiran. Evolutionary synthesis of galaxies at high spectral resolution with the code PEGASE-HR. Metallicity and age tracers. *A&A*, 425:881–897, October 2004. doi: 10.1051/0004-SeminaireBourbaki.6361:200400044. 28, 54

T. Lejeune, F. Cuisinier, and R. Buser. Standard stellar library for evolutionary synthesis. I. Calibration of theoretical spectra. *A&AS*, 125:229–246, October 1997. doi: 10.1051/aas:1997373. 54

T. Lejeune, F. Cuisinier, and R. Buser. A standard stellar library for evolutionary synthesis. II. The M dwarf extension. *A&AS*, 130:65–75, May 1998. doi: 10.1051/aas:1998405. 54

A. G. Lemaître. Contributions to a British Association Discussion on the Evolution of the Universe. *Nature*, 128:704–706, October 1931. doi: 10.1038/128704a0. 11

G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Societe Scietifique de Bruxelles*, 47: 49–59, 1927. 8

M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein, B. Flaugher, K. Honscheid, R. Kron, O. Lahav, P. McDonald, N. Roe, D. Schlegel, and representing the DESI collaboration. The DESI Experiment, a whitepaper for Snowmass 2013. *ArXiv e-prints*, August 2013. 47, 48, 78, 86

J. Loveday. The APM Bright Galaxy Catalogue. *MNRAS*, 278:1025–1048, February 1996. 16

R. Lutz, S. Schuh, R. Silvotti, S. Dreizler, E. M. Green, G. Fontaine, T. Stahn, S. D. Hügelmeyer, and T.-O. Husser. Light Curve Analysis of the Hybrid SdB PulsatorsHS 0702+6043 and HS 2201+2610. In U. Heber, C. S. Jeffery, and R. Napiwotzki, editors, *Hot Subdwarf Stars and Related Objects*, volume 392 of *Astronomical Society of the Pacific Conference Series*, page 339, 2008. 70

D. Lynden-Bell. Galactic Nuclei as Collapsed Old Quasars. *Nature*, 223:690–694, August 1969. doi: 10.1038/223690a0. 17

C. J. MacDonald and G. Bernstein. Photometric Redshift Biases from Galaxy Evolution. *Publications of the Astronomical Society of the Pacific*, 122:485–489, April 2010. doi: 10.1086/651702. 55

D. P. Machado, A. Leonard, J.-L. Starck, F. B. Abdalla, and S. Jouvel. Darth Fader: Using wavelets to obtain accurate redshifts of spectra at very low signal-to-noise. *ArXiv e-prints*, September 2013. 62, 63, 94

D. C. Martin and GALEX Science Team. Constraints on Evolution from the Blue to Red Sequence using GALEX and SDSS. In *American Astronomical Society Meeting Abstracts*, volume 37 of *Bulletin of the American Astronomical Society*, page 1235, December 2005. 93

A. I. Merson, C. M. Baugh, J. C. Helly, V. Gonzalez-Perez, S. Cole, R. Bielby, P. Norberg, C. S. Frenk, A. J. Benson, R. G. Bower, C. G. Lacey, and C. d. P. Lagos. Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection. *MNRAS*, 429:556–578, February 2013. doi: 10.1093/mnras/sts355. 49

A. A. Michelson and E. W. Morley. On the Relative Motion of the Earth and of the Luminiferous Ether. *Sidereal Messenger, vol. 6, pp.306-310*, 6:306–310, November 1887. 5

C. J. Miller, C. Genovese, R. C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore. Controlling the False-Discovery Rate in Astrophysical Data Analysis. *Astrophysical Journal*, 122:3492–3505, December 2001. 32, 72

H. Minkowski. Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 53–111, 1907. 6

W. W. Morgan and P. C. Keenan. Spectral Classification. *ARA&A*, 11:29, 1973. doi: 10.1146/annurev.aa.11.090173.000333. 24

W. W. Morgan, P. C. Keenan, and E. Kellman. *An atlas of stellar spectra, with an outline of spectral classification*. 1943. 24

H. Nyquist. Certain topics in telegraph transmission theory. *Proceedings of the IEEE*, 90(2):280–305, 2002. doi: 10.1109/5.989875. URL http://dx.doi.org/10.1109/5.989875. 43

P. Panuzzo, O. Vega, A. Bressan, L. Buson, M. Clemens, R. Rampazzo, L. Silva, J. R. Valdés, G. L. Granato, and L. Danese. The Star Formation History of the Virgo Early-Type Galaxy NGC 4435: The Spitzer Mid-Infrared View. *Astrophysical Journal*, 656:206–216, February 2007. doi: 10.1086/510147. IX, 16, 65

D. Parkinson, S. Riemer-Sørensen, C. Blake, G. B. Poole, T. M. Davis, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-h. Li, B. Madore, D. C. Martin, K. Pimbblet, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: Final data release and cosmological results. *Phys. Rev. D*, 86(10):103518, November 2012. doi: 10.1103/PhysRevD.86.103518. 90, 93

A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965. doi: 10.1086/148307. 12

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, June 1999. doi: 10.1086/307221. 14

E. C. Pickering. The Draper Catalogue of stellar spectra photographed with the 8-inch Bache telescope as a part of the Henry Draper memorial. *Annals of Harvard College Observatory*, 27:1–388, 1890. 23

S. Pires, J. B. Juin, D. Yvon, Y. Moudden, S. Anthoine, and E. Pierpaoli. Sunyaev-Zel'dovich cluster reconstruction in multiband bolometer camera surveys. *Astronomy & Astrophysics*, 455, August 2006. 741-755. 72

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. I. Overview of products and scientific results. *ArXiv e-prints*, March 2013a. 13

Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. *ArXiv e-prints*, March 2013b. 13, 14

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974. doi: 10.1086/152650. 17

P. Prugniel and C. Soubiran. A database of high and medium-resolution stellar spectra. *A&A*, 369: 1048–1057, April 2001. doi: 10.1051/0004-SeminaireBourbaki.6361:20010163. 54

A. Refregier, A. Amara, T. D. Kitching, A. Rassat, R. Scaramella, J. Weller, and f. t. Euclid Imaging Consortium. Euclid Imaging Consortium Science Book. *ArXiv: 1001.0061*, January 2010. 102

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116: 1009–1038, September 1998. doi: 10.1086/300499. 14

H. P. Robertson. Kinematics and World-Structure. *ApJ*, 82:284, November 1935. doi: 10.1086/143681. 8

H. P. Robertson. Kinematics and World-Structure II. *ApJ*, 83:187, April 1936a. doi: 10.1086/143716. 8

H. P. Robertson. Kinematics and World-Structure III. *ApJ*, 83:257, May 1936b. doi: 10.1086/143726. 8

T. P. Robitaille and B. A. Whitney. The Present-Day Star Formation Rate of the Milky Way Determined from Spitzer-Detected Young Stellar Objects. *ApJ*, 710:L11–L15, February 2010. doi: 10.1088/2041-SeminaireBourbaki.8205/710/1/L11. 52

V. C. Rubin and W. K. Ford, Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, 159:379, February 1970. doi: 10.1086/150317. 13

V. C. Rubin, W. K. J. Ford, and N. . Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *ApJ*, 238:471–487, June 1980. doi: 10.1086/158003. 13

E. E. Salpeter. The Luminosity Function and Stellar Evolution. *ApJ*, 121:161, January 1955. doi: 10.1086/145971. 26, 50, 51

J. Scalo, E. Vazquez-Semadeni, D. Chappell, and T. Passot. On the Probability Density Function of Galactic Gas. I. Numerical Simulations and the Significance of the Polytropic Index. *ApJ*, 504:835, September 1998. doi: 10.1086/306099. 51

D. Schlegel, F. Abdalla, T. Abraham, C. Ahn, C. Allende Prieto, J. Annis, E. Aubourg, M. Azzaro, S. B. C. Baltay, C. Baugh, C. Bebek, S. Becerril, M. Blanton, A. Bolton, B. Bromley, R. Cahn, P. . Carton, J. L. Cervantes-Cota, Y. Chu, M. Cortes, K. Dawson, A. Dey, M. Dickinson, H. T. Diehl, P. Doel, A. Ealet, J. Edelstein, D. Eppelle, S. Escoffier, A. Evrard, L. Faccioli, C. Frenk, M. Geha, D. Gerdes, P. Gondolo, A. Gonzalez-Arroyo, B. Grossan, T. Heckman, H. Heetderks, S. Ho, K. Honscheid, D. Huterer, O. Ilbert, I. Ivans, P. Jelinsky, Y. Jing, D. Joyce, R. Kennedy, S. Kent, D. Kieda, A. Kim, C. Kim, J. . Kneib, X. Kong, A. Kosowsky, K. Krishnan, O. Lahav, M. Lampton, S. LeBohec, V. Le Brun, M. Levi, C. Li, M. Liang, H. Lim, W. Lin, E. Linder, W. Lorenzon, A. de la Macorra, C. Magneville, R. Malina, C. Marinoni, V. Martinez, S. Majewski, T. Matheson, R. McCloskey, P. McDonald, T. McKay, J. McMahon, B. Menard, J. Miralda-Escude, M. Modjaz, A. Montero-Dorta, I. Morales, N. Mostek, J. Newman, R. Nichol, P. Nugent, K. Olsen, N. Padmanabhan, N. Palanque-Delabrouille, I. Park, J. Peacock, W. Percival, S. Perlmutter, C. Peroux, P. Petitjean, F. Prada, E. Prieto, J. Prochaska, K. Reil, C. Rockosi, N. Roe, E. Rollinde, A. Roodman, N. Ross, G. Rudnick, V. Ruhlmann-Kleider, J. Sanchez, D. Sawyer, C. Schimd, M. Schubnell, R. Scoccimaro, U. Seljak, H. Seo, E. Sheldon, M. Sholl, R. Shulte-Ladbeck, A. Slosar, D. S. Smith, G. Smoot, W. Springer, A. Stril, A. S. Szalay, C. Tao, G. Tarle, E. Taylor, A. Tilquin, J. Tinker, F. Valdes, J. Wang, T. Wang, B. A. Weaver, D. Weinberg, M. White, M. Wood-Vasey, J. Yang,

X. Y. C. Yeche, N. Zakamska, A. Zentner, C. Zhai, and P. Zhang. The BigBOSS Experiment. *ArXiv e-prints*, arXiv:1106.1706, June 2011. 47, 78, 86

A. Secchi. *Le stelle : saggio di astronomia siderale.* 1877. 4, 19, 23

Ivan W. Selesnick. Wavelets, a modern tool for signal processing. *Physics Today*, 60(10):78–79, 2007. URL http://eeweb.poly.edu/iselesni/pubs/WaveletQuickStudy_expanded.pdf. 42

C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949. ISSN 0096-8390. doi: 10.1109/jrproc.1949.232969. URL http://dx.doi.org/10.1109/jrproc.1949.232969. 43

M. Shensa. The discrete wavelet transform: wedding the a trous and mallat algorithms. *Signal Processing, IEEE Transactions on*, 40(10):2464–2482, 1992. ISSN 1053-587X. doi: 10.1109/78.157290. 44

Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi. The jpeg 2000 still image compression standard. *IEEE Signal processing Magazine*, 18:36–58, 2001. 44

V. M. Slipher. Spectrographic Observations of Nebulae. *Popular Astronomy*, 23:21–24, January 1915. 4, 11, 19

S. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, C. M. Rockosi, M. A. Carr, F. Leger, K. S. Dawson, M. D. Olmstead, J. Brinkmann, R. Owen, R. H. Barkhouser, K. Honscheid, P. Harding, D. Long, R. H. Lupton, C. Loomis, L. Anderson, J. Annis, M. Bernardi, V. Bhardwaj, D. Bizyaev, A. S. Bolton, H. Brewington, J. W. Briggs, S. Burles, J. G. Burns, F. Castander, A. Connolly, J. R. Davenport, G. Ebelke, H. Epps, P. D. Feldman, S. Friedman, J. Frieman, T. Heckman, C. L. Hull, G. R. Knapp, D. M. Lawrence, J. Loveday, E. J. Mannery, E. Malanushenko, V. Malanushenko, A. Merrelli, D. Muna, P. Newman, R. C. Nichol, D. Oravetz, K. Pan, A. C. Pope, P. G. Ricketts, A. Shelden, D. Sandford, W. Siegmund, A. Simmons, D. Smith, S. Snedden, D. P. Schneider, M. Strauss, M. SubbaRao, C. Tremonti, P. Waddell, and D. G. York. The Multi-Object, Fiber-Fed Spectrographs for SDSS and the Baryon Oscillation Spectroscopic Survey. *ArXiv e-prints*, arXiv:1208.2233, August 2012. 77

M. J. Sparnaay. Attractive Forces between Flat Plates. *Nature*, 180:334–335, August 1957. doi: 10.1038/180334b0. 15

V. Springel, T. Di Matteo, and L. Hernquist. Black Holes in Galaxy Mergers: The Formation of Red Elliptical Galaxies. *ApJ*, 620:L79–L82, February 2005a. doi: 10.1086/428772. 17

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005b. doi: 10.1038/nature03597. 15, 49

J.-L. Starck and F. Murtagh. Image restoration with noise suppression using the wavelet transform. *Astronomy & Astrophysics*, 288:342–348, August 1994. 68

J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis.* Springer, 2006. 2nd edn. 66, 68, 71, 72

J.-L. Starck, A. Bijaoui, and F. Murtagh. Multiresolution support applied to image filtering and deconvolution. *CVGIP: Graphical Models and Image Processing*, 57, 1995. 420–431. 71

J.-L. Starck, A. Claret, and R. Siebenmorgen. ISOCAM data calibration. Technical report, CEA, 1996a. 66

J.-L. Starck, F. Murtagh, B. Pirenne, and M. Albrecht. Astronomical image compression based on noise suppression. *Publications of the Astronomical Society of the Pacific*, 108, 1996b. 446–455. 66

J-L Starck, J. Fadili, and F. Murtagh. The undecimated wavelet decomposition and its reconstruction. *Image Processing, IEEE Transactions on*, 16(2):297–309, 2007. ISSN 1057-7149. doi: 10.1109/TIP.2006.887733. 44

J.-L. Starck, F. Murtagh, and M.J. Fadili. *Sparse Image and Signal Processing*. Cambridge University Press, 2010. 44, 67, 68, 70, 72

C. C. Steidel, A. E. Shapley, M. Pettini, K. L. Adelberger, D. K. Erb, N. A. Reddy, and M. P. Hunt. A Survey of Star-forming Galaxies in the $1.4 \lesssim z \lesssim 2.5$ Redshift Desert: Overview. *ApJ*, 604: 534–550, April 2004. doi: 10.1086/381960. 19

C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles, F. J. Castander, A. J. Connolly, D. J. Eisenstein, J. A. Frieman, G. S. Hennessy, R. B. Hindsley, Ž. Ivezić, S. Kent, P. Z. Kunszt, B. C. Lee, A. Meiksin, J. A. Munn, H. J. Newberg, R. C. Nichol, T. Nicinski, J. R. Pier, G. T. Richards, M. W. Richmond, D. J. Schlegel, J. A. Smith, M. A. Strauss, M. SubbaRao, A. S. Szalay, A. R. Thakar, D. L. Tucker, D. E. Vanden Berk, B. Yanny, J. K. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, M. Bartelmann, S. Bastian, A. Bauer, E. Berman, H. Böhringer, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, L. Carey, M. A. Carr, B. Chen, D. Christian, P. L. Colestock, J. H. Crocker, I. Csabai, P. C. Czarapata, J. Dalcanton, A. F. Davidsen, J. E. Davis, W. Dehnen, S. Dodelson, M. Doi, T. Dombeck, M. Donahue, N. Ellman, B. R. Elms, M. L. Evans, L. Eyer, X. Fan, G. R. Federwitz, S. Friedman, M. Fukugita, R. Gal, B. Gillespie, K. Glazebrook, J. Gray, E. K. Grebel, B. Greenawalt, G. Greene, J. E. Gunn, E. de Haas, Z. Haiman, M. Haldeman, P. B. Hall, M. Hamabe, B. Hansen, F. H. Harris, H. Harris, M. Harvanek, S. L. Hawley, J. J. E. Hayes, T. M. Heckman, A. Helmi, A. Henden, C. J. Hogan, D. W. Hogg, D. J. Holmgren, J. Holtzman, C.-H. Huang, C. Hull, S.-I. Ichikawa, T. Ichikawa, D. E. Johnston, G. Kauffmann, R. S. J. Kim, T. Kimball, E. Kinney, M. Klaene, S. J. Kleinman, A. Klypin, G. R. Knapp, J. Korienek, J. Krolik, R. G. Kron, J. Krzesiński, D. Q. Lamb, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, B. McLean, K. Menou, A. Merelli, H. J. Mo, D. G. Monet, O. Nakamura, V. K. Narayanan, T. Nash, E. H. Neilsen, Jr., P. R. Newman, A. Nitta, M. Odenkirchen, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. S. Peterson, D. Petravick, A. Pope, R. Pordes, M. Postman, A. Prosapio, T. R. Quinn, R. Rechenmacher, C. H. Rivetta, H.-W. Rix, C. M. Rockosi, R. Rosner, K. Ruthmansdorfer, D. Sandford, D. P. Schneider, R. Scranton, M. Sekiguchi, G. Sergey, R. Sheth, K. Shimasaku, S. Smee, S. A. Snedden, A. Stebbins, C. Stubbs, I. Szapudi, P. Szkody, G. P. Szokoly, S. Tabachnik, Z. Tsvetanov, A. Uomoto, M. S. Vogeley, W. Voges, P. Waddell, R. Walterbos, S.-i. Wang, M. Watanabe, D. H. Weinberg, R. L. White, S. D. M. White, B. Wilhite, D. Wolfe, N. Yasuda, D. G. York, I. Zehavi, and W. Zheng. Sloan Digital Sky Survey: Early Data Release. *Astronomical Journal*, 123:485–548, January 2002. doi: 10.1086/324741. 47, 65, 70

M. A. Strauss, D. H. Weinberg, R. H. Lupton, V. K. Narayanan, J. Annis, M. Bernardi, M. Blanton, S. Burles, A. J. Connolly, J. Dalcanton, M. Doi, D. Eisenstein, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. S. J. Kim, G. R. Knapp, R. G. Kron, J. A. Munn, H. J. Newberg, R. C. Nichol, S. Okamura, T. R. Quinn, M. W. Richmond, D. J. Schlegel, K. Shimasaku, M. SubbaRao, A. S. Szalay, D. Vanden Berk, M. S. Vogeley, B. Yanny, N. Yasuda, D. G. York, and I. Zehavi. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *Astronomical Journal*, 124:1810–1824, September 2002. doi: 10.1086/342343. 78, 90

J. Tonry and M. Davis. A survey of galaxy redshifts. I - Data reduction techniques. *Astronomical Journal*, 84:1511–1525, October 1979. doi: 10.1086/112569. 65

A. Toomre. Theories of spiral structure. *ARA&A*, 15:437–478, 1977. doi: 10.1146/annurev.aa.15.090177.002253. 17

C. M. Urry and P. Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *PASP*, 107:803, September 1995. doi: 10.1086/133630. 17

P. G. van Dokkum. Evidence of Cosmic Evolution of the Stellar Initial Mass Function. *ApJ*, 674:29–50, February 2008. doi: 10.1086/525014. 51

A. G. Walker. On the formal comparison of Milne's kinematical system with the systems of general relativity. *MNRAS*, 95:263–269, January 1935. 8

J. A. Wheeler and K. Ford. *Geons, black holes and quantum foam : a life in physics*. 1998. 7

S. D. M. White and C. S. Frenk. Galaxy formation through hierarchical clustering. *ApJ*, 379:52–79, September 1991. doi: 10.1086/170483. 17

A. N. Witt, H. A. Thronson, Jr., and J. M. Capuano, Jr. Dust and the transfer of stellar radiation within galaxies. *ApJ*, 393:611–630, July 1992. doi: 10.1086/171530. 27

Isao Yamada. The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In D. Butnariu, Y. Censor, and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*. Elsevier, 2001. 70

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R.

Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, N. Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, September 2000. doi: 10.1086/301513. 47

I. B. Zeldovich, J. Einasto, and S. F. Shandarin. Giant voids in the universe. *Nature*, 300:407–413, December 1982. doi: 10.1038/300407a0. 5

J. Zoubian and J.-P. Kneib. Designing Future Dark Energy Space Mission: III. New calibrations with ZCOSMOS and luminosity functions. in prep., 2013. 77

F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. *ApJ*, 86:217, October 1937. doi: 10.1086/143864. 13