UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

Identificação e redshifts fotométricos para quasares do tipo-I com sistemas de filtros de bandas médias e estreitas

Carolina Queiroz de Abreu Silva

Orientador: Dr. Luis Raul Weber Abramo

Dissertação de mestrado apresentada ao
Instituto de Física da Universidade de São Paulo
para a obtenção do título de Mestre em Ciências

Banca examinadora:
Prof. Dr. Luis Raul Weber Abramo (IFUSP)
Prof. Dr. Marcos V. B. Teixeira Lima (IFUSP)
Prof. Dr. Claudia L. M. Oliveira (IAGUSP)

São Paulo
2015

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação do Instituto de Física da Universidade de São Paulo

Silva, Carolina Queiroz de Abreu

Identificação e redshifts fotométricos para quasares do tipo-I com sistemas de filtros de bandas médias e estreitas. São Paulo, 2015.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Física. Departamento Física Matemática.

Orientador: Prof. Dr. Luis Raul Weber Abramo

Área de Concentração: Quasares.

Unitermos: 1.Cosmologia; 2. AGNs; 3. Quasares; 4. J-PAS; 5. Estatística Bayesiana.

USP/IF/SBI-120/2015

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA

Identification and photometric redshifts for type-I quasars with medium- and narrow-band filter surveys

Carolina Queiroz de Abreu Silva

Advisor: Dr. Luis Raul Weber Abramo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Science at University of São Paulo

Examination board:
Prof. Dr. Luis Raul Weber Abramo (IFUSP)
Prof. Dr. Marcos V. B. Teixeira Lima (IFUSP)
Profa. Dra. Claudia L. M. Oliveira (IAGUSP)

São Paulo
2015

To my loving mother

“Twinkle, twinkle, quasi-star,

Biggest puzzle from afar.

How unlike the other ones,

Brighter than a trillion Suns.

Twinkle, twinkle, quasi-star,

How I wonder what you are!”

—George Gamow

Resumo

Quasares são objetos valiosos para diversas aplicações cosmológicas. Em particular, eles podem ser usados para localizar alguns dos halos mais massivos e suas luminosidades intrinsecamente elevadas permitem que eles sejam detectados a altos redshifts. Isso implica que quasares (ou núcleos ativos de galáxias, de um modo geral) possuem um grande potencial para mapear a estrutura em larga escala. Entretanto, esse potencial ainda não foi completamente atingido, porque instrumentos que se baseiam no imageamento por bandas largas para pré-selecionar alvos espectroscópicos perdem a maioria dos quasares e, consequentemente, não são capazes de separar adequadamente quasares com linhas de emissão largas de outras fontes pontuais (como estrelas e galáxias de baixa resolução). Esse trabalho é uma tentativa inicial de investigar os ganhos reais na identificação e separação de quasares e estrelas quando são usados filtros de bandas médias e estreitas. A principal novidade desse método é o uso de priors Bayesianos tanto para a distribuição angular de estrelas de diferentes tipos no céu quanto para a distribuição de quasares como função do redshift. Como a evidência desses priors é uma convolução entre a dependência angular das estrelas e a dependência em redshift dos quasares, isso permite que a degenerescência entre esses objetos seja levada em consideração. Entretanto, nossos resultados ainda são inconclusivos para quantificar a eficiência da separação entre estrelas e quasares utilizando esse método e, portanto, alguns refinamentos críticos são necessários.

Abstract

Quasars are valuable sources for several cosmological applications. In particular, they can be used to trace some of the heaviest halos and their high intrinsic luminosities allow them to be detected at high redshifts. This implies that quasars (or active galactic nuclei, in a more general sense) have a huge potential to map the large-scale structure. However, this potential has not yet been fully realized, because instruments which rely on broad-band imaging to pre-select spectroscopic targets usually miss most quasars and, consequently, are not able to properly separate broad-line emitting quasars from other point-like sources (such as stars and low resolution galaxies). This work is an initial attempt to investigate the realistic gains on the identification and separation of quasars and stars when medium- and narrow-band filters in the optical are employed. The main novelty of our approach is the use of Bayesian priors both for the angular distribution of stars of different types on the sky and for the distribution of quasars as a function of redshift. Since the evidence from these priors convolve the angular dependence of stars with the redshift dependence of quasars, this allows us to control for the near degeneracy between these objects. However, our results are inconclusive to quantify the efficiency of star-quasar separation by using this approach and, hence, some critical refinements and improvements are still necessary.

Acknowledgements

I would like to acknowledge the support and assistance from all of those who have contributed to this work or have helped me to keep the motivation during the process. I am very grateful to all of you.

I would like to acknowledge my advisor, Dr. Luis Raul Abramo, for his guidance, caring, patience, for proving me an excellent atmosphere and great ideas for developing the research project, and for the opportunity of internships at Princeton University and at CEFCA, in Spain, as well as of working with J-PAS. I would like to thank Dr. Michael Strauss for advising me at Princeton and for providing the stellar spectra from the SDSS Everything Survey; Dr. Silvia Bonoli, who has provided me with a very productive working environment at CEFCA; Dr. Txitxo Benítez, for providing the BPZ code and for the initial guidance on the technical details about the code. I would like to thank especially Dr. Alberto Molino, for all the assistance and fruitful discussions about including quasars into BPZ, which were very enlightening, and for the useful advice about using the available ALHAMBRA data.

Secondly, I would like to thank: Dr. Teresa Villegas, for helping us to understand the contamination between quasars and stars and for presenting the Besançon Model of the stellar population of the Milky Way; Dr. Jose Ignacio Gonzalez, from Instituto de Física de Cantabria (IFCA), for receiving me in Santander and providing valuable information about type-II quasars, which will be very useful in our next steps; Dr. Paula Coelho, for the discussions about the galactic stellar population and presenting the TRILEGAL model; and Dr. Jesus Varella, for providing the latest version of BPZ. Several other people, who are also interested in the issues treated by this project, have also contributed for the development of this work.

I would also like to acknowledge the research foundations for the indispensable financial grants: *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq); *Fundação de Amparo à Pesquisa do Estado de São Paulo*, for the Master's fellowship (2012/25503-2) and the BEPE fellowship

(2014/21148-9), which made my internship in Spain possible; and the USP-Princeton bilateral agreement.

I acknowledge the contribution of the University of São Paulo, the *Instituto de Física* and the *Departamento de Física Matemática* for the infrastructure and facilities, the internationalized academic environment and the prestigious faculty. In particular, I would like to acknowledge all of the professors who have contributed to my academic development.

I would also like to extend my thanks to my colleagues, especially Arthur Loureiro, Lucas Secco and Hugo Camacho, for the assistance with numerical programming, and Jonás Montero, for the assistance with the LePhare code and for proving some quasar spectra from ALHAMBRA.

I would like to thank my family, for always being comprehensive and supportive, and my cat, for providing me with the so needed relaxing moments.

Finally, I would like to express my gratitude to my parents for their unconditional support and encouragement throughout my studies and work. Their love gives me confidence to follow my choices, their friendship comforts me in the most difficult moments, and their wisdom inspires me to accomplish all of my ambitions.

Contents

Chapter 1

Introduction

In general, the main contribution to the light of normal galaxies is attributed to stars in hydrodynamical equilibrium, with small contributions of gas and dust. However, a small fraction of the galaxies have a broad energy distribution, with emissions that range from the radio wavelengths to the X-rays and even Gamma-rays. This energy release seems to originate from a small central region inside the galaxy, and since evidence shows that most galaxies harbor a black hole in their centers (Kormendy & Richstone 1995), in such active galaxies the central engine is believed to be a (supermassive) black hole of mass $M \sim 10^6$-10^{10} $M_\odot$. These are the so-called active galactic nuclei (AGNs).

AGNs form a family composed of different types of active galaxies, believed to be the same objects viewed from different directions and which differ in their spectral properties. In some AGNs, the radio jets may reach $\gtrsim$ 1 Mpc (1pc = 1 parsec $\sim$ 3.26 ly), and some may present variabilities in their luminosities in time-scales of days or months.

Quasars, also known as quasi-stellar objects (QSOs), are the most luminous members of the class of AGNs. Due to their intrinsic high luminosities (which can exceed the luminosities of “normal” galaxies by a factor of at least a thousand), they can be detected at large distances. Accordingly, quasars constitute a powerful tool for locating high redshift galaxies, providing means of investigating the history and evolution of the galaxies.

Quasars also contribute to the ultraviolet (UV) ionizing background and have a role in the reionization of the Universe. In addition, they are also believed to inhabit the centers of very massive halos (with $M \gtrsim 10^{13}$ $M_\odot$). As a result, quasars can be used to probe the conditions in the early Universe, and consequently, help us map the structures on the largest scales.

However, instruments which rely on broad-band imaging to pre-select spectroscopic targets (e.g., SDSS, PSF) usually miss most quasars and AGNs. In particular, the SDSS photometric sample contains $\gtrsim 10^6$ quasar candidates, but only $\sim 18\%$ of those were spectroscopically confirmed (Richards et al. 2009; Alam et al. 2015). The main reason for this low efficiency is that the fibers in SDSS are only allocated to the brightest, most clearly distinguished quasars, due to the fact that quasars can be confused with stars in color-color diagrams, especially at intermediate redshifts. Thus when the target selection is based on the colors and magnitudes in broad-band filters, the quasar colors cannot be properly distinguished from the colors of stars and, sometimes, of unresolved galaxies, reflecting the redshift-dependent degeneracy between quasars and stars in color-color diagrams.

On the other hand, the construction of a high-purity catalog of quasars, with accurate photometric redshifts, can be much more efficient with medium- or narrow-band filter (FWHM$\sim$100-200 Å) surveys, such as ALHAMBRA (Molino et al. 2014), and the upcoming J-PAS (Benítez et al. 2014) and J-PLUS .

The Lyman break photometric "drop-out" technique as first used for the detection of high redshift quasars, but it was much more finely tuned as a detection technique for high redshift galaxies in the mid-1990s. The technique relies on the large break in the continuum flux from an object that occurs at the 912 Å Lyman limit from neutral hydrogen absorption in the line-of-sight. Multi-band images of a field containing high redshift galaxies can be used to identify those objects that have very red colors as a result of the redshifted Lyman limit falling between any two filters.

This technique was extensively employed by Steidel et al. (1996) to detect $z \sim 3$ objects by their lack of flux in the u-band. The technique was often applied by using three or four filters and defining a region in the two-color plane in which such "drop-outs" were most likely to occur. A key advantage of this technique is that it is essentially free of selection effects, with little contamination from low redshift objects. All high redshift objects above a given magnitude limit will be detected, provided the signal-to-noise ration (S/N) is high enough in all the bands, particularly the bluest band where an upper limit must be established.

Therefore, the u-band dropout technique is efficient to detect quasars at $z \gtrsim 2$. Similarly, employing narrow-band systems is sufficient to resolve the broad emission lines of type-I quasars (as well as most broad absorption line objects), and also to detect the narrow lines of many type-II's and AGNs. In addition, if the properties of extragalactic objects and stellar populations are known *a priori*, then applying Bayesian priors may optimize the identification of these populations.

The goal of this dissertation is to develop tools to separate type-I quasars from stars, and improve the precision with which the photometric redshifts for quasars are estimated. In particular, we study how Bayesian priors can be used to improve the typing and photometric redshift extraction of quasars, within the context of medium- and narrow-band filter surveys. We investigate the efficiency of these tools by characterizing the purity and completeness of the data sample before and after the inclusion of priors. Note that faint and unresolved galaxies also constitute an important contaminating population, however we are not including them in our analysis.

In the following sections we discuss about the main features of quasars and stars.

1.1 Quasars

The observation of strong and very broad emission lines in galaxies dates to the beginning of the twentieth century, but it was in 1963 that the radio galaxies 3C48 and 3C273 were identified as point-like sources by Thomas Matthews, Allan Sandage and Maarten Schmidt (Schneider 2006).

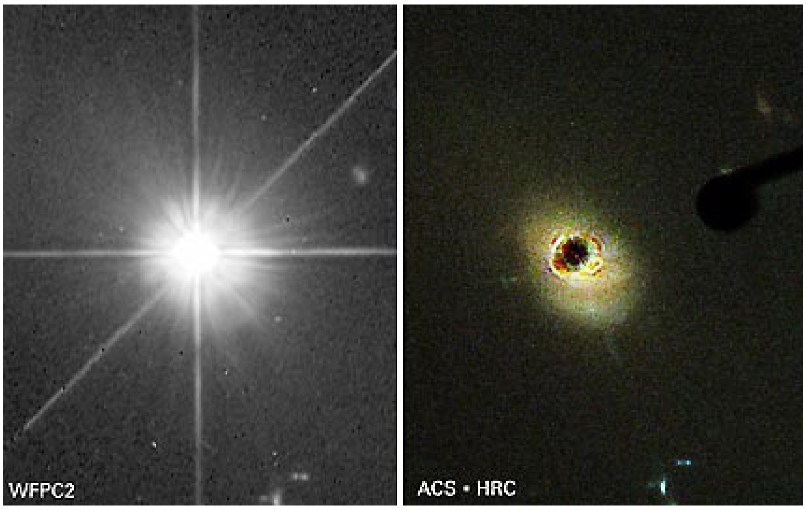


Figure 1.1: Quasar 3C273 at $z = 0.158$, located in an elliptical galaxy in the direction of the constellation of Virgo. This pair of images shows the quasar as the extremely luminous nucleus of an ordinary galaxy. Note the “star-like” appearance on the left panel and the host galaxy with reddened dust lanes on the right panel. Left: Quasar as imaged by Hubble’s Wide Field and Planetary Camera 2 (WFPC2). Right: This image was taken by Hubble’s ACS Camera, using a small occulting disk to block the light coming from the quasar. (Credit: ESA/Hubble & NASA)

Quasars can be included in the unification model scenario for AGNs, which is not complete but is one of the most accepted models nowadays and relates the energy release to the accretion of galactic matter (gas and stars in the vicinity) onto a supermassive black hole ($M{\sim}10^6$-10^{10} $M_\odot$) located at the very center of the galaxy. The accretion of matter forms an accretion disk, which feeds the black hole. This unification model (shown in Fig.1.2) accounts for the several intrinsically similar objects and classifies them according to their appearance, emitted radiation and orientation angle.

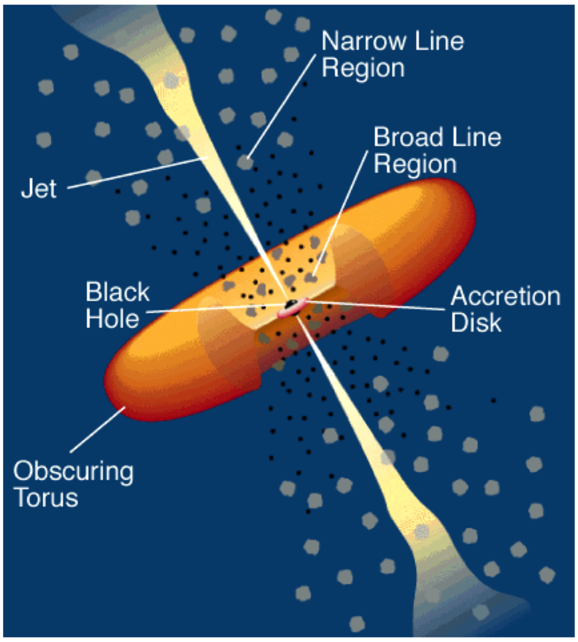


Figure 1.2: The unification model for AGNs (Urry & Padovani 1995).

Centaurus A is the closest AGN to the Milky Way ($\sim$ 3.4 Mpc) and is presumed to be a source of ultrahigh energy cosmic rays—with energies $\gtrsim 10^{19}$ eV (Yuksel et al. 2012).

In the unified model, each member of the AGN family receives a different nomenclature depending on its orientation with respect to the line-of-sight, and this classification could be an evidence of evolutionary relationship between the types. In Fig.1.2 the central region is surrounded by an extended, dusty, molecular toroidal region, involving the broad line region (BLR). Broad emission lines originate close to the torus in clouds orbiting above the disk at high velocity, and are due to the Doppler effect. Another important component is the narrow-line region (NLR), which is further from the torus and has lower density. We can also note two-sided jets of relativistic particles emanating perpendicular to the plane of the accretion disk. The mechanism that originates these jets is not fully understood yet (Torres & Anchordoqui 2004).

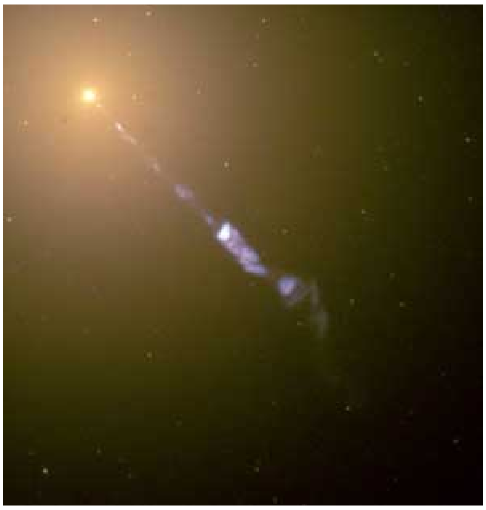


Figure 1.3: Side view from the jet streaming out from the center of the elliptical galaxy M87. This active galaxy is one of the brightest radio sources on the sky. (Credit: NASA and The Hubble Heritage Team)

In the following we investigate how this family is organized. If the axis is close to the line-of-sight, the source is radio-loud. If the jet points directly to the observer, then we see a featureless spectrum produced by synchrotron radiation and this object is called a blazar. BL Lacertae (BL Lac) objects are included in the blazar classification. At small angular offsets from the line of the jets, the broad-line region is hidden by the torus, but the narrow-line is visible, and the corresponding class is a type-II quasar or a Seyfert 2 galaxy (or "edge on" AGN). If the orientation angle is $\sim 30^\circ$, a regular (type-I) quasar or a Seyfert 1 galaxy is observed; in this case, both the broad-line and the narrow-line are visible and there is no obscuration. We say it is a "face-on" AGN. At angles closer to the accretion disk, there is the torus obscuration and no emission line is seen—these are the radio galaxies. Lastly, the source is radio-quiet if the jet is not pointing towards the observer. Fig.1.4 and 1.5 show the main features of a Seyfert 1 and a type-II quasar, respectively.

Quasars are the most luminous members of the class of AGNs: they emit at all wavelengths, from the radio to the X-ray domain of the spectrum. The flux of the source varies at nearly all frequencies, where the variability time scale differs among the objects. If the optical variability occurs at short timescales, the object is referred to as an optically violently variable (OVV) quasar.

The Eddington luminosity (L_{edd}) determines the maximum luminosity in the case of spherical gas accretion onto a central object, i.e., it determines the condition for accretion to occur. This limit can be estimated assuming an equilibrium between the gravitational force on the gas and the radiation pressure force:

$$L_{edd} = \frac{4\pi G c m_p}{\sigma_T} M_* \approx 1.3\text{x}10^{38} \left(\frac{M_*}{M_\odot}\right) \text{ erg s}^{-1} \tag{1.1}$$

where G is the gravitational constant, c the speed of light, m_p the proton mass, σ_T the Thomson scattering cross-section, and M_* the mass of the central object. Typical quasars have luminosities of $\sim 10^{46}$ erg/s.

The quasar luminosity function (see section §3.2) is a very important quantity to characterize these objects, because it describes the spatial density (number per Mpc^3) of quasars per magnitude as a function of the redshift z and of the K-corrected absolute optical magnitude M. In other words: the luminosity function can be used to describe the evolution of the number density of quasars (or AGNs, in a more general sense) with time. It can also be used to describe the formation history of supermassive black holes, to understand the host galaxy evolution and the contribution of quasars to reionization (at $z = 6$).

In general, quasars have a very blue optical spectrum: for instance, most quasars at $z \lesssim 2$ have $U - B < -0.3$. For comparison, only hot white dwarfs have a similar blue color index. This may cause a degeneracy while trying to distinguish quasars and stars in color-color diagrams.

Due to the blue continuum, techniques such as the u-band dropout are efficient to detect quasars at $z \gtrsim 2$. The u-band dropout is based on a strong spectral break at λ= 912 Å, due to the ionization of hydrogen, seen in the bluer filters. Besides the blue continuum, strong and very broad emission lines are characteristic of the optical spectrum, some corresponding to transitions of very high ionized atoms.

The morphology of a quasar consists of two radio lobes which are approximately symmetrical around its optical position. The lobes are frequently connected to the central core by jets, which can reach distances up to 1 Mpc, and are originated on small spatial regions.

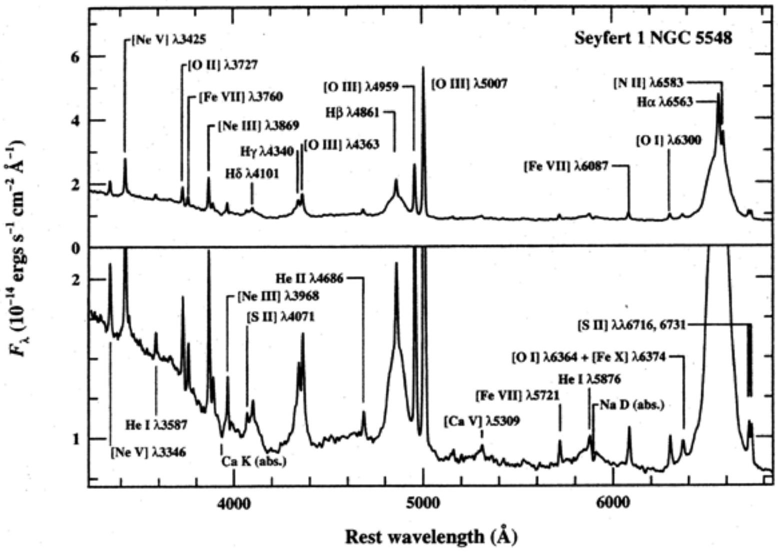


Figure 1.4: Optical spectrum of the Seyfert 1 galaxy NCG 5548 located at $z = 0.017$. The prominent broad and narrow emission lines are shown. The lower panel is an expanded version of the upper panel to facilitate the visualization of the weaker features. (Credit: A. V. Filippenko)

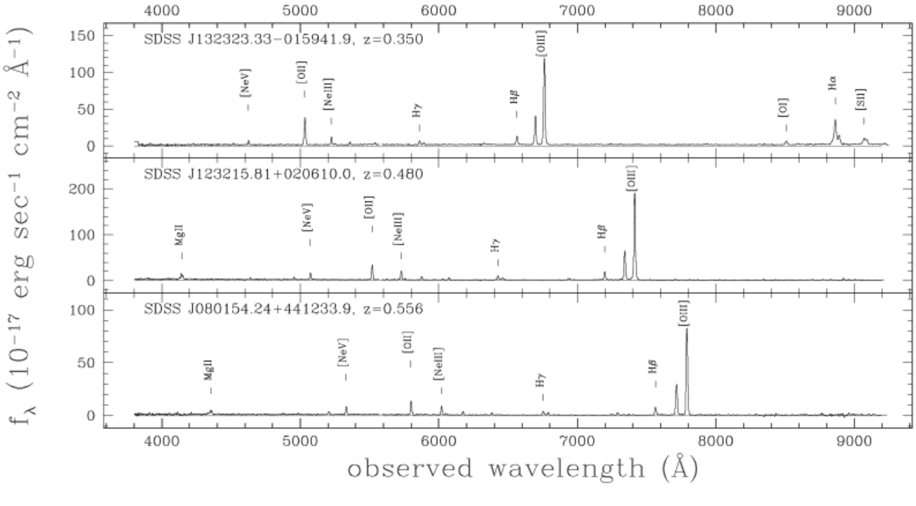


Figure 1.5: Optical spectra of type-II quasar candidates from SDSS (Zakamska et al. 2003).

Emission line intensities and emission line ratios supply information on the physical conditions in the line emitting gas. The electron density and temperature, the degree of ionization and excitation and the chemical composition, can all be deduced from line ratio analysis. There are some 20 broad lines, and a similar number of narrow lines, that can be measured in a single AGN, and the amount of information conveyed by the many line ratios is very large. The emission lines that are commonly detected in the spectrum of a quasar are Lyα (1215 Å), CIV (1549 Å), MgII (2799 Å), H$_\beta$ (4863 Å), and OIII (4959 Å).

In addition to the strong emission lines that characterize the optical and UV spectrum of quasars, the spectrum may also contain absorption lines, which may be caused by the absorbing material in the host galaxy itself, or may arise during the journey of the light from the quasar to us, due to the intervening gas in the intergalactic medium (IGM) and the interstellar medium (ISM) of the Milky Way.

These absorption lines may constitute metal systems, in which the doublets (i.e., two transitions that occur at the same redshift) MgII ($\lambda = 2795\AA$ and $2802\AA$) and CIV ($1548\AA$ and $1551\AA$) are the most frequent and easily identified. In general, they are narrow lines and their redshift is $0 < z_{abs} < z_{em}$, which means that they are due to cosmologically distributed gas along the line-of-sight and are not associated with the quasar.

Neutral hydrogen intersected by the line-of-sight to the quasar will produce numerous narrow absorption lines at $\lambda_{obs} \lesssim \lambda_{obs}(\mathsf{Ly}\alpha) = (1 + z_{em})\ 1216\ \AA$. The set of these absorption lines is denoted as the Lyman-α forest. In 1965, Gunn & Peterson found that this spectral region of reduced flux can put upper limits on the amount of intergalactic neutral hydrogen.

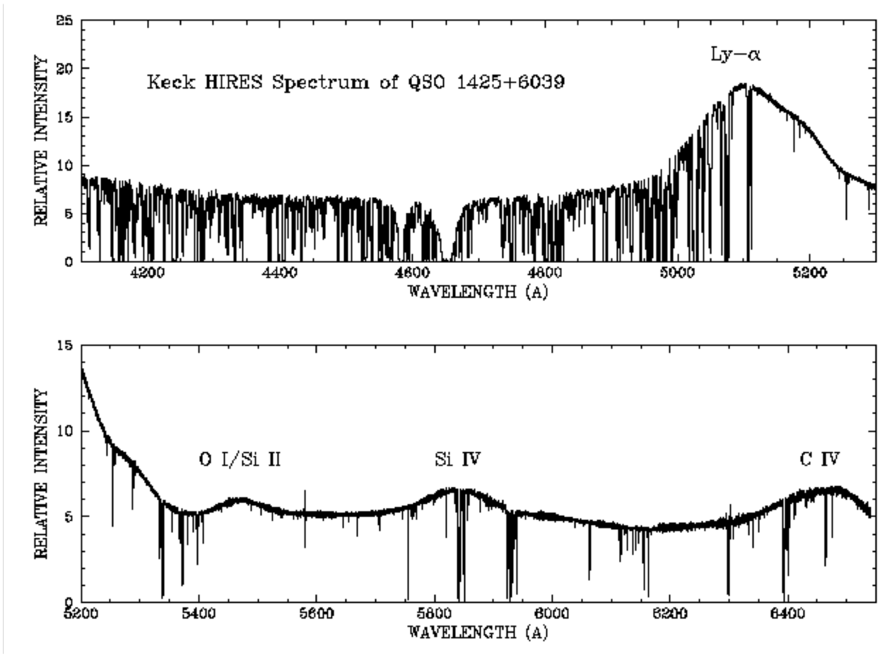


Figure 1.6: Spectrum of the quasar Q1425+6039 obtained with the High Resolution Echelle Spectrograph (HIRES) on the Keck-I telescope. The quasar has an emission redshift of $z_{em} = 3.18$. Upper panel: The Lyα forest originates from the absorption of neutral hydrogen by intervening intergalactic clouds. The strong line at 4650 Åis a damped Lyα feature (caused by a high column density of HI) at $z = 2.82$. Lower panel: Series of absorption lines.

Another class of absorption are the broad absorption lines, known as BALs, which can be found in about 40% of the quasars (Shankar et al. 2008) and originate from material in the AGN itself. In general, the redshift of the BALs is slightly lower than that of the emission lines; this means that the absorbing gas must be moving towards us and, for this reason, they are associated with sources in which material is flowing out at high velocities.

Since most absorption lines in quasar spectra (except perhaps for BALs) are not physically related to the AGN phenomenon, they can provide us with an opportunity to probe the matter along the line-of-sight to the quasar.

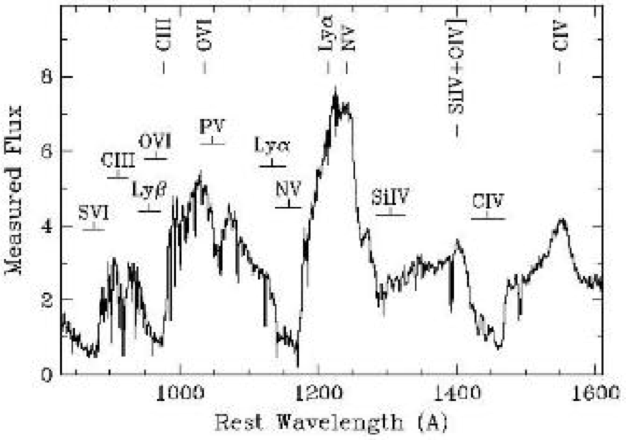


Figure 1.7: BALs in the quasar PG 1254+047 (Hamann 1998). On the blue side of every strong emission line very broad absorption is visible, believed to be caused by outflowing material. The flux is in units of 10^{-15} erg s^{-1} cm^{-2} $Å^{-1}$.

Studying in detail the space density of quasars with cosmic time is a powerful tool to establish a better assessment about the meaning of the priors we define in §3.4. In particular, some models of galaxy and quasar evolution can provide a better understanding about some intrinsic properties of these objects and the environments where quasars can be found, placing important constraints on early structure formation.

For instance, Hopkins et al. (2005) found that the lifetime and peak luminosities above any magnitude (independently of the imaging band) vary systematically according to the black hole mass. So, in this scenario quasar activity would be tied to the self-regulating process of the supermassive black hole in the center of the galaxy. Also, hydrodynamic simulations (e.g. Mayer et al. 2010) suggest that the supermassive black holes may have been formed by direct collapse of (metal-free) gas at the center of protogalaxies.

Based on the Blandford-Znajek mechanism (1977), in which the free energy associated with the black hole spin can be tapped by large-scale magnetic field lines, and be carried away in an electromagnetic jet, magnetohydrodynamic models have been used to study the formation of jets and the evolution of accretion disks in the surroundings of the supermassive black hole (Kato 2006), and more recently, to explain the observed radio and gamma-ray emission produced by some AGNs (Singh et al. 2015). Recent works (Nenkova et al. 2008a,b) show evidence that the dust around AGNs, which forms a toroidal shape, is most likely distributed in clumps, instead of being homogeneously distributed. Such clumpiness suggests that the separation between a type-I and type-II objects can be explained as the transition of an obscuring cloud along the line-of-sight.

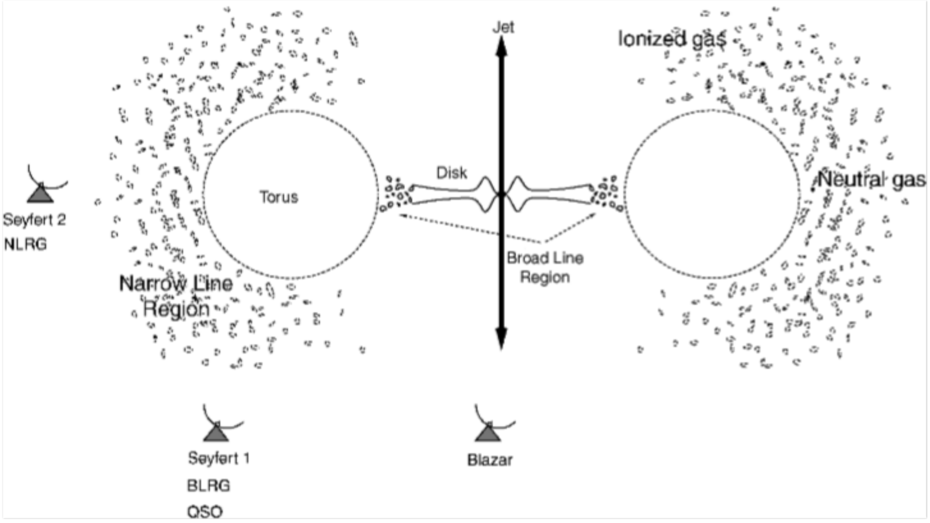


Figure 1.8: Current understanding of the unified model of AGNs. The accretion disk is surrounded by a thick torus containing dust which thus obscures the view to the center of the AGN. The difference between Seyfert 1 and Seyfert 2 is therefore merely a matter of orientation relative to the line-of-sight.

1.2 Stars

Stars are one of the most fundamental building blocks of galaxies. Their age, distribution and composition help us trace the history, dynamics and evolution of the galaxies.

Although not perfect blackbodies ("continuous spectrum"), because of the eventual presence of absorption lines, the spectra of stars are close enough to a blackbody spectrum and then the Wien's law

$$\lambda_{max} = \frac{0.29 \text{ cm K}}{T} \tag{1.2}$$

gives a good estimate of their surface temperature (T_{eff}), which can be determined from the wavelength where the emission peaks (λ_{max}).

Measuring the spectrum of a star is not always easy, but astronomers can often measure the color of a star using different filters. The color is related to the apparent brightness, which depends on the distance and is a measurement of how bright the star appears to a detector on Earth. Note that the apparent brightness is different from the luminosity, which is an intrinsic property of the star and is a measurement of the amount of light it emits from its surface.

So, using the information of the color of a star and also the fact that the absorption lines visible in the spectra of different stars are different (and have some dependence on temperature), it is possible to classify stars into different groups based on their luminosities and according to the atomic (and, in the case of cool stars, also molecular) spectral lines they present.

The spectral classes for stars are ordered with the hottest stars at the beginning of the sequence and the coolest ones at the end of the sequence. The current order of stellar spectral types is O, B, A, F, G, K, M. For instance, O stars have temperatures of the order of 50,000 K, whereas M stars are much cooler with about $3,000$ K. Each class is also divided in 10 subclasses labeled from 0 to 9. Thus a B0 star is slightly cooler than a O9 star.

The relationship between the star's luminosity and spectral classification (or, equivalently, absolute magnitude and effective temperature, respectively) can be summarized in plots known as Hertzsprung-Russell diagrams, or simply H-R diagrams. The upper left corner of an H-R diagram includes the hot, blue stars. The coolest, red stars are much fainter, and they lie at the lower right. The band connecting these two opposite regions is called the main sequence. Most of the stars occupy the main sequence region, including our Sun, which is a G star.

Another way of understanding the H-R diagram is observing that the stars are distributed according to their ages; therefore, it also provides an evidence that the main sequence is actually a mass sequence. Typically, the larger the star's mass, the shorter its lifespan. These features are shown in Fig.1.9.

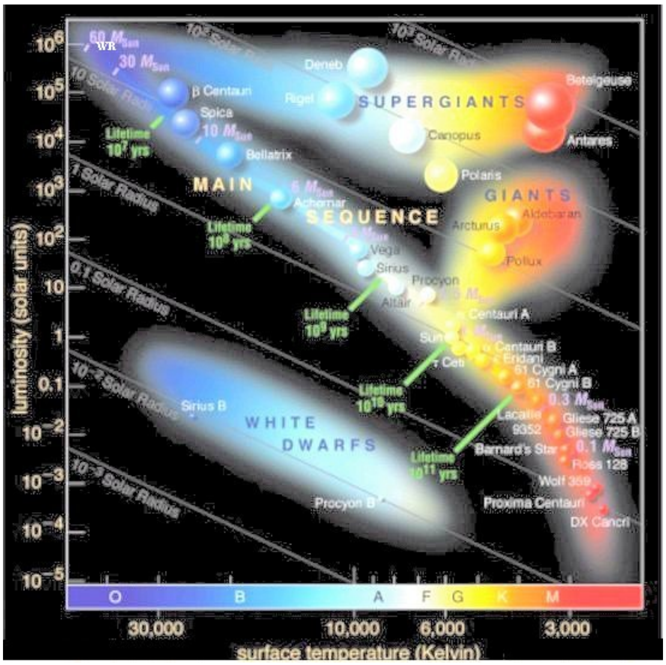


Figure 1.9: Hertzsprung-Russell diagram. The prominent diagonal region correspond to the main sequence. Above the main sequence, we find the giants and supergiants; the white dwarfs are found on the lower left side.

Looking at this diagram closely, we can note that most of the stars are concentrated in the main sequence. We can understand this sequence in terms of the luminosity-temperature relationship for blackbodies:

$$L = 4\pi R^2 \sigma T^4 \tag{1.3}$$

where σ is the Stephan-Boltzmann constant. Assuming that R is approximately constant in this band, i.e., assuming that all stars have roughly the same size, this equation tells us that the hotter a star is, the brighter it will be, and since the luminosity L scales as T^4 (temperature), small differences in T will cause large variations in L. Hence, we expect hot, blue stars to be much brighter than cool, red stars.

There are also stars outside the main sequence, in the upper right side and the lower left side of the diagram. The objects in the upper right region have the same temperature as M stars, but are much brighter. Considering again Eq.1.3, if two stars have the same temperature, the only way that one can be brighter than the other is if it has a larger radius R. Therefore, these stars are much larger than those directly below in the main sequence and, since they are red, we refer to them as red giants. Using the same argument, the stars in the lower left region have the same temperatures as O, B and A stars, but are much less luminous. Thus, these stars are much smaller than the ones directly above in the main sequence and are consequently called white dwarfs.

The observations and theory of stars show that the stars are not eternal: they follow a lifecycle. The process of star formation starts inside relatively dense concentrations of interstellar gas and dust known as molecular clouds. These regions are extremely cold, with temperatures of about 10 to 20 K. At these temperatures, gases become molecular, which means that the atoms bind together. However, there exists a minimum mass—known as the Jean mass—for the cloud to have its internal pressure balanced by gravity. When the cloud exceeds the Jean mass, it becomes unstable to gravitational collapse.

Star formation begins when the denser parts of the cloud core[1] collapse under their own gravity. Typically, these cores have a mass of about 10^{14} $M_\odot$. At some point, the core of the collapsing clump becomes so dense that the radiation generated inside the clump becomes trapped, i.e., it becomes opaque, causing the temperature of the core to increase quickly. At this moment, the core can be referred to as a protostar.

By the time the core of a cloud fragment has become a protostar, it has reached a temperature of several million Kelvin, and hence it becomes an IR source. At this temperature, the hydrogen in the core will be a plasma, a “soup” of ions and electrons, and the protons are packed together very tightly and are moving very rapidly. When the temperature reaches about 10^6 K, the protons are moving so fast inside the core that the electrical repulsion cannot prevent them from colliding. Once they collide, they fuse together in a process that generates energy.

Inside the core of stars like the Sun, fusion proceeds via a process called the proton-proton chain. In this multi-step process, six protons fuse together and the product is a helium nucleus and two protons. Therefore, energy is generated in the star’s core by converting hydrogen into helium. This energy provides enough radiation pressure to finally balance the inward pull of gravity, and then we say that the star is in hydrostatic equilibrium. At this point, the star will lie on the main sequence, its first stage of evolution.

The temperature that the core of a protostar reaches depends on its mass. The more massive the protostar, the hotter it gets. If the core reaches a high enough temperature (more than 20 million Kelvin), a different set of fusion reactions proceed more efficiently than the proton-proton chain. This process, called the CNO (carbon-nitrogen-oxygen) cycle, occurs in stars more massive than the Sun. The CNO cycle still requires hydrogen to proceed, so even in these stars the main fuel

---

[1] The cores are denser than the outer cloud, so they collapse first.

for the fusion reaction is hydrogen. In both the proton-proton chain and the CNO cycle, one element is being converted into another via nuclear fusion; this process of creating new elements is called nucleosynthesis.

The first step of the proton-proton chain also generates neutrinos. Therefore, since it is not possible to directly observe the core of a star, the detection of solar neutrinos provides evidence that nuclear fusion is indeed the power source of the Sun.

The lifecycle of a star depends on its mass. In the following, we present an evolutionary track that is appropriate for Sun-like stars.

When the star has used up the majority of hydrogen in its core, the main sequence phase ends and its subsequent evolution phase begins. When the star's core becomes filled with helium, the star will fall out of equilibrium. As the total pressure decreases, gravity will once more dominate. Consequently, the star begins to contract, increasing the pressure and the temperature in the core. The helium core will continue to generate energy by gravitational contraction.

Although fusion has turned the hydrogen in the core into helium, most of the outer layers of the star are made of hydrogen, including the layer immediately surrounding the core. Thus, when the core reaches a critical density and temperature during its contraction, it can ignite hydrogen fusion in a thin shell outside the helium core. Hence, the first stage after the main sequence can be refered as the hydrogen shell fusion stage.

The energy being generated in the core will be more intense than during the core hydrogen fusion phase, so the outer layers of the star will experience a larger pressure, causing them to expand significantly. As a side effect of this expansion, the outer layers will cool down because they are now further away from the energy source (the hydrogen shell around the core).

The observable changes in the outer layers of the star will occur in two phases. First, the star will appear to cool slowly and will undergo a modest increase in luminosity. During this phase, the path the star will follow in the H-R diagram is almost horizontal to the right of its position on the main sequence. Stars in this phase are usually referred to as subgiants. Next, the star will grow to as much as 100 times its original size, which will cause a significant increase in luminosity with only a small decrease in temperature, so the star will move almost vertically in the H-R diagram. Stars in this area of the H-R diagram are usually referred to as red giants.

During the red giant phase, the core is not in equilibrium: all of the fusion is occurring in a shell outside the helium core, so there is no energy generation or outward radiation pressure to support the helium core. For this reason, the core of the star continues to collapse during this phase. In many low mass stars (from about 0.5 - 3 $M_\odot$), the core can be compressed to the point that it becomes a degenerate gas.

At some point after the core has become degenerate, the core temperature reaches approximately 100 million Kelvin, creating the proper conditions for three helium nuclei to fuse together to form one carbon nucleus. This is the triple-alpha process. The star will convert all of its core helium into carbon and oxygen, and then fusion will end once again. With no radiation pressure to support it, the core will again begin to collapse inward. Because there is still so much helium and hydrogen outside of the core of the star, after core helium fusion ends, the increased temperature can once again ignite shell helium fusion just outside of the carbon/oxygen core, and shell hydrogen fusion can continue outside the helium shell. During this second phase of shell fusion, the outer layers of the star will expand again, but this time by an even larger amount. In this phase, the star can be called an asymptotic giant branch star, or sometimes a red supergiant star.

The core of the star is no longer undergoing nuclear fusion of any variety, so it is once again collapsing. When the core reaches a size approximately equal to that of the Earth, the collapse will stop. Since no fusion is occurring, the carbon/oxygen remnant of the stellar core will not generate any new energy. Instead, it will simply cool off slowly by radiating light, getting fainter and fainter until it no longer gives off enough light to be visible, in a process that can take billions or even trillions of years. While the object is still visible, it is called a white dwarf, and it occupies the lower left side of the H-R diagram because of its high temperature and faint luminosity. The Chandrasekhar limit imposes an upper limit of 1.4 $M_\odot$ for the white dwarf to remain in equilibrium resisting gravitational collapse.

The lifecycle of high mass stars diverges from that of low mass stars after the stage of carbon fusion. In low mass stars, once helium fusion has occurred, the core will never get hot or dense enough to fuse any additional elements, so the star begins to die. However, in high mass stars, the temperature and pressure in the core can reach high enough values that carbon fusion can begin, followed subsequently by oxygen fusion, and then even heavier elements—like neon, magnesium, and silicon—can undergo fusion, continuing to power the star.

When iron builds up in the core of a high mass star, there are catastrophic consequences: the process of fusing iron requires the star's core to use energy, which causes the core to cool. This causes the pressure to go down, which speeds up the gravitational collapse of the core. After the core collapses, it rebounds. A large quantity of neutrinos get created in reactions in the core, and the rebounding core and the newly created neutrinos go flying outward, expelling the outer layers of the star in a gigantic explosion called supernova (to be precise, a type II or core collapse supernova).

When the core of a star collapses at the beginning of a Type II supernova explosion, a neutron star is created. The neutron star is supported by neutron degeneracy pressure, but in the case of the most massive stars (M > 25-50 $M_{\odot}$), not even neutron degeneracy pressure can stop the collapse of the core. In this case, matter collapses into a single point referred to as a singularity. Even light cannot escape from this object, and for this reason it is called a black hole.

Chapter 2

Cosmological surveys

Cosmology is the field of study that analyzes the Universe as a whole: its origin, structure, composition and evolution over time. Its four pillars are the expanding Universe (Hubble, 1929); the origin of the cosmic microwave background radiation (CMB, Penzias & Wilson, 1965; Smoot et al. 1992), i.e., the radiation emanating from the surface of last scattering, the last physical interaction of photons with matter when the Universe was about 3.8x10^5 years old; the nucleosynthesis of light elements; and the formation of galaxies and the large-scale structure.

These cosmological observables can provide us with information about the composition of the Universe and in which proportions each of its constituents appear. According to the current most accepted model, ΛCDM, the main ingredients are the baryonic matter (ordinary matter), relativistic species (i.e., electromagnetic radiation and neutrinos), cold dark matter (CDM) and dark energy (believed to be responsible for the present phase of accelerated expansion). The standard model (Planck Collaboration 2015) asserts that the Universe is flat and its present composition is about 2% of baryonic matter; 30% of cold dark matter, which comprises the halos surrounding galaxies and clusters of galaxies and is said to be cold because it is non-relativistic during the era of structure formation; and 68% of dark energy.

With the advent of larger telescopes and more sensitive instruments the distant Universe has become more accessible to us. The key to the the great achievements on cosmology in recent years is that models can now be tested by observational data with unprecedented precision.

Yet, many fundamental questions remain without an answer, and although the Universe can be mapped in a wide range of scales and frequencies, high-quality observations still come up against the detection limits of the instruments. Clearly, the selection of the area is a very important part of the strategy of any survey, and most combine, in different degrees, two complementary techniques (spectroscopy and photometry) to select their targets. There is always a trade-off between the area and the depth of the survey, which have consequences on how the sources—in particular, the fainter ones—are detected.

Each photometric band has its own biases due to selection effects and, therefore, the use of different imaging bands provides different perspectives for the detection of sources. On the other hand, the use of imaging to pre-select targets for spectroscopy usually miss the fainter sources. Therefore, the main challenges consist on understanding better the systematics of the data sets, in order to obtain catalogs of higher completeness and reliability, and to minimize the contamination between different classes of objects.

In the following sections we present some features of the main recent cosmological surveys and a discussion about the differences between two techniques of redshift estimation.

2.1 Distance measures

In this section we make a brief review of cosmological distance measures. This discussion is based on the following references: Hogg 2000, Dodelson 2003 and Weinberg 2008.

The cosmological principle states that the Universe is homogeneous and isotropic. The Friedmann-Lemaître-Robertson-Walker (FLRW) metric incorporates this principle into its symmetries and the distance in the four dimensional space-time is

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{2.1}$$

where $a(t)$ is the scale factor, and k can assume positive or negative values.

All evidence shows that the Universe is expanding. Due to this expansion, the distances between two objects or events evolve with time, so that an observer on Earth is actually looking back in time while looking out in distance. In cosmology, there are different sorts of distances, but it is always convenient to define a comoving distance which remains fixed with time for objects moving with the Hubble flow. The distance on the comoving grid is then established as a fundamental distance measure, and all the other distances (e.g., angular diameter distance and luminosity distance) can be derived from it.

The redshift z is one of the fundamental observables in cosmology. In an expanding Universe, the wavelength of light is stretched out, meaning that the observed wavelength is larger than the emitted one. So the relation

$$\frac{\lambda_o}{\lambda_e} \equiv 1 + z = \frac{1}{a} \tag{2.2}$$

enables us to determine the shift (z) of the observed atomic line (λ_o) with respect to the rest frame emission line (λ_e). The shift z is related to the relative velocity between the source and the observer; accordingly, it can be either positive or negative. In cosmology, the term *redshift* is widely used to describe the recession of galaxies, even when this shift is negative (also denoted as blueshift[1]).

At small distances, or equivalently at small redshifts, the relation $z \approx \frac{v}{c}$ stands and the recession velocity v is linearly proportional to the distance. The constant of proportionality is the Hubble constant $\mathrm{H_0}$ (usually written as $\mathrm{H_0}$ = 100 h km $\mathrm{s^{-1}}$ $\mathrm{Mpc^{-1}}$) and then

$$v = H_0\, d \tag{2.3}$$

where the subscripted "0" refers to the present epoch. This relation was first put forward by Hubble, in 1929, and it was considered an evidence to the theoretical model of an expanding Universe as a solution to Einstein's equations. The inverse of the Hubble constant gives the Hubble time $t_H = \frac{1}{H_0} = 9.78\text{x}10^9\, h^{-1}$ yr ; we can also define the Hubble distance $D_H = \frac{c}{H_0} = 3000\, h^{-1}$ Mpc. These quantities set scales to the Universe.

Although the Hubble diagram is still the most direct evidence of an expanding Universe, at high redshifts the linear relation between distances and velocities is no longer valid because the evolution of the scale factor with the cosmic time t becomes more important. Then, the Friedmann equation determines the evolution of the scale factor:

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{3a^2} \tag{2.4}$$

---

[1] The term blueshift is used when the light of an object moves closer to us. Interestingly, the Andromeda galaxy, the closest neighbor of the Milky Way, has a negative shift of z = -0.001, since the two galaxies are approaching each other.

where ρ is the total energy density, which include contributions from various distinct components. It is convenient to express the densities in units of the critical density $\rho_c = \frac{3H^2}{8\pi G}$, by introducing the density parameter:

$$\Omega_i = \frac{\rho_i}{\rho_c} = \frac{8\pi G \rho_i}{3H^2} \tag{2.5}$$

where the index i denotes the individual components.

The energy density of each distinct component evolves as the Universe expands; so, we may define equations of state of the form

$$p_i = w_i \, \rho_i \tag{2.6}$$

where p is the pressure, and w is a constant (or, in many dark energy models, some function of redshift).

Without going into the details, we can use the energy-momentum conservation equation $\nabla_\mu T^{\mu\nu} = 0$ to demonstrate that the energy density has a power-law dependency on the scale factor:

$$\rho_i \propto a^{-3(1+w_i)} \tag{2.7}$$

where w assumes the following values: 0 for matter, 1/3 for radiation, -1/3 for curvature and -1 for the cosmological constant. The matter component accounts for baryons and dark matter, a sort of matter that does not emit nor absorb any kind of electromagnetic radiation. Radiation accounts for all of the relativistic particles, such as photons and neutrinos. The curvature characterizes the geometry of the Universe, and there are three options for the geometry of the space: flat ($\Omega_k = 0$), open ($\Omega_k < 0$) or closed ($\Omega_k > 0$). The cosmological constant is believed to be one form of dark energy in which the energy density is constant.

Finally, we can rewrite Eq. 2.4 as a function of redshift:

$$H(z) = H_0 E(z) \tag{2.8}$$

where $E(z) = \sqrt{\Omega_m (1+z)^3 + \Omega_\Lambda}$, Ω_m and Ω_Λ are the matter and cosmological constant density parameters, respectively. Based on cosmological observations, which give evidence that the Universe is flat and that the current radiation density is much smaller when compared with the mass density, we assume contributions only from matter and cosmological constant.

The total line-of-sight comoving distance out to a distant emitter is given by

$$D_C(z) = c\int_0^z \frac{dz'}{H(z')}. \tag{2.9}$$

We can also define the transverse comoving distance (D_M), which provides the comoving distance between two events at the same redshift but separated by some angle $\delta\theta$:

$$D_M = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh\left(\frac{\sqrt{\Omega_k} D_C}{D_H}\right) & , \ \Omega_k > 0 \\ D_C & , \ \Omega_k = 0 \\ \frac{D_H}{\sqrt{|\Omega_k|}} \sin\left(\frac{\sqrt{|\Omega_k|} D_C}{D_H}\right) & , \ \Omega_k < 0 \end{cases} \tag{2.10}$$

where D_H is the Hubble distance. Note that D_M reduces to D_C in a flat Universe.

Physical distances are proportional to the comoving distance times the scale factor a, which at earlier times was smaller than today and whose present value is set to 1.

The angular diameter distance (D_A) is defined as the ratio of the physical transverse size to the angular size. This distance does not increase indefinitely as $z \to \infty$: it turns over at $z \sim 1.5$, so that distant objects appear larger in angular size.

$$D_A = \frac{D_M}{1+z}. \tag{2.11}$$

Another usual way of inferring distances is by measuring the flux (F) from a source of known luminosity. The flux at a distance d from the source scales as $F \propto d^{-2}$, since the total luminosity through a spherical shell of area $4\pi d^2$ is constant. On the comoving grid, due to expansion, the number of photons that cross the shell in some fixed time interval will be smaller today than at the emission by a factor of a. Therefore, we can generalize the observed flux as $F \propto D_M^2$, and then the luminosity distance is defined as

$$D_L(z) = (1+z)D_M. \tag{2.12}$$

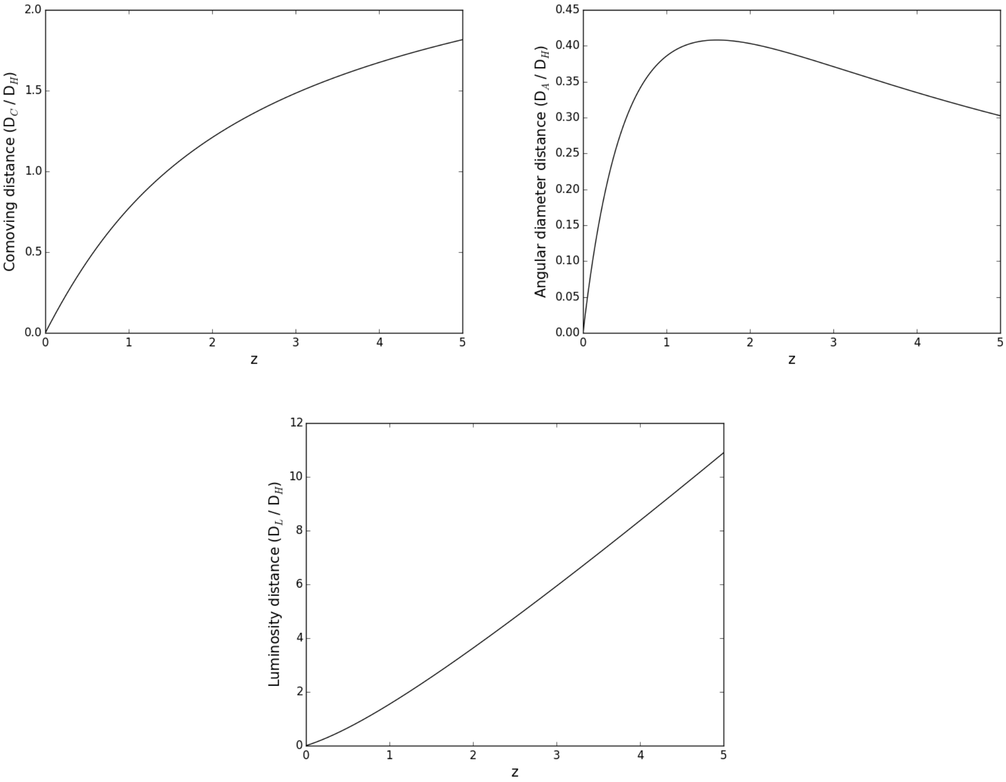


Figure 2.1: The dimensionless comoving, angular diameter and luminosity distances. The curves are for the ΛCDM model with parameters $(\Omega_m, \Omega_\Lambda) = (0.3,\ 0.7)$.

Lastly, we define the comoving volume element in the solid angle $d\Omega$ and redshift interval dz as

$$dV_C = D_H \frac{(1+z)^2 D_A^2}{E(z)} d\Omega\, dz. \tag{2.13}$$

Integrating over redshift and across the whole sky gives the total comoving volume, which in a flat Universe is given simply by $V_C = \frac{4\pi}{3} D_M^3$.

The luminosity distance and comoving volume are the relevant quantities to develop the luminosity function for quasars in §3.2.

2.2 Cosmic frontier

At large distances, even the most intrinsically luminous galaxy would appear to be faint. This means that the quality of the observations is directly linked to the properties of the astronomical instruments, which in turn reflects the science purposes of the survey. Hence, being aware of the limitations imposed by the instrumentation is fundamental to comprehend the effectiveness of the survey.

The sensitivity specifies how dim a source can be to still be observable in a given integration time; it depends on the aperture of the instrument (i.e., its photon collecting area), on the efficiency of the detector and also on the sky background. The angular resolution, which in general is limited by turbulence in the atmosphere (an effect called seeing), specifies down to which angular separation two sources on the sky can still be separated by the detector. The spectral resolution specifies the capability of separating different spectral features, in terms of the wavelength. Finally, the field of view of the camera, the readout noise of the CCD, the number of clear nights at the astronomical site, and the effective exposure times of the imaging bands are factors that also influence the efficiency of observations. Thus, ideally, telescopes with large apertures and high angular resolution are preferable for extragalactic astronomy, and this applies to all wavebands.

Redshift surveys are one of the primary tools of observational cosmology (Colless 2000) and usually rely on both spectroscopic and photometric strategies to acquire large numbers of sources across huge volumes allowing us to measure, e.g., the matter power spectrum, and to estimate how galaxies and other tracers of large-scale structure are related to the underlying dark matter density.

The current and forthcoming surveys such as DES (The Dark Energy Survey Collaboration 2005), J-PAS and J-PLUS (Benítez et al. 2009, Benítez et al. 2014), PFS (Ellis et al. 2012), LSST (Ivezic et al. 2008), and Euclid (Laureijs et al. 2011; Amendola et al. 2012), promise to deliver greater depth, better image quality, and photometric accuracy, as well as large numbers of objects.

The Two-Degree Field facility (2dF) (Lewis et al. 2001) at the Anglo-Australian Telescope is a 400-fibre optical spectrograph with a 2 degree diameter field of view. All of its innovative features, such as a robotic fibre-positioner and the two focal planes, have made the 2dF ideally suited for massive redshift surveys. In fact, this wide-field spectroscopic facility was designed to provide large numbers of spectra and generate targets for the next generations of telescopes. The 2dF Galaxy Redshift Survey (2dFGRS, Colless et al. 2001), and its companion—the 2dF QSO Redshift Survey (2QZ, Boyle et al. 2000), which were designed to provide a characterization of the galaxy and quasar population on the nearby Universe and at large-scales, have obtained spectra for almost $250,000$ objects and redshifts for $221,414$ $b_J < 19.45$ galaxies (Colless et al. 2003) and over 25,000 $b_J < 21$ quasars (2QZ team).

The Sloan Digital Sky Server (SDSS) is a dedicated 2.5 m telescope that has being working since 2000 (York et al. 2000). Located at the Apache Point Observatory, in New Mexico, the telescope is equipped with a mosaic format camera with 30 CCDs (arranged in six rows of five CCDs each), which scans the sky in five optical bands, the so-called *ugriz* system. This photometric system was developed so that the transmission curves overlap as little as possible. Besides the camera, the telescope has also four digital spectrographs to obtain the spectra of objects selected from the imaging data.

The instrument operates in a drift scan mode (SDSS Collaboration): the camera slowly reads the CCDs as the data is being gathered, while the telescope moves along great circles on the sky so that images of objects move along the columns of the CCDs at the same rate the CCDs are being read. It takes about 54 s for an object to move from the beginning of a CCD to the end, so the effective integration time per filter is 54 s. However, since there is some space between the rows of the CCDs, actually each object has one image in each filter taken at 71.7 s intervals. In the end, it takes about 5.7 minutes for passage over the entire photometric array.

The imaging bands are employed to select the targets whose spectra will be taken, and a magnitude limit of $i \sim 20.1$ is imposed for these candidates. Therefore, the objects are detected based on their colors and magnitudes, and classified as point-like or extended sources. At this moment, degeneracies between quasars and stars may arise, because the selection of quasars is allowed to be close to the stellar locus around $z \sim 2.8$, where quasar colors approach those of late A and early F type stars (York et al. 2000).

With its new techniques and remarkable discoveries, the SDSS has established a new paradigm for the subsequent generation of surveys. Its major contributions include high-precision maps of large scale structure; detection of quasar samples thousands of times larger than the ones that already existed; studies that demonstrated the bimodal distribution of galaxies divided into star-forming populations and passive galaxies; acquisition of relatively good photometric redshifts for luminous red galaxies (LRGs); identification of rare and intrinsically faint stellar populations; detection of over 10,000 variable and transient sources; identification and color measurements for thousands of asteroids and minor objects. This wide range of discoveries is only a sketch of all of the SDSS science results.

Currently, the SDSS is in its last generation (SDSS-IV, 2014-2020): the APOGEE-2 (a stellar survey of the Milky Way) and eBOSS (a cosmological survey of galaxies and quasars) surveys will extend the precision of the cosmological measurements to the early phase of cosmic history, extend the infrared spectroscopy of the Galaxy, and construct spatially resolved maps to explore the internal structure of nearby galaxies.

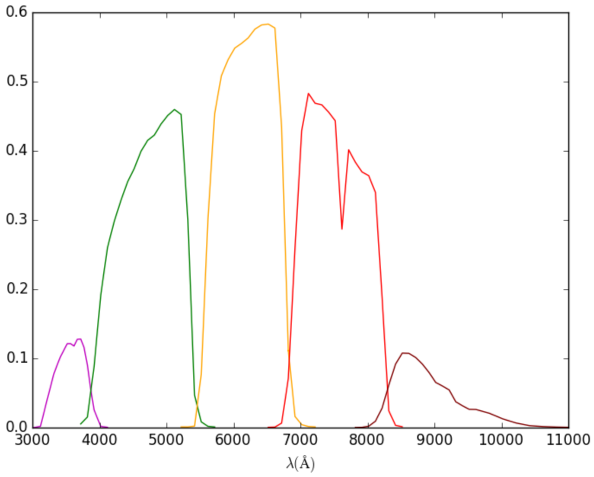


Figure 2.2: SDSS *ugriz* filter system.

The phenomenal successes of the Sloan Digital Sky Survey motivated the implementation of new surveys. In particular, the Baryon Oscillation Spectroscopic Survey (BOSS, Schlegel et al. 2007), one of the four surveys that comprise the SDSS-III, was designed to map out the baryon acoustic oscillation (BAO) signature with unprecedented accuracy on large scales. It has introduced a pioneer method of measuring the BAO scale at high redshift with quasars, by building a catalog of Lyα absorption systems that spans $2.15 < z < 3.5$ (Dawson et al. 2012).

The Subaru Prime Focus Spectrograph (PFS), expected to have the first light in 2017, is a multi-object spectrograph that will mostly map emission-line galaxies (ELGs) up to $z \sim 2.4$. With an 8.2 meter aperture, the Subaru telescope will be powerful for obtaining spectra of faint objects, and it will also detect many thousands of galaxies and quasars.

In general, broad-band filters detect only "breaks" (Bolzonella et al. 2000), because they are not sensitive to the presence of emission lines (except when the contribution to the total flux in a given filter is higher than, or of the same order as the photometric errors, as it happens, for instance, for some AGNs). Projects such as COMBO-17 (Wolf et al. 2008) and ALHAMBRA (Molino et al. 2014), which have carried out a medium-band imaging, appear as an alternative to the broad-band filters and show the potential of this new approach.

COMBO-17 is a low-resolution multi-object spectroscopic survey carried out in 17 optical filters: 5 broad-band filters (UBVRI) and 12 medium-band filters, in 1 deg^2 at high latitudes. This survey was performed to study the evolution of galaxies at $z \lesssim 1$ and quasars at $1 \lesssim z \lesssim 5$. The highlights of COMBO-17 include the detection of almost 25,000 galaxies with a precision of $\sim 0.02(1+z)$ in the photometric redshift (photo-z) errors and robust constraints on the evolution of the luminosity function for quasars.

The ALHAMBRA (Advanced, Large, Homogeneous Area, Medium-Band Redshift Astronomical) survey was designed to map 4 deg^2 over eight deep sky areas for the study of the cosmic evolution and cosmic variance using 20 contiguous medium-band filters (3500-9700 Å) plus 3 standard broad-band NIR filters (JHK) with limiting AB magnitudes that reach $\sim$ 24.5. The data was gathered between 2005 and 2012 with the LAICA camera and the NIR instrument Omega-2000 on the 3.5 m telescope at the Calar Alto Observatory. The use of NIR filters is important for breaking the color-redshift degeneracies between low and high redshift galaxies. This degeneracy is due to a possible confusion between the Balmer and Lyman breaks, one of the most salient features in the spectral energy distributions (Moles et al. 2008). If no infrared information is available, then it is not possible to determine the slope of the rest-frame red end of the spectrum and, consequently, it is not possible to distinguish the differences between these families.

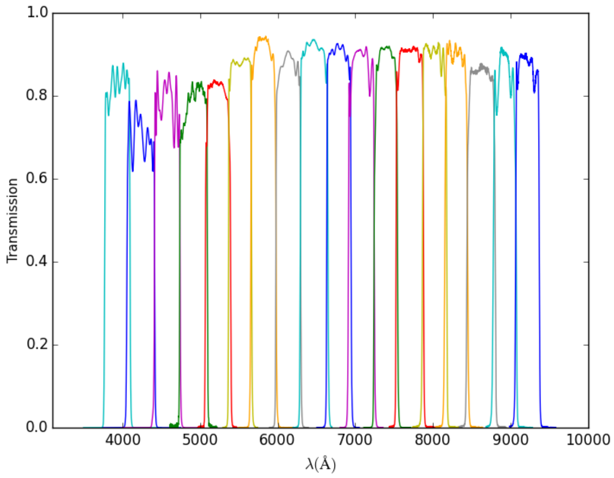


Figure 2.3: ALHAMBRA filter system with 18 contiguous filters.

Benítez et al. (2009) showed that narrow-band filter systems are even more effective with respect to the photo-z depth. In this context, it appears J-PAS (Javalambre Physics of the Accelerating Universe Astrophysical Survey), which will rely on a 54 narrow-band filter system, of width ∼145 Å, plus 2 medium-band and 3 standard broad-band (*ugr*) filters, using a dedicated 2.5 m telescope (T250), a 80 cm telescope (T80-North) to carry out large sky photometric surveys and a camera with 1.2 Gpix (Benítez et al. 2014). This is more than enough to resolve the broad emission lines of type-I quasars (as well as most broad absorption line objects), and is sufficient to detect the narrow lines of many type-II's and AGNs.

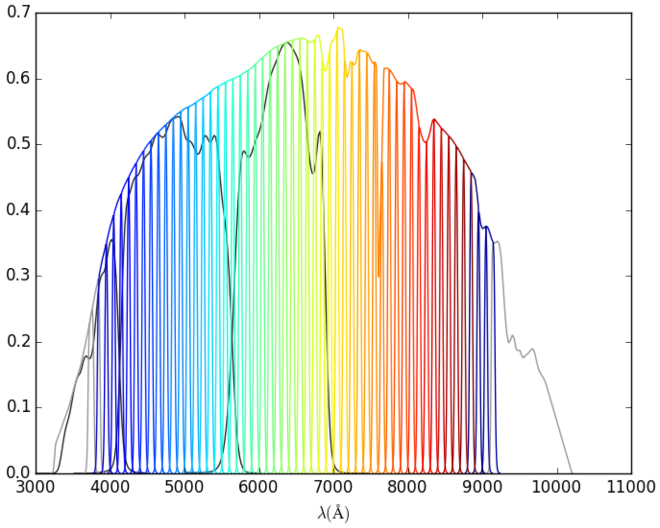


Figure 2.4: J-PAS filter system.

In addition to J-PAS, which is expected to start taking data in 2016, the survey J-PLUS is already operational, taking science data. J-PLUS employs the JAST/T80 telescope to observe more than 8,500 square degrees of sky employing an optimized system of four broad-band (*griz*) and eight narrow-band filters (Benítez et al. 2014). J-PLUS reaches about one magnitude deeper than SDSS, being ideal to properly recover stellar parameters (such as T_{eff}, log g, and [Fe/H]) through the fitting of flux calibrated models of the observed stars.

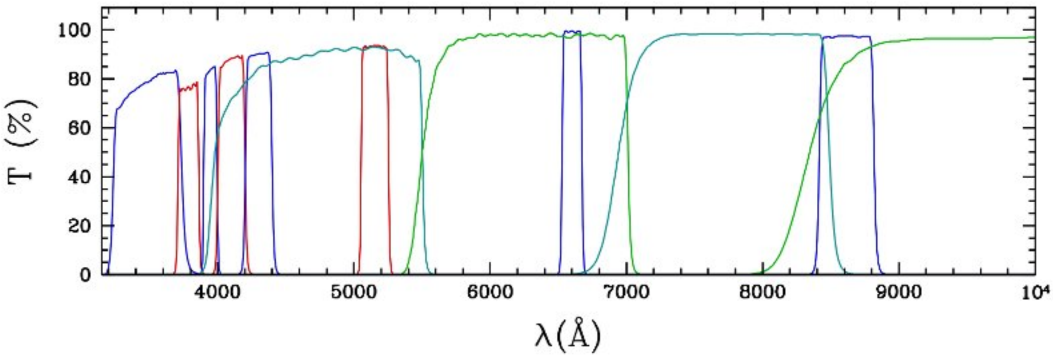


Figure 2.5: J-PLUS filter system (J-PAS Collaboration).

The present and next generation of surveys point to new frontiers in cosmology, where greater volumes will be covered, achieving higher completeness and improving the photometric accuracy.

2.3 Spectroscopy versus photometry

Redshifts can be estimated through two different but complementary techniques. In spectroscopy, the light of the astrophysical object is spread out into its wavelengths, generating a spectrum. Emission and/or absorption features in this spectrum tells us about the composition and the motion of the object. Since the atoms and molecules have characteristic spectral signatures (rest-frame frequency) due to their quantum mechanical properties, and assuming they are the same whether they are located at Earth or in outer space, it is possible to identify known spectral lines and compare the measured wavelengths with them. This relation is given by Eq. 2.2.

In spectroscopy, the light from the galaxy is separated into narrow wavelength bins (resolution elements), which for most cosmological applications have a few angstroms (Å) width. Each bin then receives only a small fraction of the total light emitted from the object. So, to achieve a sufficiently high signal-to-noise ratio in each bin, very long integration times are required, and since this technique requires a deep and wide imaging to select the targets, it becomes very time consuming (Connolly et al. 1997). This task is especially hard when the method is applied to faint objects or the sample is too large and a complete spectroscopic coverage is not practical.

In view of these difficulties, using the photometry is an alternative, where the redshift is derived from the colors of different standard imaging bands of an object, rather than from its spectrum. With this technique, the redshifts can be measured much faster and in larger quantities than their spectrometric counterparts. In photometry, the wavelength "bins" are much larger: typically of the order of hundreds or thousands of angstroms, which requires a shorter exposure time to obtain the same signal-to-noise ratio. In addition, imaging detectors cover a greater area of the sky than multi-object spectrographs. Thus the redshifts of more objects can be measured simultaneously.

Of course, these advantages come at the expense of higher redshift errors. In order to control these uncertainties, imaging surveys use large spectroscopic samples as calibration sets.

We can illustrate the typical integration times required by each of these techniques with the SDSS. As discussed in the previous section, the SDSS camera continually sweeps the sky in circles, so that a single point on the sky passes through the five filters in succession. As a result, the effective integration time per filter is 54.1 s, and it takes 5.7 minutes (~ 340 s) for the passage over all filters. Conversely, the spectrum of an object is obtained after an exposure time of the order of 900 s.

Therefore, albeit with a reduced spatial resolution in the radial direction due to the photometric redshift error (Blake & Bridle 2005), from a statistical perspective the photometry has the advantage of providing larger volumes and densities at reduced times. As a result, this technique has become increasingly popular for the redshift estimation of galaxies.

2.4 Photometric redshift estimation

The technique of estimating photometric redshifts is usually divided in two main groups (Abdalla et al. 2008): template fitting (Loh & Spillar 1986) and empirical training set (proposed originally by Connolly et al. 1995, 1997).

The template fitting technique derives the photo-z by fitting a set of reference spectra (which constitute the template library) to the data sample. The comparison between the observed spectral energy distribution (SED) and the energy distribution of the reference spectra, using a photometric system, can be made with the χ^2 test:

$$\chi^2(z) = \sum_i^N \left[\frac{f_i - b f_i^t}{\sigma_i} \right]^2 \tag{2.14}$$

where f_i and σ_i are the observed fluxes and their corresponding uncertainty in filter i, f_i^t are the template fluxes, and b is a normalization constant. So, for each object there is a distribution of possible values of photo-z. Thus the best photo-z will be the one that minimizes the χ^2 value in Eq. 2.14.

The template library can be composed of either synthetic or empirical spectra, but it must include SEDs for a variety of types (Dahlen et al. 2013) to represent different redshifts, star-formation histories, chemical abundances, mixture of dust and stars, and so on, and correspond to all kinds of objects that compose the sample. Therefore, given an observed SED, the method consists of finding the best combination of SEDs (allowing interpolation between them) and the best set of redshifts to describe the observations. This suggests that, in order to avoid potential mismatches, the template library should ideally display the same properties of the sample for which one wants to estimate the photo-z. Thus, the SED fitting technique is very sensitive to template choice, and its efficiency depends on the detection of strong spectral features and on the characteristics of the photometric system.

The other technique uses a training set, i.e., a subsample of objects with known magnitudes and spectroscopic redshifts, to derive an empirical relation between magnitudes (and/ or colors) and redshifts:

$$z_{photo} = f(m, C). \tag{2.15}$$

This parametrization can be applied to objects for which no spectroscopic information is available, yielding an estimative for the photo-z. This method has the advantage of not making any assumptions regarding to the types of objects or their evolution, so it is preferred when great statistical precision is required.

However, this approach lacks flexibility, since a new empirical relation between redshift and magnitude must be computed for each survey, and it might also introduce some bias while computing the redshift for the fainter sources, due to the fact that the training set is typically composed of the brightest objects.

Today, a large number of codes that use one (or both) techniques are publicly available. Codes based on template fitting approach include: HyperZ (Bolzonella et al. 2000), BPZ (Benítez 2000), LePhare (Arnouts et al. 1999; Ilbert & Arnouts 2006), and ZEBRA (Feldmann et al. 2006). Codes that include the empirical training set approach: ANNz (Collister & Lahav 2004), Multilayer Perceptron Artificial Neural Network (Vanzella et al. 2004), ArborZ (Gerdes et al. 2010), “Random Forests” (Carliles et al. 2010). There are also codes combining features from both techniques, like GOODZ (Dahlen et al. 2010), and EAZY (Brammer et al. 2008), which can use a training set to derive corrections to zero-points to minimize systematic differences between the templates and the observed SEDs. EAZY actually performs an iterative adaptation of the templates by introducing a template error function to account for the wavelength-dependent template mismatches.

In general, the efficiency of the photometric redshift estimation is quantified by the precision and the outlier fraction. A popular way of estimating the precision is the normalized median absolute deviation (NMAD):

$$\sigma_{\mathsf{NMAD}} = 1.48\mathsf{x}\mathsf{median}\left|\frac{\Delta z - \mathsf{median}(\Delta z)}{1+z_{spec}}\right| \tag{2.16}$$

where $\Delta z = z_{photo} - z_{spec}$. Note that the term $\mathsf{median}(\Delta z)$ already constitutes a way of estimating the accuracy of the measure. The outlier fraction is defined in terms of some factor of the σ_{NMAD}; for the analysis of our results in Chapter 5, we adopt 2xσ_{NMAD}:

$$\left|\frac{\Delta z}{1+z_{spec}}\right| > 2\ \sigma_{\mathsf{NMAD}}. \tag{2.17}$$

A comparative study between six photometric redshift codes was performed by Abdalla et al. 2008, using 5,482 LRGs from the 2SLAQ sample. In Fig.2.6, we illustrate the performance of three of these codes: two template-fitting (BPZ and LePhare) and the neural network method ANNz. Each code was run several times in order to obtain the set of parameters that optimized the photo-z estimation. So, this is actually a “code plus library comparison”. However, the same redshift resolution was applied to all codes to ensure that the uncertainties come from the photo-z estimation. In a general sense, the three codes perform similarly, however LePhare seems to work better at small redshifts, and ANNz seems to perform slightly better when compared to the template-fitting methods, which may be related to the use of a complete and large training set. The plots are color-coded and the scale is exponential. We can note there is a large number of outliers in the sample.

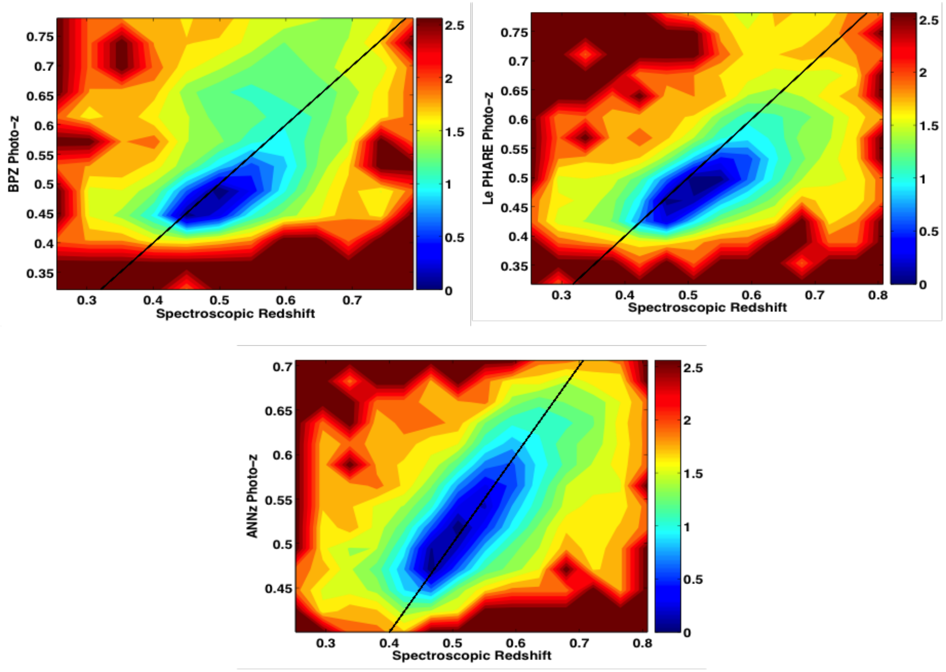


Figure 2.6: Distribution of the photometric redshift as a function of the spectroscopic redshift for three photo-z codes: BPZ, LePhare and ANNz. The plots are color-coded and the scale is exponential. A color difference of one is equivalent to the density being decreased by a factor of e. The solid black lines show where the photometric redshift equals the spectroscopic redshift. (Abdalla et al. 2008)

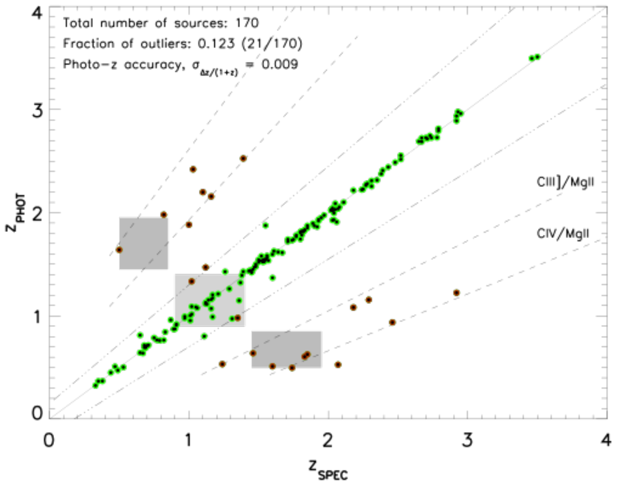


Figure 2.7: Comparison between the photometric redshift and the spectroscopic redshift for a sample of 170 quasars in the ALHAMBRA fields using LePhare. The solid line gives the $z_{photo} = z_{spec}$ relation; the dashed line represents the boundary between good solutions (green dots) and outliers (red dots). The dashed lines correspond to the degenerescence lines between pairs of emission lines. The central light grey square shows the precision of the photo-z estimation in the interval z= [0.9, 1.4]; the dark grey squares show the presence of a color-redshift degeneracy, which implies in photo-z solutions which are overestimated or underestimated. (Matute et al. 2012)

In Fig.2.7 we show the photometric redshift estimation for a sample of 170 quasars detected in the ALHAMBRA fields (Matute et al. 2012). The analysis was performed with LePhare, using all of the 23 band filters and 50 templates (see more details in Chapter 5). In the paper they found a photo-z uncertainty of 0.009 and that the fraction of outliers is 12.3% (for a threshold of $\frac{\Delta z}{1+z} > 0.15$).

Although the ANNz code, an empirical training set approach, performs slightly better when compared to other photo-z codes, we have chosen to work with two template fitting codes: BPZ (*Bayesian Photometric Redshift*) and LePhare (*PHotometric Analysis for Redshift Estimation*). This choice is related to the fact that both codes are widely employed by the members of the J-PAS Collaboration; thus using these codes enables us to obtain independent results and eventually compare them with the ones obtained by the other members.

BPZ and LePhare are very similar to each other with respect to the usage: they both receive as input a set of template SEDs and compute the χ^2 minimization between the data and the set of templates. Both codes also allow interpolations between pairs of templates; here, we have allowed interpolation between the quasar templates, but not between the star templates.

LePhare allows adding the contribution of the emission lines and includes various prescriptions to correct for galactic extinction, with the possibility of using different extinction laws, and it is also already set to work with galaxies, stars and quasars. However, LePhare has a size limitation with respect to the dimension of some vectors and does not provide all the redshift likelihoods for stars, so using this code may not be feasible when we have great photometric precision ($\Delta z \sim 0.001$) in a photometric system composed of several filters ($N_f > 25$). This means that, for now, including priors for the classification of quasars and stars in a mixed sample with many narrow-bands is not possible with LePhare.

BPZ, on the other hand, works better with galaxies, and as a result some adaptations were necessary in order to include templates for quasars and stars when using mixed catalogs. We actually run BPZ twice with the mixed catalog, first only with the list of quasar templates and then only with the star templates; in the end, we perform a joint analysis with the results, where the corresponding minimum χ^2 gives us the classification of the object.

A χ^2 fitting procedure can perform better when additional constraints are applied. Both LePhare and BPZ allow the inclusion of a prior information on the redshift distribution, with the purpose of avoiding catastrophic errors which place an extragalactic source (galaxy, quasar, etc) at an unrealistic redshift or fitting the object with an infrequent template. In particular, BPZ incorporates the priors within a Bayesian framework (see Chapter 4). Following the discussion in Benítez (2000), the probability of an extragalactic source with apparent magnitude m and color C having a redshift z is given by

$$p(z|m,C) = \frac{p(z|m)p(C|z)}{p(C)} \propto p(z|m)p(C|z) \tag{2.18}$$

where $p(z|m)$ brings in the prior knowledge of the magnitude redshift distribution, $p(C|z)$ is the redshift likelihood and $p(C)$ is just a normalization.

BPZ actually uses two methods for the photo-z estimation: the maximum likelihood method picks the maximum likelihoods over all the likelihoods in redshift and types, and the Bayesian method picks the maximum of all of the likelihoods after weighting by the prior probabilities. In the absence of priors, these methods should converge.

In Chapter 5, we investigate some sets of templates and compare the performances of BPZ and LePhare for simulated samples of quasars and stars. We also compare the classification of quasars and stars for mixed mock catalogs before and after the inclusion of Bayesian priors, using only BPZ.

Chapter 3

Mock catalogs for quasars and stars

In the previous chapter we bring up a discussion about the main features of quasars and stars, which prove to be very different objects.

In this chapter, we develop the formalism for the construction of mock catalogs for quasars and stars. Such catalogs must take into account both the distribution of quasars and AGNs as a function of redshift, and the angular distribution of stars of different types on the sky. These density distributions are related to the luminosity function, which essentially returns the number of objects as a function of luminosity (or, equivalently, of magnitude).

As we have seen, quasars are members of the family of active galactic nuclei (AGNs). The realization that all members of this family contain a supermassive black hole at their center means that AGNs and, in particular, quasars are likely to share intrinsic characteristics. For this reason, the luminosity function for quasars is the same as the one derived for AGNs, up to some constant geometrical factors. In the analysis, we use the luminosity function of Croom et al. (2009).

As for stars, we obtain their angular distribution from an empirical model for the stellar population of the Milky Way called Besançon Model (Robin et al. 2003).

In the end, we have the expected number of quasars as a function of redshift and magnitude, and the number of stars per square degree as a function of types and magnitude. We shall use these number densities both to select the objects to compose the mock catalogs, and as our Bayesian priors (see Chapter 4).

3.1 Luminosity functions

The bolometric luminosity is the total energy rate output integrated over all wavelengths. However, since the Earth atmosphere is opaque to most wavelengths, it is difficult to determine precisely the bolometric luminosity. Instead, astronomical luminosities are measured in one or more wavelength bands.

The magnitude is a measure of the brightness of an object viewed through a certain filter. Magnitudes are measured in a logarithmic scale, because the magnitudes perceived by the human eye scale roughly logarithmically with the radiation flux.

The apparent magnitude is determined by the apparent brightness as observed with no consideration to how the distance affects the observation. It is convenient to measure the magnitudes in the AB magnitude system, based on flux measurements calibrated in absolute units. In cgs units, the AB magnitude is defined as

$$m_{AB} = -2.5\log_{10}(f_{\nu}) - 48.6 \tag{3.1}$$

where f_{ν} is the flux in units of erg s^{-1} cm^{-2} Hz^{-1}. To give a quantitative idea of the magnitude system, some very bright objects, such as the Sun, the Moon, and the Vega star, can have magnitudes of 0 or even negative values; the faintest star the naked human eye can perceive has about +6 magnitudes; and very faint objects have magnitudes greater than +6.

The absolute magnitude M of a star is the magnitude the object would have if it was placed at a distance of 10 parsecs (pc) from Earth. So, by considering objects at a fixed distance, astronomers can compare the real intrinsic) brightness of different objects. To convert the observed (apparent) magnitude to an absolute magnitude, we need to know the luminosity distance to the object. These calculations are made using the distance modulus μ:

$$\mu \equiv m - M = 5\log_{10}\left(\frac{D_L}{1\ \mathrm{Mpc}}\right) + 25 \tag{3.2}$$

where we have converted the distance D_L in units of Mpc. For instance, the Sun has an apparent magnitude of -26.8 and an absolute magnitude of 4.83 in the V-band. The difference between apparent and absolute magnitude is independent of the filter choice, and it equals the distance modulus if no extinction is present. In fact, a K-correction must be applied, because the observed radiation was shifted from its rest-frame (see section §2.2).

The luminosity function provides the number density of a class of objects as a function of the luminosity (or, equivalently, of the absolute magnitude). Denoted by $\Phi(M)dM$, the luminosity function defines the number counts of galaxies with absolute magnitude in the interval $[M,\ M+dM]$. The total density of galaxies (per Mpc^3) is given by

$$\Phi = \int_{-\infty}^{+\infty} dM \Phi(M). \tag{3.3}$$

Accordingly, $\Phi(L)$ provides the number density of galaxies with luminosity in the interval $[L,\ L+dL]$.

The study of the luminosity function is very useful for understanding the formation, evolution and properties of galaxies. However, the task of determining the luminosity function depends on the measurement of the flux of the galaxy, and consequently, a precise measure of its distance. It also depends on a representative sample of galaxies, and for this a large volume must be surveyed. In addition, there is the *Malmquist bias*—an effect of over representing luminous galaxies, which are visible at larger distances, in flux-limited surveys—and hence a correction is needed.

The distribution of galaxies is represented to good approximation by the Schechter luminosity function (1976):

$$\Phi(L) = \left(\frac{\Phi_*}{L_*}\right)\left(\frac{L}{L_*}\right)^{\alpha} \exp\left(-\frac{L}{L_*}\right) \tag{3.4}$$

where Φ_* has units of number density and specifies the normalization of the distribution, L_* is a characteristic luminosity above which the distribution decreases exponentially, and α is the slope of the luminosity function for small L, typically assuming values of $-1.5 < \alpha < -1$. The galaxy luminosity function may have different parameters for different populations and environments; it is not a universal function.

This relation is derived from the Press-Schechter theory, which predicts the number density of virialized halos of a certain mass. Therefore, the mass function and the luminosity function are very correlated: the greater the luminosity of the galaxy, the greater the mass of the halo where it is found.

The equivalent in magnitude of Eq.3.4 is

$$\Phi(M)dM = \Phi(L)\left|\frac{dL}{dM}\right|dM =$$
$$0.4\ln(10)\Phi_* 10^{-0.4(1+\alpha)(M-M_*)}\exp\left[-10^{-0.4(M-M_*)}\right]dM \tag{3.5}$$

where the following substitutions are applied:

$$\Phi(M)dM = \Phi(L)d(-L) \tag{3.6}$$

$$M - M_* = -2.5\log\left(\frac{L}{L_*}\right). \tag{3.7}$$

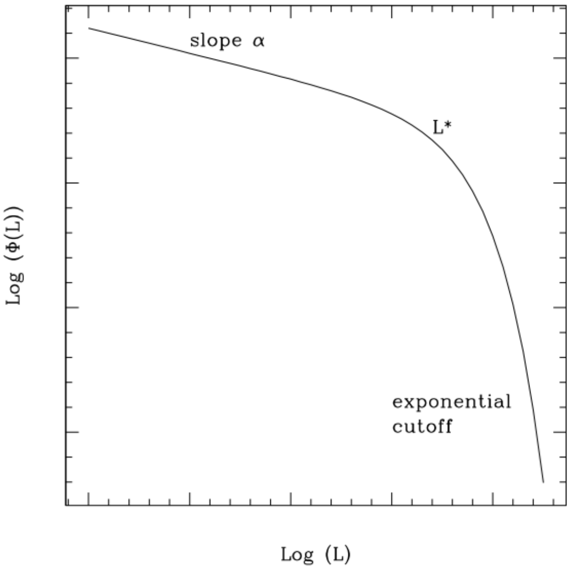


Figure 3.1: Schematic Schechter luminosity function, showing the slope α, the characteristic luminosity L_* and the exponential cut-off for bright stars.

Typical values derived from B-band measurements are:
$\phi_* = (1.6 \pm 0.3)$ x10^{-2} h^{-3} Mpc^{-3}; $M_B^* = -19.7 \pm 0.1 + 5\log 5$; $\alpha = -1.07 \pm 0.07$; and $L_B^* = (1.2 \pm 0.1)$ h^{-2}x10^{10} $L_\odot$.

Although the Schechter function seems an adequate representation for the total distribution of galaxies, different spectral types are expected to have distinct luminosity functions and, thus, deviations from the Schechter function are common.

3.2 Luminosity function for quasars

The quasar luminosity function (QLF) provides the volumetric density of quasars per unit magnitude as a function of the redshift. In other words, it can be used to describe the evolution of the number density of AGNs with time.

The QLF is often parametrized by a standard double power-law form (Boyle et al. 2000) and we are adopting the following expression from Croom et al. (2009)

$$\Phi_{0,q} \equiv \frac{d^2\bar{N}_q}{dmdV} = \frac{\Phi_0}{10^{-0.4\text{x}2.33\text{x}[M(m,z)-M_g^*(z)]} + 10^{-0.4\text{x}0.41\text{x}[M(m,z)-M_g^*(z)]}} \tag{3.8}$$

where $\Phi_0 = 1.45\text{x}10^{-6}$ Mpc^{-3}; $M_g^*(z) = -22.2 - 2.5(1.44z - 0.32z^2)$ is the break magnitude in the g-band; $M(m,z) = m - 25 - 5\log_{10}(D_L)$ is the absolute magnitude corresponding to the apparent magnitude m for an object at redshift z, and D_L is the luminosity distance in Mpc [we have adopted the ΛCDM model with parameters $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$]; m is the apparent magnitude which must be K-corrected through $m_k(z) = m - K(z)$, where

$$K(z) = -2.5(1+\alpha_\nu)\log_{10}(1+z). \tag{3.9}$$

We assume a power-law slope of $\alpha_\nu \simeq -0.5$. The K-correction is necessary because we are comparing the fluxes measured through a single filter for quasars at different redshifts. Therefore, applying this correction is equivalent to obtaining the flux measurement in the rest frame of the quasar.

These parameters were obtained for a sample of quasars within the range $0.4 < z < 2.6$ (Croom et al. 2009) and are defined for the g-band. Here we have the particular interest of computing the QLF as a function of the i-band magnitude, because when the i-band of J-PAS is used as the detection band, it goes deeper than the g-band. Then, we must apply the following conversion:

$$M_i(z) = M_g(z) - 2.5\alpha_\nu \log\left(\frac{4770\,\mathring{A}}{7625\,\mathring{A}}\right). \tag{3.10}$$

Note that Eq. 3.9 and 3.10 are expressed in a very similar format. However, they account for different effects: the first one is due to the shift of the observed radiation from its rest frame emission, while the second is due to a shift in wavelength and relates the central wavelengths of two band filters.

The QLF as a function of the absolute magnitude at different redshifts is shown in Fig.3.2. The x-axis was inverted because brighter objects have smaller magnitudes (or higher fluxes).

In Fig.3.2, we can notice that at a given redshift the number of quasars decreases with brightness. This effect can be attributed to the hierarchical structure formation theory (S. White 1996), in which the luminosity and lifetime of the quasar are governed by the central black hole mass and the supply of cold accreting gas. Since the abundance and evolution of supermassive black holes are linked to the evolution of the mass-function of the galactic halo, we expect the luminosity function to have a behavior similar to that of the mass-function. In this context, the brightness would be linked to the enclosed mass; since there are few very massive halos (or equally, few very massive black holes), the spatial density of very bright objects will also be low

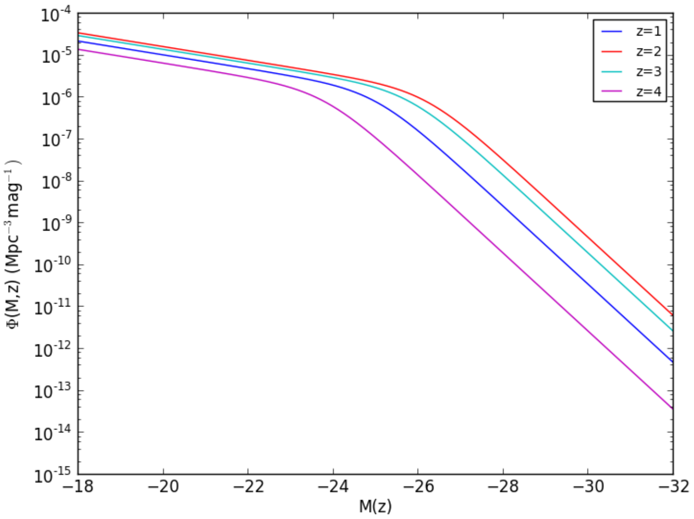


Figure 3.2: Quasar luminosity function at different redshifts.

Objects observed through our survey may have two sorts of magnitude limitations. Since we deal with flux-limited surveys, we must apply a cut in the apparent magnitude, which we take to be $i < 23$. This means that only objects brighter than a magnitude of 23 in the i-band have any chance to be detected. It is also prudent to apply a cut in absolute magnitude, which for us will be $M_i < -22$, because the host galaxy may overshadow the light of quasars fainter than this. This cut affects mainly quasars at low redshifts.

In Fig.3.3 we show the volumetric density of quasars in terms of a limiting i-band apparent magnitude ($i < 23$) and a limiting absolute magnitude ($M_i < -22$)

as a function of the redshift. Note that the volumetric density $n(z)$ is given simply by integrating $\Phi_{0,q}$ up to the corresponding limiting magnitude.

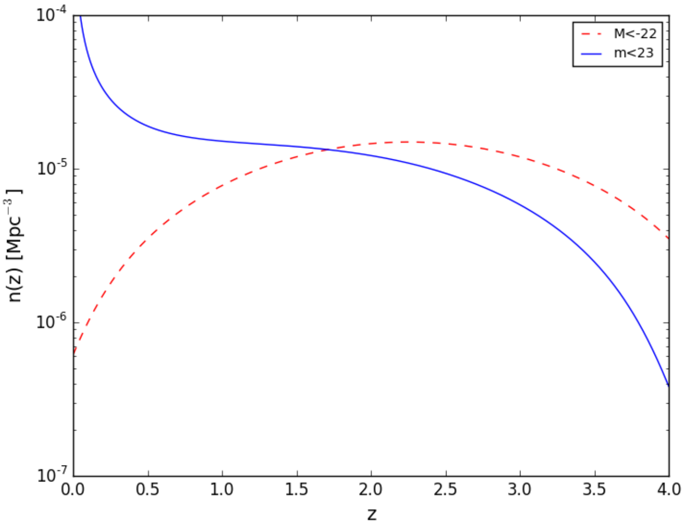


Figure 3.3: Volumetric density of quasars for limiting $i < 23$ apparent magnitude (blue solid line) and $M_i < -22$ absolute magnitude (red dashed line), as a function of redshift.

Taking these two limiting magnitudes into account, the luminosity function for quasars becomes $\Phi_q(m,z) = \Phi_{0,q}(m,z)[\theta_H(M_{lim} - M(m,z)) + \theta_H(m_{lim} - m)]$, where θ_H is the Heaviside step function. So, the number of quasars in an area of solid angle $\Delta\Omega$, in the magnitude bin m_i can be defined as

$$\frac{d\bar{N}_q}{dz}(m_i,z) = \frac{dV}{dz}\int_{m_i}^{m_i+\Delta m_i} dm\Phi_q(m,z) \tag{3.11}$$

where dV/dz is the comoving volume corresponding to an area of solid angle $\Delta\Omega$ between the redshifts z and $z+dz$.

We define the total number of quasars with magnitude m_i in an area of solid angle $\Delta\Omega$ at any redshift as

$$\bar{N}_q(m_i) = \int_0^\infty dz\frac{dV}{dz}\int_{m_i}^{m_i+\Delta m_i} dm\Phi_q(m,z). \tag{3.12}$$

From now on, the "i" in m_i denotes the index of the magnitude bin.

Finally, the total number of quasars expected in this area is

$$\bar{N}_q = \int_o^\infty dz\frac{dV}{dz}\int_{-\infty}^{m_{lim}} dm\Phi_q(m,z) \equiv \sum_{i=1}^{lim}\bar{N}_q(m_i). \tag{3.13}$$

Considering a perfect selection of objects, we find that a large-area flux-limited ($i < 23$) survey could yield about 1.7x10^6 quasars up to $z = 4$.

We can also define the probability of finding a quasar with apparent magnitude m_i at redshift z_j as

$$P^q_{ij} = \frac{1}{\bar{N}_q} \int_{z_j}^{z_j+\Delta z} dz \frac{d\bar{N}_q(m_i, z_j)}{dz}. \tag{3.14}$$

So the probability of finding a quasar with magnitude m_i at *any* redshift is given by $\sum_j P^q_{ij} = 1$.

3.3 Besançon Model

The number density (per unit solid angle) of stars has an angular dependence and a dependence on the apparent magnitude, but no redshift dependence. So, on a given direction $\hat{n}$ on the sky, the luminosity function of a star of type s is

$$\Phi_s(m, \hat{n}) = \frac{d\bar{N}_s}{dm d\Omega}. \tag{3.15}$$

We denote the total number of stars of type s on the field of area $\Delta\Omega$ centered around the direction $\hat{n}$, whose reference magnitudes lie within the bin m_i, by:

$$\bar{N}_s(m_i, \hat{n}) = \Delta\Omega \int_{m_i}^{m_i+\Delta m_i} dm \, \Phi_s(m, \hat{n}). \tag{3.16}$$

The luminosity function of stars, which is truly their number densities, can be estimated in some regions of the sky through direct observations (using, for example, the "Spectra of Everything" catalog of point-like sources, from SDSS), or through empirical models, such as Besançon Model (Robin et al. 2003) and TRIdimensional modeL of thE GALaxy (TRILEGAL; Girardi et al. 2005).

These models are stellar population synthesis codes for simulating the dynamical and evolutionary aspects of the Milky Way. Of course, each code has its own specificities, however both run Monte Carlo simulations to predict the expected number density of stars in every direction of the sky in the bands of several photometric systems. The default values in the codes correspond to the respective paper calibrations, but they can also be changed from simulation to simulation, the most relevant ones being the photometric system, the coordinates and total area of the field of interest, and the extinction calibration.

In particular, TRILEGAL allows more constraints on the input parameters, especially on the components of the Galaxy (thin and thick disk, halo and bulge), and it also provides more detailed information about the physical parameters of the simulated stars, such as age, effective temperature, luminosity, surface gravity, and the apparent photometry for the selected system. So, from the output parameters one may produce a Hess diagram, which translates the distribution of stars in H-R diagrams into a color-magnitude diagram, to obtain the stellar counts. However, a non-specialist in stellar dynamics might find this process unpractical, because it does not provide a direct measure of the star counts with magnitude.

Besançon, on the other hand, allows the selection of the count numbers as a direct function of the luminosity class (from supergiants to T Tauri stars, including main sequence stars and white dwarfs). Therefore, this model is more adequate for the purposes of our analysis, and we use the Besançon model to draw the statistical distributions of stars at a certain area on the sky using the i band as the detection band. This model provides the stellar counts in terms of the luminosity class and also of the subclasses, nevertheless we have still opted for simplifying the classification into eight "broad" main types.

Gao et al. (2012) study the ability of these models to reproduce the star counts of SDSS data at the north Galactic pole. They find that both Besançon and TRILEGAL reproduce adequately the SDSS luminosity function; however, they fail in reproducing the stellar distributions at high latitudes, with medium deviations reaching from 20% to 53% in the case of Besançon, and of the order of 26% in the case of TRILEGAL. Although these deficits are significant, our analysis is restricted to an area of 10 deg^2 around the latitude of 70°, thus we rely on the stellar distributions provided by Besançon.

At a first moment, we considered a central point of coordinates (RA, dec) = (200, 70) as our reference point on the sky. Comparing the outputs of these models in 1 deg^2 around this central coordinate, we obtained that Besançon retrieves 2719 stars, while TRILEGAL retrieves 2793 stars. Those are main sequence stars plus white dwarfs located on the thin and thick components of the disk. The differences in the star counts are of the order of 2.6%, and thus we are confident that we are not loosing too much information by using Besançon instead of TRILEGAL.

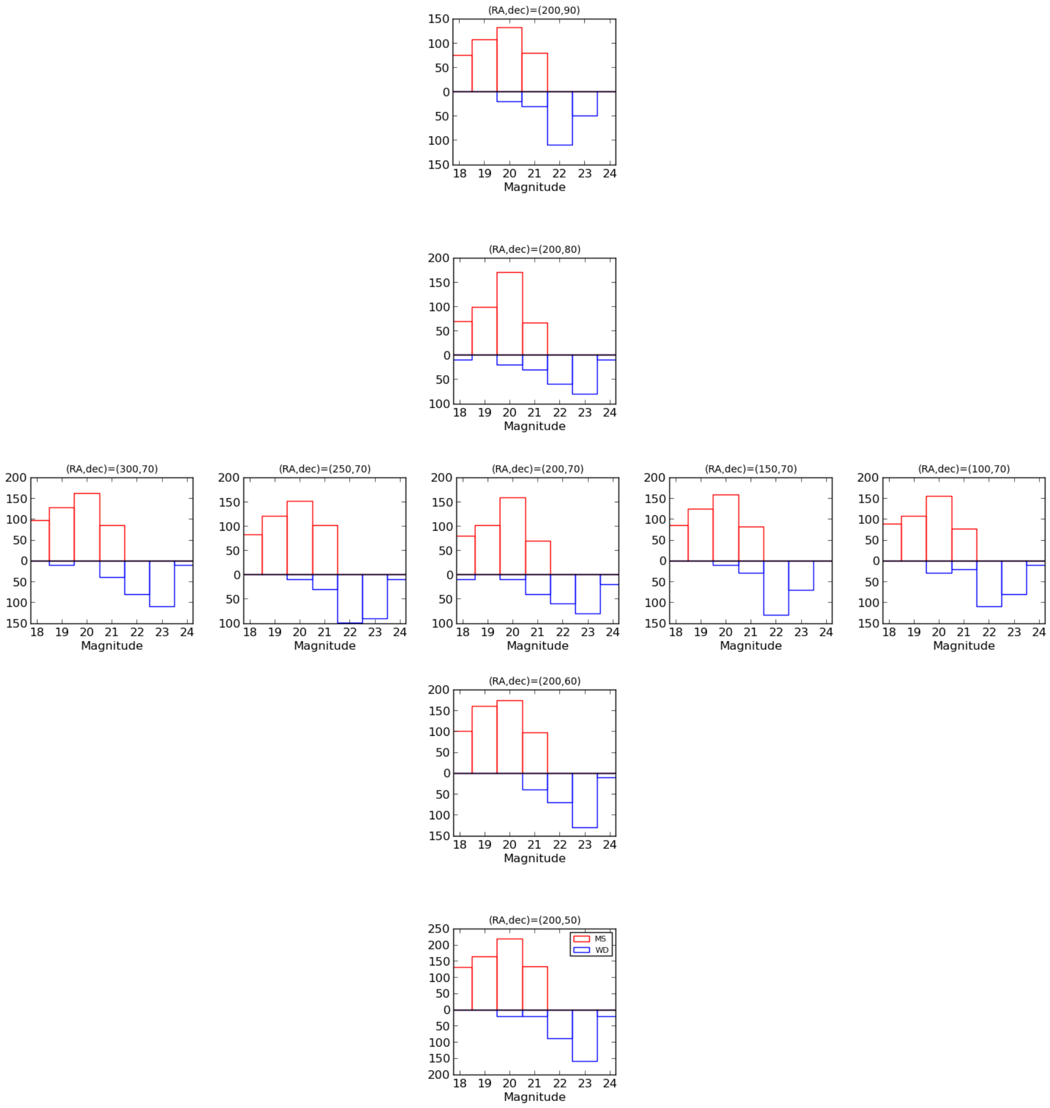


Figure 3.4: Count of main sequence stars (red, upward bars) and white dwarfs (blue, downward bars) per square degree as a function of the apparent magnitude in the i-band at different coordinates on the sky. These counts were obtained with the Besançon Model. In order to get a better visualization, the counts of white dwarfs have been multiplied by 10. Note that the count numbers vary more along the declination (for a given fixed right ascension) than along the right ascension (for a given fixed declination).

3.4 Simulating fluxes

Photometric techniques have been widely employed in astronomy and cosmology. However, these techniques do not measure exactly the flux of the object, because the emitted photons may suffer different sorts of interactions in the path between the source and the instrument, resulting in a probability distribution for the counts of photons, even in the (hypothetical) case of an efficiency of detection of 100%. Therefore, within this framework it arises the concern about properly understanding the different sources of noise that could reduce the quality of the image.

Currently, one of the most employed observational methods is the photometry with charge-coupled devices (CCDs), which works through a physical effect similar to the photoelectric effect and, therefore, converts the incident photons into a charge that can be measured and recorded. Typically, a measurement with these devices returns the incident count of photons in the detector after a time $\Delta\tau$ of exposure:

$$C = A\Delta\tau \int d\lambda \frac{f_\nu(\lambda)T(\lambda)}{h\lambda} \tag{3.17}$$

$$C = A\Delta\tau \int d\lambda \frac{f_\lambda(\lambda)T(\lambda)\lambda}{hc} \tag{3.18}$$

where A is the effective area of the primary mirror of the telescope, and $T(\lambda)$ is the total transmission function of the filter system (which includes the atmospheric transmission, the efficiency of the detector and the telescope, and transmission of the filter used on the observation). We shall use Eq. 3.17 or 3.18 whether the spectrum of the object is given in units of erg s^{-1} cm^{-2} Hz^{-1} or in units of erg s^{-1} cm^{-2} Å^{-1}, respectively.

Suppose that N is the count of photons that were detected by our instrument. Since the photons produced by an astrophysical object follow a Poisson distribution, there are random variations in the number of photons that strike the CCD each second, and the statistical noise of this measure is given by $\sqrt{N}$. However, there are also other sources of noise that should be taken into account, the main ones being the sky background and the readout noise in the CCD.

Then we can define the signal-to-noise ratio as

$$\frac{S}{N} = \frac{C}{\sqrt{C + \sigma_{sky}^2 + \sigma_r^2}} \tag{3.19}$$

where C is the count of photons, σ_s is the sky background, given by

$$\sigma_s^2 = A\Delta\tau P^2 \int d\lambda \frac{f_{sky,\nu}(\lambda)T(\lambda)}{h\lambda} \tag{3.20}$$

where P is the binned pixel size; and σ_r is the CCD readout noise, given by

$$\sigma_r^2 = n_\tau \mathrm{x} N_e^2 \mathrm{x} N_P \tag{3.21}$$

where n_τ is the number of exposures for each filter, N_e is the readout noise (electrons/pixel) and N_P is the number of pixels (which depends on the factor by which the filter imaging will be binned).

Our initial SDSS catalog contains about $10,000$ quasars and $22,000$ stars from the *Spectra of Everything*. Starting from the spectra of our initial sample, we constructed synthetic fluxes for three different filter systems (which will be specified in the next section) by adding the level of signal-to-noise (SNR) expected for each system. In this case, we are interested in the flux measured in each filter of the new system, which is obtained by simply convolving the SDSS spectrum with the filter transmission function:

$$f_{\alpha,k} = \frac{1}{c}\frac{\int d\lambda T_\alpha(\lambda) S_{\nu,k}(\lambda)\lambda^2}{\int d\lambda T_\alpha(\lambda)} \tag{3.22}$$

where $f_{\alpha,k}$ is the flux of the object k measured in the filter α, $T_\alpha(\lambda)$ is the transmission function of the filter α and $S_{\nu,k}$ is the SED of the object (which in the case of SDSS spectra is given in units of 10^{-17} erg s^{-1} cm^{-2} Hz^{-1}).

Now that we have the flux of the quasars and stars in the filters of different systems, we will consider that the signal in the detector will be proportional to this flux, where the constant of proportionality is given by calibrating the SED of the object with some broad-band magnitude (e.g., we have chosen the i-band).

If we assume the signal of the object to be subdominant when compared to other sources of noise such as the sky background and the CCD read noise (which, in general, is true for most objects except very bright stars), then we have a simplified expression for the signal-to-noise ratio expected for this object. In the filter α, in terms of the limiting magnitude at some level of significance (e.g. 5σ) we have:

$$\left.\frac{S}{N}\right|_{\alpha(k)} = 5\mathrm{x}10^{-0.4(m_{\alpha,k}-m_\alpha^{5\sigma})} \tag{3.23}$$

where $m_{\alpha,k}$ is the AB magnitude of the object k at the filter α and $m_{\alpha}^{5\sigma}$ is the limiting AB magnitude at the filter α at 5σ.

Finally, the simulated flux $f_{\alpha,k}^{S}$ for each object at the filter α can be obtained by adding the expected level of noise given by Eq. 3.23. Therefore,

$$f_{\alpha,k}^{S} = f_{\alpha,k}[1 + G(0, \Delta SNR^{-1})] \tag{3.24}$$

where $G(0, \Delta SNR^{-1})$ is a random number selected from a Gaussian distribution centered on zero and with dispersion $(\Delta SNR)^{-1}$. Note that

$$\frac{1}{\Delta SNR} = \sqrt{\left(\frac{1}{SNR}\right)^2 + \left(\frac{1}{SNR_D}\right)^2} \approx \frac{1}{SNR} \tag{3.25}$$

where SNR_D is the signal-to-noise ratio coming from the data (original spectrum). In the case of J-PAS filters and exposure times, and SDSS spectra, the SNR_D is much larger than SNR, and so its contribution to the noise is negligible.

In Fig.3.5 and 3.6 we show the ideally high SNR fluxes (red dots) and the simulated fluxes (black dots) with the J-PAS filter system for templates of type-I (Vanden Berk 2001) and type-II quasars (Zakamska et al. 2003), respectively.

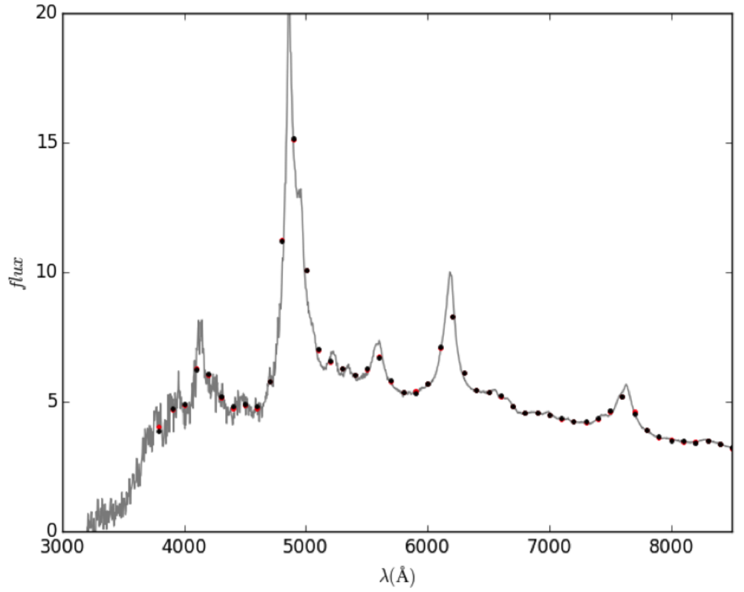


Figure 3.5: Ideally high SNR fluxes (red dots) and simulated fluxes (black dots) at the J-PAS filter system for the Vanden Berk template of type-I quasar at $z = 3$.

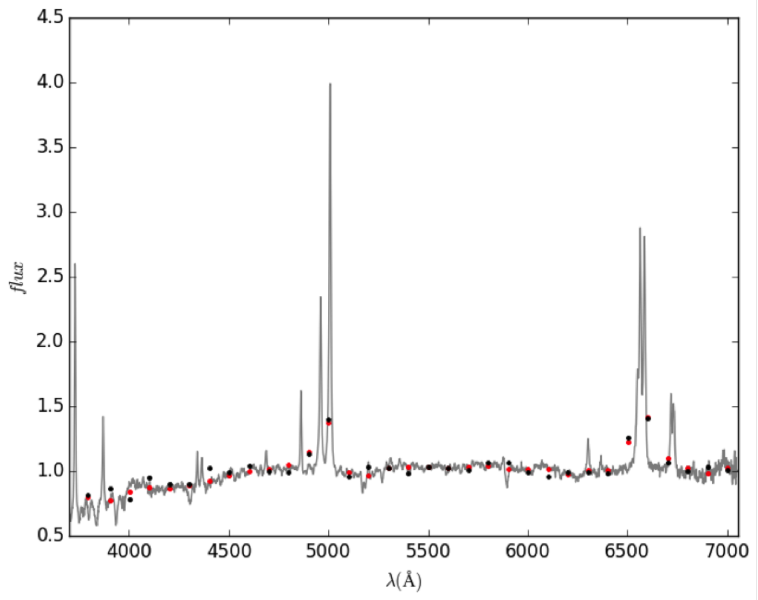


Figure 3.6: Ideally high SNR fluxes (red dots) and simulated fluxes (black dots) at the J-PAS filter system the Zakamska template of type-II quasar at rest frame.

3.5 Data set

Our sample contains the average number of quasars and stars expected to be found in 1 deg^2. The average number of quasars comes directly from the luminosity function (§ 3.2). As for the stars, we selected an area of about 10 deg^2 around our reference point of coordinates (RA, dec) = (200, 70), and the Besançon Model has provided us with the average number of stars expected to be found in 1 deg^2 at this region of the sky. So, in the end, we have 218 quasars selected from a SDSS catalog of about $\sim 10,000$ quasars (Fig.3.7), in the redshift range $[0.5, 3.5]$ and with i-band magnitudes in the range $[17, 23]$, and $2,719$ stars selected from the SDSS *Spectra of Everything* (Fig.3.8), which includes the programs “merged48” and “merged73” and is supposed to provide a fair sample of about $22,000$ stars with reddening-corrected i-band PSF magnitudes brighter than 19.1 over an area of approximately 240 square degrees. For simplicity, the selected stars were divided into 8 different types (O, B, A, F, G, K, M and WD—white dwarf) and they also cover the i magnitude range $[17, 23]$.

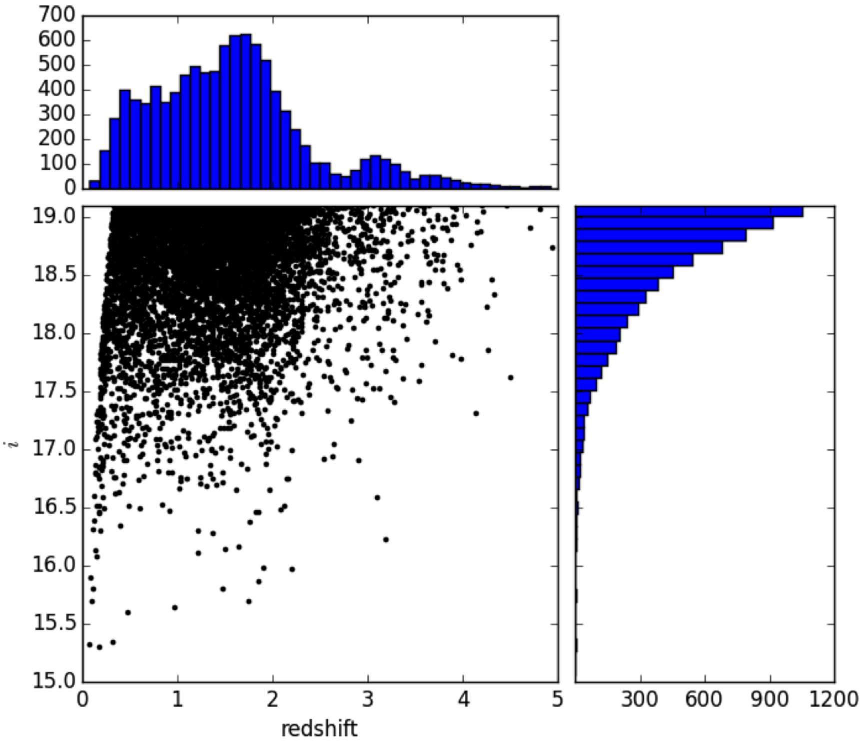


Figure 3.7: Magnitude-redshift distribution for the catalog of about $10,000$ quasars from SDSS from which our sample of quasars was selected.

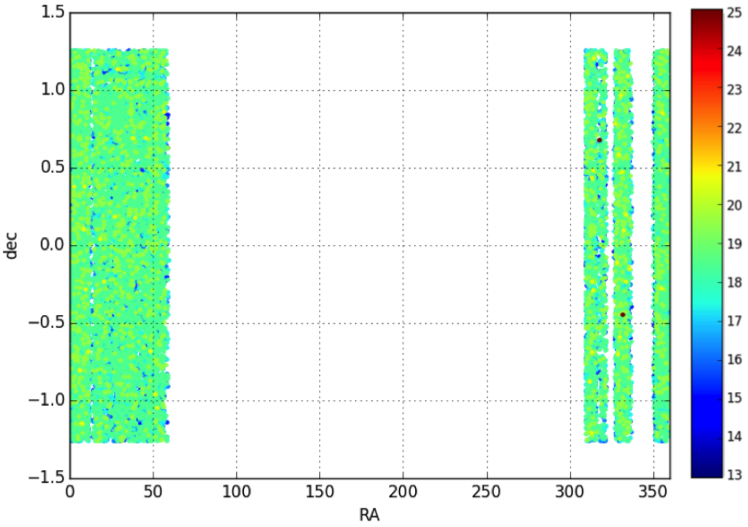


Figure 3.8: Stars from the SDSS *Spectra of Everything*. The color code indicates the magnitude in the i-band.

The original SDSS sample of quasars and stars goes only up to magnitude $i \sim 19$. In order to extend this magnitude to a fainter limit ($i < 23$), we have put the original bright spectra to fainter magnitudes by introducing into the spectra the level of noise expected for each photometric system while simulating the fluxes. Our final sample is characterized in the following.

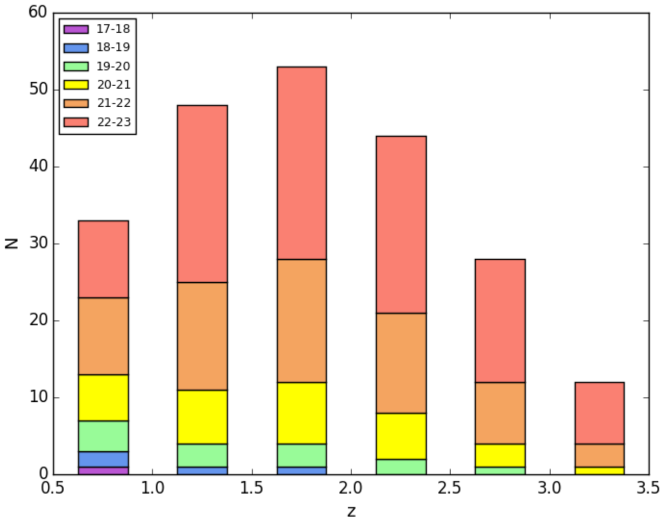


Figure 3.9: Histogram of the redshift distribution as a function of the *i*-band magnitude for our sample of quasars.

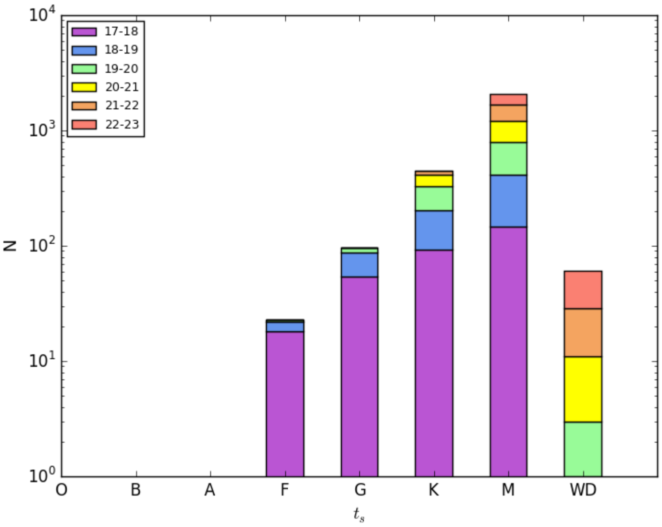


Figure 3.10: Histogram of the distribution of types as a function of the *i*-band magnitude for our sample of stars. The number counts are in log scale.

The number densities in Fig.3.9 and 3.10 reproduce the properties of quasars and stars that we expect to find, on average, on 1 deg^2 at these ranges of redshift and magnitude. These histograms represent the way by which we have selected the objects to compose our sample and as our Bayesian priors. From the Besançon

Model, we do not expect to find any stars of the types O, B and A in the region of the sky chosen to obtain the average number of stars.

In Fig.3.11, we show the real spectra of some of the quasars in our sample.

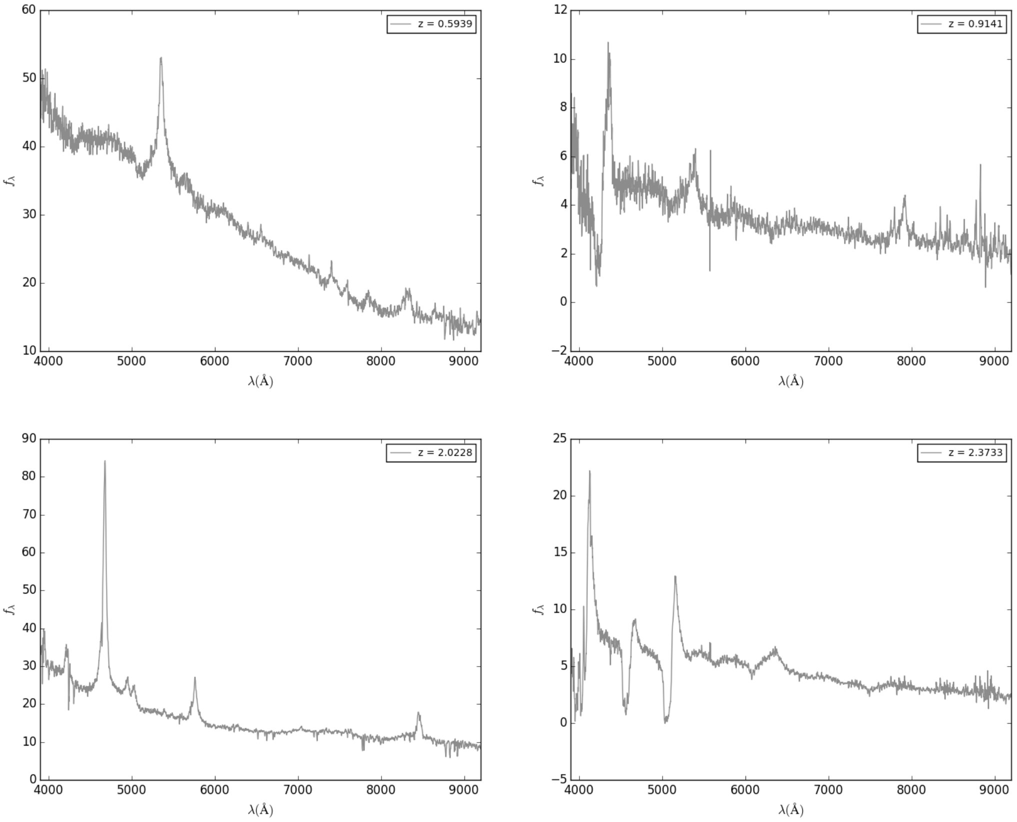


Figure 3.11: Real spectra of quasars from our catalog. The fluxes are in units of 10^{-17} erg s^{-1} cm^{-2} Hz^{-1}. From left to right and from top to bottom, the quasars have the following apparent magnitudes in the *i* band: 17.2, 22.1, 22.1 and 22.9. Upper left panel: This quasar has an extremely blue continuum. Upper right panel: This quasar presents an absorption line at $\lambda \sim 4100$ Åfollowed by a strong emission line. Lower left panel: Typical quasar spectrum, with the broad emission lines and several narrow absorption lines. Lower right panel: BAL quasar.

Chapter 4

Bayesian priors

Presently, galaxy catalogs with accurate redshifts and large areas (> 1000 deg^2) reach only up to $z \sim 0.7$. BOSS has built a catalog of Ly-α absorption systems that spans $1.5 \lesssim z \lesssim 2.5$, but there are many statistical and systematic uncertainties associated with this technique—and there are no overlapping galaxy catalogs that could calibrate those results. On the other hand, based on estimates computed with existing luminosity functions, a reasonably complete (volumetric densities $\gtrsim 10^{-4}$ h^3 Mpc^{-3}) catalog of quasars could span redshifts over a continuous interval $0.5 \lesssim z \lesssim 4$. Such a catalog would allow us to measure the clustering of matter over the largest distances of any survey by far.

However, instruments which rely on imaging to pre-select spectroscopic targets usually miss most quasars and AGNs, due to the redshift dependent degeneracy of quasars with stars in color-color diagrams.

In Fig.4.1 and 4.2, the stellar locus is represented by the black dots and the colored dots (in color version) represent the quasar loci at different redshifts. Quasars with $z > 2.5$ leave the stellar locus and begin to be distinguishable from stars. However, in the range $1.5 < z < 3.0$ most of the quasars overlap the stellar locus and then we can identify a sample contamination. York et al. 2000 show that quasar colors approach those of early F and late A type stars at $z \sim 2.8$. However, in the right panel of Fig.4.2, we identify that in our sample the contamination comes mainly from G-type stars and white dwarfs.

In any case, it is clear that there is some degeneracy between quasars and stars in color-color diagrams (also present in color-magnitude diagrams), especially at intermediate redshifts ($z \sim 1.5 - 2.5$), where quasars overlap with the stellar locus.

For instance, the SDSS photometric sample (the largest in existence today) has $\sim 10^6$ quasar candidates, but only $\sim 18\%$ of those had their spectra actually taken (Richards et al. 2009; Leistedt et al. 2013). The main reason for this low efficiency is that, with broad (*ugriz*) filters, quasars are easily confused with stars (and even with unresolved galaxies), hence the fibers in SDSS are only allocated to the brightest, most clearly distinguished quasars.

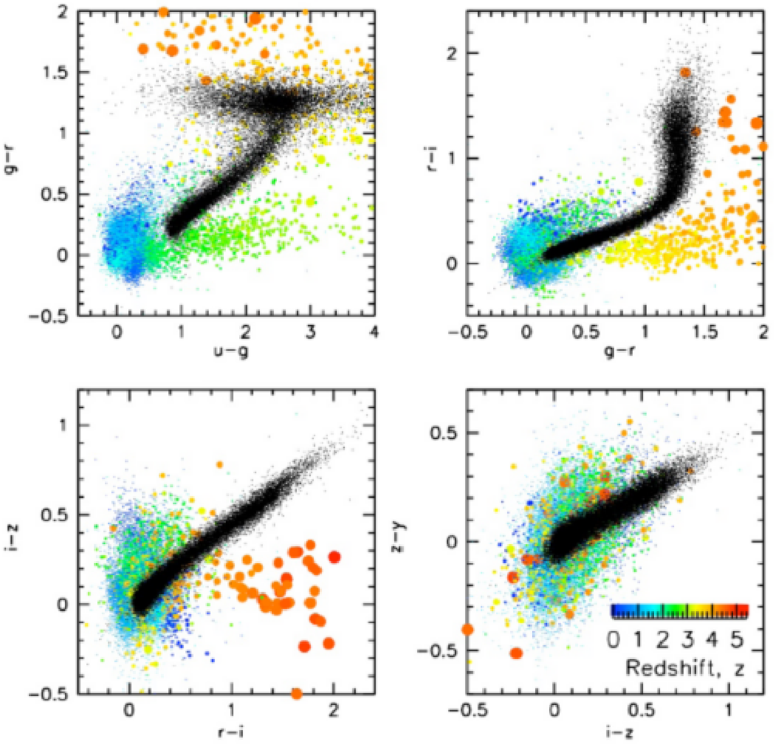


Figure 4.1: Color-color diagrams of known quasars from SDSS (colored dots) and stars (black dots) in the LSST photometric system. The quasars are color coded by redshift (LSST Science Book 2000).

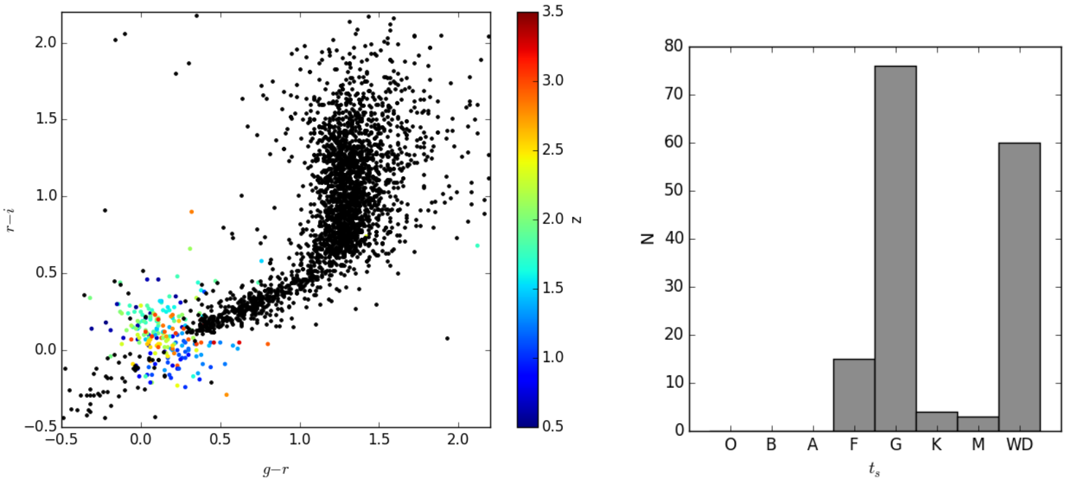


Figure 4.2: Color-color diagram for our sample of quasars (colored dot) and stars (black dots). The quasars are color coded by redshift. The right panel is a histogram with the types of stars that satisfy $|g-r| \leq 0.5$ and $|r-i| \leq 0.5$.

These issues are not restricted to broad-band imaging surveys; they must also be taken into consideration and properly treated in the context of medium/ narrow-band filters, in order to obtain good photometric redshifts and to minimize the contamination of quasar catalogs.

However, the situation is not necessarily so dramatic with narrow-band filters. Abramo et al. 2012 showed that a narrow-band survey of type-I quasars has the potential to identify those objects with high completeness and purity, and to determine their photometric redshifts with high accuracy. We provide a more detailed analysis in Chapter 5.

In this chapter, we discuss the Bayesian statistics, how it differs from the frequentism approach and finally introduce and develop our formalism of Bayesian priors for quasars and stars.

4.1 Bayesian statistics

The main rules of the probability theory are the sum and the product rules. The sum rule states that

$$P(X|H) + P(\bar{X}|H) = 1 \tag{4.1}$$

where $P(X|H)$ is the probability of a proposition X being true and $P(\bar{X}|H)$ is the probability of X being false, and H denotes the model (i.e., the "null" hypothesis) being tested.

The product rule states that

$$P(X,Y|H) = P(X|Y,H)P(Y|H) \tag{4.2}$$

where $P(X,Y|H)$ is the probability of both X and Y being true, $P(X|Y,H)$ is the conditional probability of proposition X being true given that the proposition Y is true, and $P(Y|H)$ is the probability of proposition Y being true.

In statistical analysis, there are two fundamental approaches: Bayesianism and frequentism. The fundamental difference between these approaches concerns to the definitions of probability (VanderPlas 2014). For frequentists, the probability is related to the frequencies of repeated events, while Bayesians relate the probability to the degree of uncertainty about a statement. This means that frequentists generally quantify the properties of data-derived quantities considering fixed model parameters, while Bayesians generally quantify the properties of unknown model parameters in light of the observed data.

The frequentism, as the name suggests, interprets the probability as the frequency of the outcome of a repeated experiment. In this view, any given experiment can be considered as one realization out of infinitely many possible realizations of the same experiment, so that each realization produces statistically independent results. In other words, the data are random and the hypotheses are fixed.

In frequentist statistics, the set of all possible outcomes constitutes the sample space of the experiment. Thus, an event will be a particular subset of the sample space and there are only two possibilities: the event occurs or it does not (the hypothesis is true or false, respectively). Measuring the probability of a given event means obtaining the relative frequency for that event. The relative frequency $P(Z|H)$ of an event Z occurring only converges to the real frequency as the number N_T of trials (or realizations) approaches infinity:

$$P(Z|H) = \lim_{N_T \to \infty} \frac{N_Z}{N_T} \tag{4.3}$$

where N_Z is the number of realizations in which the event Z occurred.

Therefore, in frequentist statistics the probability is associated with the data, instead of being derived for a hypothesis or parameter.

The Bayesian statistics has its fundaments on the Bayes' theorem, which can be derived from the product rule. Substituting X by the set of parameters Θ and Y by the observed data D, this theorem states that

$$P(\Theta|D,H) = \frac{P(D|\Theta,H)P(\Theta|H)}{P(D|H)} \propto P(D|\Theta,H)P(\Theta|H). \tag{4.4}$$

In Eq. 4.4, we identify the term $P(D|\Theta,H)$ as the likelihood, denoted by $\mathscr{L}(D|\Theta,H)$, which describes the probability of the data being observed given a set of model parameters. Similarly, we identify the term $P(\Theta|H)$ as the prior probability distribution $\pi(\Theta|H)$, which represents our belief of what the parameters of the model are likely to be before obtaining the data. $P(D|H)$ is the model evidence, and constitutes a normalization factor (Feroz 2008). Thus the Bayes' theorem becomes

$$P(\Theta|D,H) = \frac{\mathscr{L}(D|\Theta,H)\Pi(\Theta|H)}{\int \mathscr{L}(D|\Theta',H)\Pi(\Theta'|H)d\Theta'}. \tag{4.5}$$

The gain of using this equation is evident: all knowledge about Θ is summarized by the posterior probability distribution $P(\Theta|D,H)$.

Bayesian statistics supplies no special rule for determining the priors, since, in principle, it may reflect the researcher's subjective state of knowledge about the parameters Θ before the measurement is carried out. However, the priors must be and, in general, are constructed in the basis of a careful, objective analysis. For example, they may reflect the spectral distribution of galaxies in a certain region of the sky, and these number densities should reproduce with great fidelity the real ones. In this sense, the subjectivity of the researcher's degree of knowledge about the parameters is actually related to the approach chosen to derive these priors.

In the following section we develop the Bayesian priors for our analysis of separating quasars and stars in medium- and narrow-band filter surveys. Our priors reflect the distribution of quasars as a function of redshift and magnitude, obtained from the luminosity function for quasars, and the angular distribution of stars as a function of magnitude and types, given by an empirical model.

4.2 Bayesian priors

The main goal of this work is to determine how the different abundances of stars in different regions of the sky affect the quasar catalog—especially at the faint end of the survey. The approach chosen to treat this degeneracy is based on Bayesian priors. Assuming that the point-like object observed in the imaging survey can be classified either as a quasar or as a star (we can also extend this classification and consider other types of objects, such as galaxies), the method consists in computing the posterior probability of the object belonging to each class over a specific area of the sky, for objects of different magnitudes.

Here we will assume that the observations consist of the fluxes of the images taken with the narrow-band filters f_k ($k = 1, 2, \ldots, N_f$, where N_f is the number of filters). These fluxes can be used to compute the likelihood of this object being of type t (e.g., a quasar q at redshift z) through the following procedure (similar to a χ^2):

$$\mathscr{L}(\mathbf{f}|t,z) = \exp\left[-\sum_k \left(\frac{f_k - f_k^t}{\sigma_k}\right)^2\right] \tag{4.6}$$

where f_k^t is a theoretical template for an object of type t (at redshift z, if applicable), and σ_k[1] are the uncertainties in the fluxes f_k. In addition to these fluxes, we will also assume that each object has one well-measured "reference" (or "detection") broad-band magnitude. This magnitude (chosen to be the i-band) shall be used for the priors, but not for the likelihoods.

Within the Bayesian framework, our problem consists in calculating the posterior predictive distribution for stars and quasars at different redshift intervals. Suppose that we have a certain field on the sky, centered on the direction $\hat{n}$, with a (small) solid angle $\Delta\Omega$, and assume that we are considering only objects whose magnitudes lie within some interval $m_i < m < m_i + \Delta m$. We want to jointly predict both the number of stars and the number of quasars per unit redshift on that field (and for that magnitude interval), taking into account the fact that the different possible types of objects may not be clearly distinguished in an imaging survey – i.e., the measured fluxes of the objects may fit well both some star template and a quasar (or galaxy) template at some redshift.

In order to calculate these posterior probabilities, we first need the prior probabilities and the marginal likelihoods (or evidences) for these objects on that area of the sky, and for that magnitude interval. We must also notice that the distribution of stars is a function of magnitude and angular position, but not of redshift, whereas the distribution of quasars is a function of magnitude and redshift, but not of the angular position.

For quasars, this prior can be calculated from the QLF $\Phi_{0,q}(m,z)$ (Eq.3.8 in §3.2), which provides the volumetric density of quasars per unit magnitude, as a function of the redshift. One possibility is using the luminosity function defined by Croom et al. 2009 as shown previously, but other choices may also be used for high-redshift quasars, or for quasars of some particular sub-species whose SEDs possess some distinct features (such as broad absorption lines – BALs). Of course, the same procedure also applies for type-II AGNs, and galaxies.

In the case of stars, their density numbers $\Phi_s(m,\hat{n})$ have an angular dependence and a dependence on magnitude (but not on redshift) and can be estimated

---

[1] Here we assume that only the uncertainties in the fluxes are contributing for the uncertainty in the expression for the likelihood, but since we use a template-fitting code (BPZ), we could also suppose there is an additional uncertainty σ_T that comes from the fact that the chosen templates do not fit well the spectra of our sample. In this case, the denominator in Eq. 4.6 would be $(\sigma_k^2 + \sigma_T^2)$.

from the Besançon Model (as discussed in §3.3). In the literature, we can also find other sorts of empirical models for the stellar population in our Galaxy, such as TRILEGAL. Our stellar priors are drawn from a realistic real sample in which the stars are classified in main types and subclasses; however, we divide the stars into only eight "broad" types (O, B, A, F, G, K, M, and white dwarf—WD) because the subclasses are not relevant in our analysis for the moment.

Assuming for a moment that in a certain region of the sky we only have one type s of star, and one type q of quasar, and without any other source of information—just the reference magnitude, the prior probabilities that an object is a star or a quasar is given simply by their relative frequencies in that region:

$$\begin{aligned} \Pi(q|m_i,\hat{n}) &= \frac{\bar{N}_q(m_i)}{\bar{N}_q(m_i)+\bar{N}_s(m_i,\hat{n})}, \\ \Pi(s|m_i,\hat{n}) &= \frac{\bar{N}_s(m_i,\hat{n})}{\bar{N}_q(m_i)+\bar{N}_s(m_i,\hat{n})}. \end{aligned} \tag{4.7}$$

There are, of course, many different types of stars, quasars and AGNs. We may also wish to include unresolved galaxies, which would be classified as point sources and could potentially be typed as quasars. Hence, in a realistic situation we should consider many types t of objects, and their prior probabilities are also given in terms of their relative frequencies:

$$\Pi(t|m_i,\hat{n}) = \frac{\bar{N}_t(m_i,\hat{n})}{N_{tot}(m_i,\hat{n})} \tag{4.8}$$

where $N_{tot}(m_i,\hat{n}) = \sum_t \bar{N}_t(m_i,\hat{n})$ is the total number of objects of any type on that area of the sky, in the magnitude bin m_i. Of course, this procedure assumes that we know, to a good extent, the numbers for each one of these different types of objects, on a certain patch on the sky, and for all relevant magnitudes.

The priors above denote simply the blind guess that objects observed in some reference magnitude, at some area of the sky, are distributed according to their frequency. However, we have made no distinction between objects in our own galaxy (stars) and extragalactic objects (quasars, galaxies, etc.). Since extragalactic objects (type t_e) are also unevenly distributed with redshift, we would like to have priors for these objects as a function of the redshift as well. The same argument used above implies that these priors are:

$$\frac{d\Pi(t_e,z|m_i,\hat{n})}{dz}=\frac{1}{N_{tot}(m_i,\hat{n})}\frac{d\bar{N}_{t_e}(m_i,z)}{dz}. \tag{4.9}$$

In fact, we may also include stars in this unified description, using the following trick

$$\frac{d\Pi(t_s,z|m_i,\hat{n})}{dz}=\frac{1}{N_{tot}(m_i,\hat{n})}\bar{N}_{t_s}(m_i,\hat{n})\delta_D(z)=\frac{1}{N_{tot}(m_i,\hat{n})}\frac{d\bar{N}_{t_s}(m_i,\hat{n},z)}{dz} \tag{4.10}$$

where $\delta_D(z)$ is the Dirac delta function, or some other normalized narrow window-function narrowly peaked around $z=0$. With this definition, the total number of objects is

$$N_{tot}=\sum_t\int dz\frac{d\bar{N}_t(m_i,\hat{n},z)}{dz}. \tag{4.11}$$

Clearly, then, for all types of objects, the prior probability for an object of type t at *any* redshift is simply $\Pi(t|m_i,\hat{n})=\int dz d\Pi(t,z|m_i,\hat{n})/dz$, and the probability that an object is of *any* type is $\sum_t \Pi(t|m_i,\hat{n})=1$.

With the priors given by the equations above, we can now compute the likelihoods and the evidences, given some objects with measured fluxes $\mathbf{f}$ and with reference magnitudes in the bin m_i, all located on a patch of the sky with area $\Delta\Omega$, around the direction $\hat{n}$.

Now, assuming that the likelihoods are computed in the usual way, as in Eq. 4.6, the evidence is given by:

$$\begin{aligned}P(\mathbf{f}|m_i,\hat{n})&=\sum_t\int dz\mathscr{L}(\mathbf{f}|t,z)\frac{d\Pi(t,z|m_i,\hat{n})}{dz}\\&=\sum_{t_s}\mathscr{L}(\mathbf{f}|t_s)\Pi(t_s|m_i,\hat{n})+\sum_{t_e}\int dz\mathscr{L}(\mathbf{f}|t_e,z)\frac{d\Pi(t_e,z|m_i,\hat{n})}{dz}.\end{aligned} \tag{4.12}$$

4.3 Posterior probabilities

With the likelihood of Eq. 4.6, the priors given in Eq. 4.9-4.10, and the evidence of Eq. 4.12, we can now write the posterior probabilities. For an object with measured fluxes $\mathbf{f}$, reference magnitude in the bin m_i, in a field centered around the direction $\hat{n}$, with solid angle $\Delta\Omega$, the posterior PDF that it belongs to type t_e, and lies at some redshift z, is given by:

$$\frac{dp(t_e,z|\mathbf{f},m_i,\hat{n})}{dz} = \frac{\mathscr{L}(\mathbf{f}|t_e,z)}{P(\mathbf{f}|m_i,\hat{n})}\frac{d\Pi(t_e,z|m_i,\hat{n})}{dz}. \tag{4.13}$$

For stars, in particular, there is no redshift, so the posterior probability that the object is a star of type t_s becomes:

$$p(t_s|\mathbf{f},m_i,\hat{n}) = \frac{\mathscr{L}(\mathbf{f}|t_s)}{P(\mathbf{f}|m_i,\hat{n})}\Pi(t_s|m_i,\hat{n}). \tag{4.14}$$

These equations represent the key point of this project. It is important to notice that the evidence mixes the posterior probability of a point-like source being a quasar at redshift z with that of the object being a star of type t_s. This means that in areas where there are large numbers of stars, the posterior probability of an object being typed as a quasar will decrease, independently of its redshift. However, in the absence of the priors (and the evidence), the contamination of the quasar catalog from stars would increase.

Chapter 5

Classifying quasars and stars

In Chapter 3, we described the procedure employed for creating the mock catalog of quasars and stars. Our sample contains the mean expected number of quasars and stars on 1 deg^2 given, respectively, by the luminosity function of Croom et al. 2009 and the Besançon Model for the Milky Way stellar population. It consists of 218 quasars selected from a SDSS catalog of about $10,000$ quasars and $2,719$ stars selected from the SDSS *Spectra of Everything*, which contains a random sample of all stars covering an area of approximately 240 deg^2.

As we have seen, template-fitting codes are very sensitive to template choice. For this reason, several sets of templates for quasars were tested before we obtained the best set which minimized the error on the photo-z for quasars and maximized the correct typing of stars. These tests were performed for a sample of 63 spectroscopically confirmed quasars from ALHAMBRA. The results for the most relevant tests are presented in §5.1, as well as the final set of templates that was employed in our analysis, composed of 7 templates for quasars. The stars are categorized in 118 types and are later collected into eight "broad" types (O, B, A, F, G, K, M, WD) because, in principle, the information of subclasses is not relevant for this analysis.

For the final set of templates, we compare the performance of BPZ with that of LePhare for a sample of 8,716 quasars. We run the codes without any priors and only allow interpolations between the templates of quasars. For LePhare, we do not use any extinction law.

In Chapter 2, we saw that BPZ was developed to work with galaxies, therefore some adaptations were necessary in order to include templates for quasars and stars. LePhare, on the other hand, has a size limitation with respect to the dimension of some vectors and does not provide all the redshift likelihoods for stars.

For this reason, the comparative analysis in §5.2 was performed only for

J-PLUS and ALHAMBRA, whose photometric systems have 12 and 23 filters, respectively, of which only 11 and 18 are used because they are the ones in the wavelength range of our spectra.

In section §5.3, we study the effect of inclusion of priors, and we consider three filter systems: ALHAMBRA, with 18 medium-band filters, and two sets of filters for J-PAS, one with 42 filters and a more complete one with 55 filters. Again, these numbers correspond to the filters that cover the same wavelength ranges of the spectra of our sample. For this analysis, we use only BPZ.

Typically, the global minimum χ^2 provided by the photo-z code determines both the best photo-z and type for each object. In fact, we jointly analyze the global χ^2 given by the set of templates for quasars and the set of templates for stars, and the minimum χ^2 between both analyses gives us the classification for each object before the inclusion of priors.

We then include the priors into the output likelihoods to obtain the posterior probabilities, and once more evaluate how the objects are classified. The results are presented in terms of three level quality cuts (high, medium and low) in the diagram of magnitude versus χ^2. These criteria are arbitrary and designed to select higher SNR objects. They are also designed for each instrument: they may vary among the different filter systems, but in a way that separates more or less evenly the sample, and so they are fixed for the analyses with the templates of quasars and stars.

We also make an assessment of the initial (before including priors) and final (after the inclusion of priors) purity and completeness of the sample, and focus on objects classified at $z \geq 2.1$, which are important for studies in the Lyα forest.

Since the initial tests with the quasar templates were not performed with exactly the same objects of our sample, the chosen templates might not reproduce with great fidelity the spectra of quasars of our catalog. In §5.3 we see that this is true especially for the quasars, and for this reason an additional analysis was performed: we consider that the error of not using the most adequate templates for quasars is about 10% of the error of the fluxes in each filter. We then evaluate how the inclusion of this error affects the results.

5.1 Templates for quasars and stars

5.1.1 Sets of templates

In figure 5.1, we show a mosaic (Dr. Alberto Molino, in private communication) of a sample of 63 spectroscopically confirmed quasars detected in the first field of ALHAMBRA.

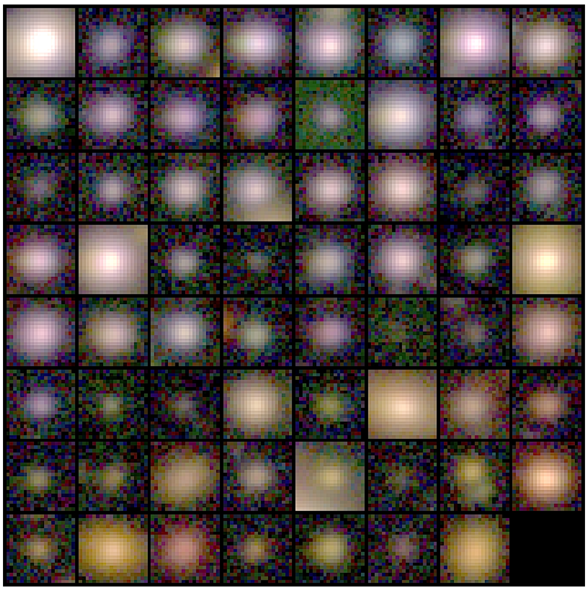


Figure 5.1: Mosaic with 63 spectroscopically confirmed quasars from ALHAMBRA (Alhambra Collaboration).

Although we have information for all 23 ALHAMBRA filters, we use only 18 filters in these tests (we do not consider the bluest filter and the NIR filters J, H and K_S). For the analysis, we use BPZ and quantify the quality of the fit with σ_{NMAD} and the fraction of quasars that have odds ≥ 0.6, where odds is a probability that quantifies how good is the redshift likelihood for the best photo-z. As we will see, the number of outliers vary little from set to set (and it is of the order of 20).

We introduce the final list of templates in Fig.5.2, and then we explain some of the tests that lead us to this list. This initial list contains 35 templates for galaxies and quasars, a subsample of the 50 templates used in Matute et al. (2012). The templates are organized as: non-active galaxies, including elliptical and spiral galaxies (#1-8 & 13); starburst galaxies (#9-12); Seyferts (#14-16); hybrid quasar templates, i.e., composite templates of galaxies with type-I and type-II quasars (#17-28); one synthetic template (#31) and quasar templates (#29-30 & 32-35).

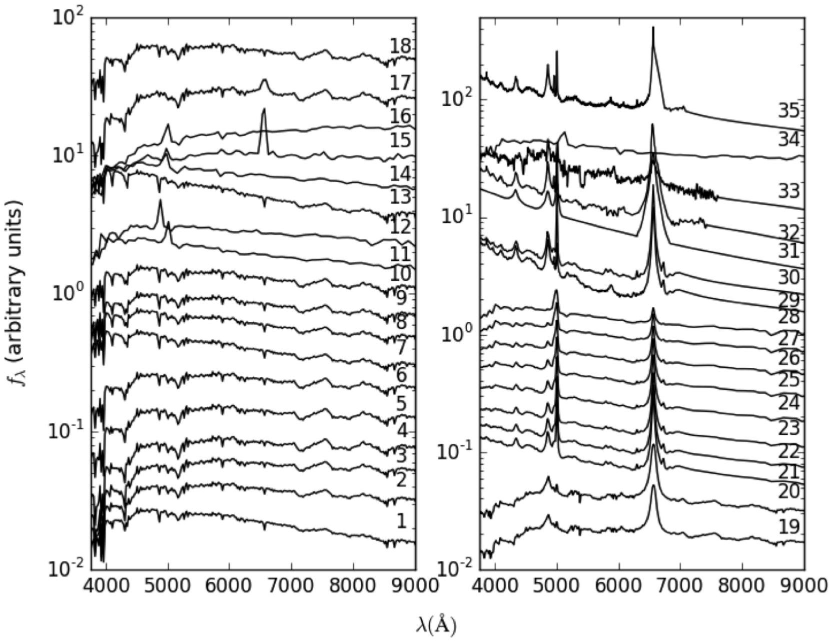


Figure 5.2: Complete set of extragalactic templates used in this work.

We start with a list of 33 templates, containing most of the templates present in the initial list, but without the hybrid galaxies and the quasar template #35. Instead, we include more synthetic templates, which in general improve the fit.

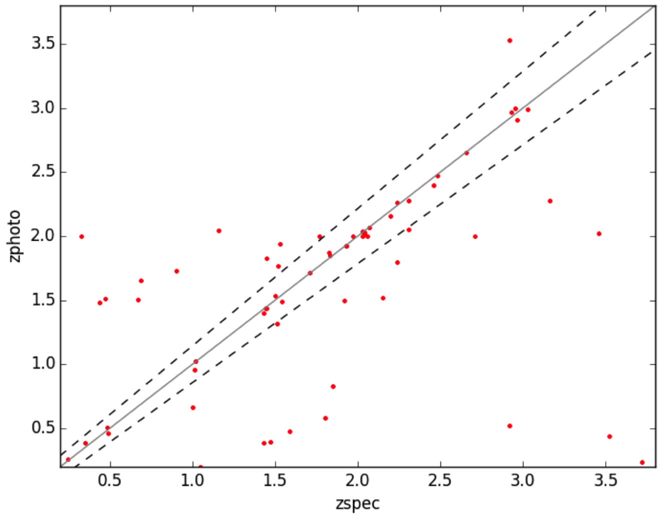


Figure 5.3: Photometric redshift as a function of the spectroscopic redshift for a sample of 63 quasars detected in the ALHAMBRA fields. These results were obtained with BPZ using a list of 33 templates for quasars and galaxies. The solid line indicates $z_{photo} = z_{spec}$, and the dashed lines indicate the outlier threshold. The photo-z uncertainty is $\sigma_{\mathrm{NMAD}} = 0.036$ and 42 quasars have odds$\geq$0.6.

We then reduced this list to 21 templates, maintaining only one template for elliptical galaxy and excluding the starburst galaxies, the hybrid templates, and all of the synthetic templates.

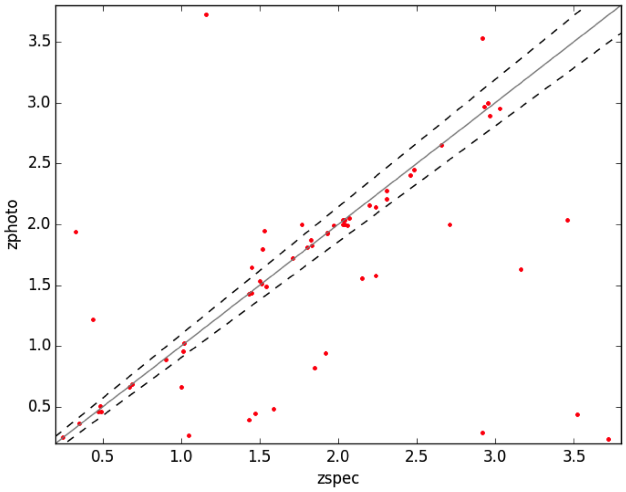


Figure 5.4: Photometric redshift as a function of the spectroscopic redshift for a sample of 63 quasars detected in the ALHAMBRA fields. These results were obtained with BPZ using a list of 21 templates for quasars and galaxies. The solid line indicates $z_{photo} = z_{spec}$, and the dashed lines indicate the outlier threshold. The photo-z uncertainty is $\sigma_{\mathrm{NMAD}} = 0.024$ and 47 quasars have odds$\geq$0.6.

Next, we maintain only one template for elliptical galaxy and exclude one hybrid template; this results in 19 templates. The result is very similar to the previous one.

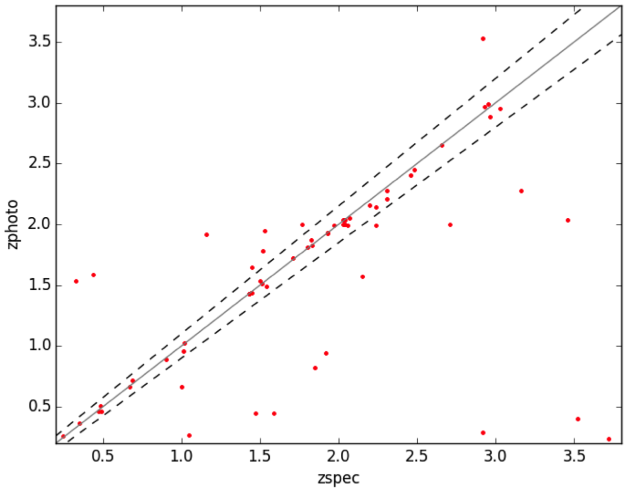


Figure 5.5: Photometric redshift as a function of the spectroscopic redshift for a sample of 63 quasars detected in the ALHAMBRA fields. These results were obtained with BPZ using a list of 19 templates for quasars and galaxies. The solid line indicates $z_{photo} = z_{spec}$, and the dashed lines indicate the outlier threshold. The photo-z uncertainty is $\sigma_{\mathrm{NMAD}} = 0.025$ and 48 quasars have odds$\geq$0.6.

Finally, instead of reducing the number of templates, we reordered the templates in the initial list with the purpose of mimicking the evolution path of galaxies: ellipticals, spirals, starburst, Seyfert and quasars. We also kept only one synthetic quasar and increased the number of hybrid templates. This gives us the 35 templates shown in Fig.5.2. This final set of templates for galaxies and quasars seems to minimize the photo-z uncertainty.

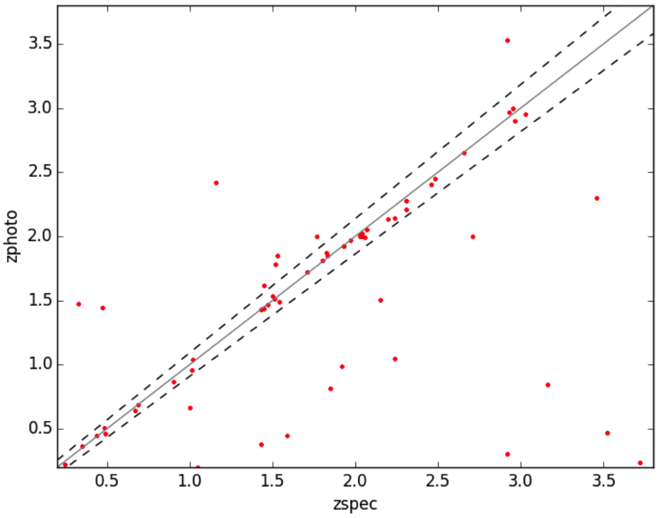


Figure 5.6: Photometric redshift as a function of the spectroscopic redshift for a sample of 63 quasars detected in the ALHAMBRA fields. These results were obtained with BPZ using a list of 35 templates for quasars and galaxies. The solid line indicates $z_{photo} = z_{spec}$, and the dashed lines indicate the outlier threshold. The photo-z uncertainty is $\sigma_{\rm NMAD} = 0.023$ and 49 quasars have odds$\geq$0.6.

In the end, our best set of templates is composed of 7 “pure” (i.e., without mixtures with galaxies) templates for type-I quasars, and from now on we use this set in our analysis. For stars, we consider 118 templates for main sequence stars and white dwarfs (or 8 categories in total), which in principle is a realistic set. As we have seen in Chapter 1, the classification of stars is much more complex in the sense that they span a wide range of luminosities and colors and, therefore, are divided in subclassifications other than just the main types (O, B, A, etc.). However, at this moment, we are not interested in a deep study of stars and allowing the stars two be divided in only 8 categories is sufficient for our purposes.

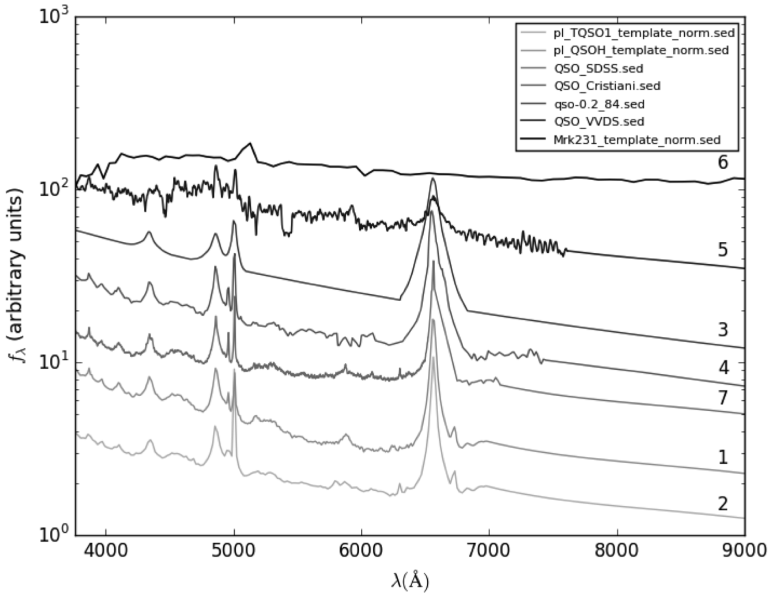


Figure 5.7: Set of templates for type-I quasars. The different colors are employed just to facilitate the visualization.

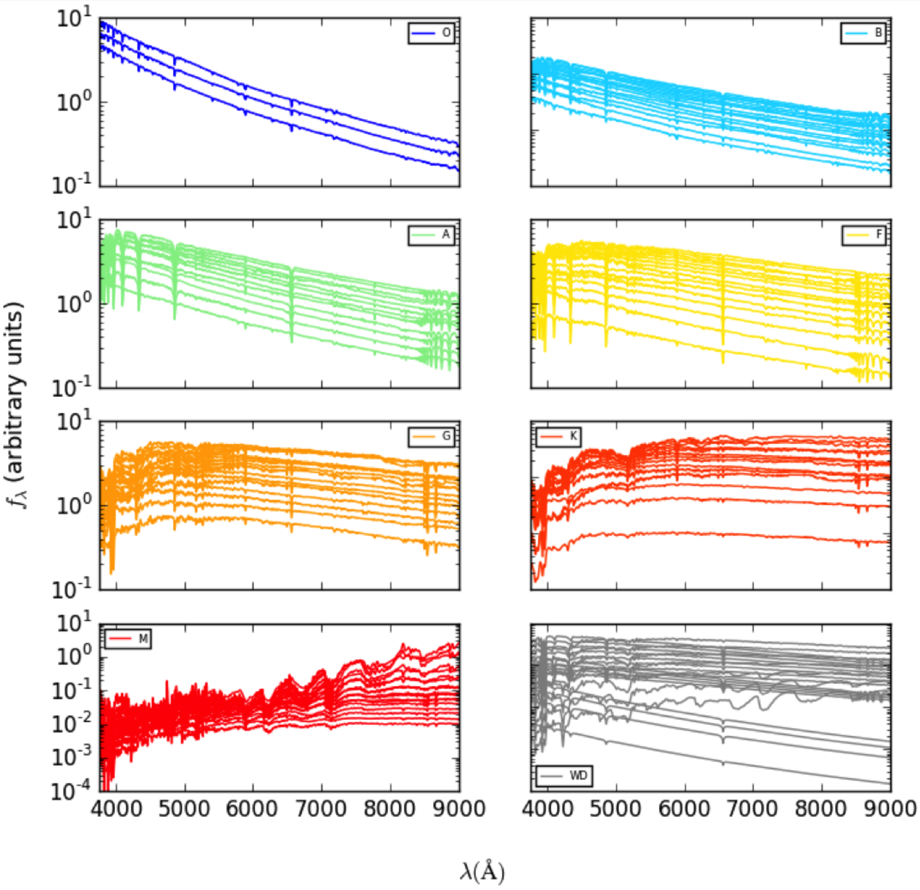


Figure 5.8: Set of stellar templates. The different colors are employed just to facilitate the visualization and divide the stars into the 8 main groups we are considering.

5.2 Comparison between BPZ and LePhare

We compare the performance of BPZ and LePhare for two photometric systems: J-PLUS, with 11 filters, and ALHAMBRA, with 18 filters. The analysis is performed for quasars with i-band magnitudes in the range $[17,\ 23]$. The fluxes were simulated from real SDSS spectra of 8,716 quasars with $0.5 \le z \le 3.5$.

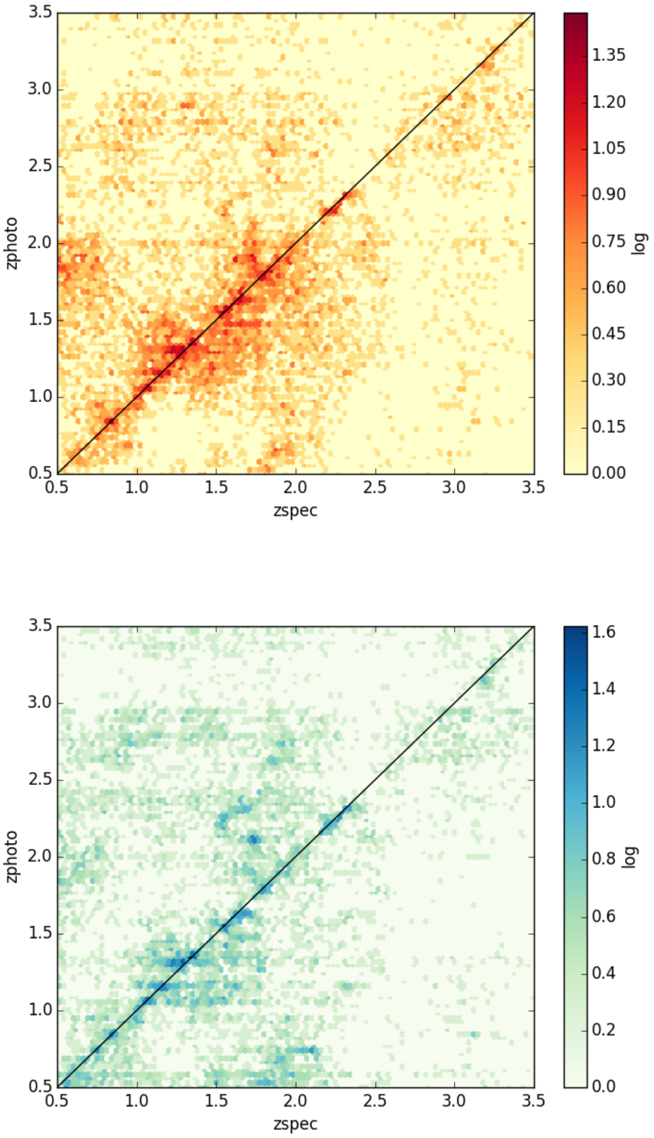
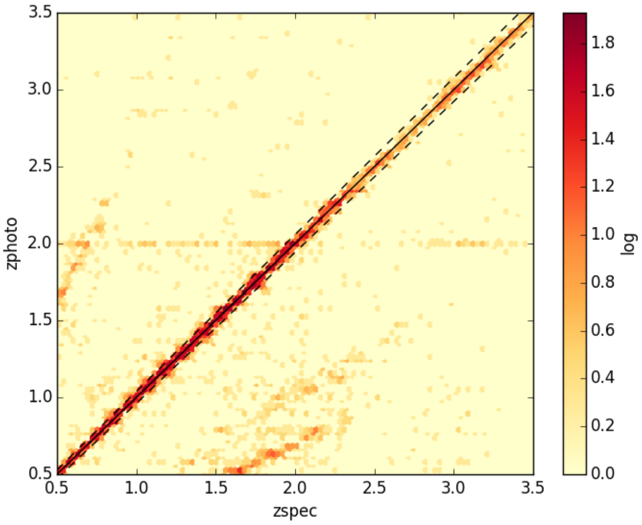


Figure 5.9: Photometric redshift as a function of the spectroscopic redshift for a sample of 8,716 quasars simulated in the in the J-PLUS photometric system. Upper panel: Results obtained with BPZ; $\sigma_{\mathrm{NMAD}} = 0.18$. Lower panel: Results obtained with LePhare; $\sigma_{\mathrm{NMAD}} = 0.23$. The solid line indicates $z_{photo} = z_{spec}$.



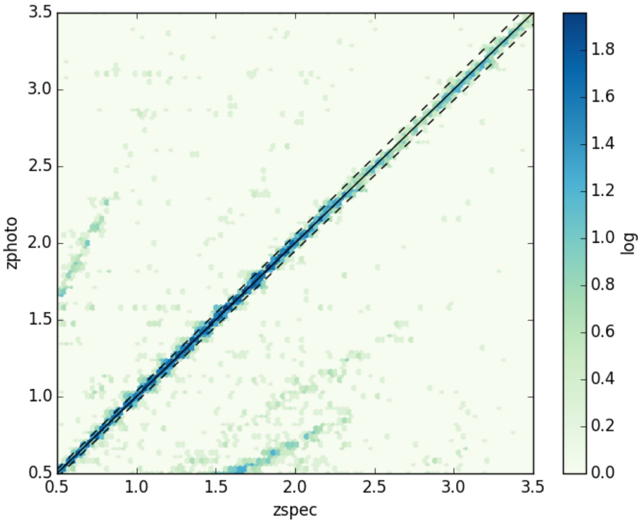


Figure 5.10: Photometric redshift as a function of the spectroscopic redshift for a sample of 8,716 quasars simulated in the in the ALHAMBRA photometric system. Upper panel: Results obtained with BPZ; $\sigma_{\mathrm{NMAD}} = 0.01$. Lower panel: Results obtained with LePhare; $\sigma_{\mathrm{NMAD}} = 0.009$. The solid line indicates $z_{photo} = z_{spec}$, and the dashed lines indicate the outlier threshold.

In Fig.5.9, we can note the large number of catastrophic redshifts (especially the horizontal features) obtained for J-PLUS. This happens because, due to the reduced number of filters (Fig.2.5), many of the characteristic spectral features are lost; as a result, templates with a similar continuum will be equally good fits, generating the observed features.

For ALHAMBRA, the number of outliers is much smaller, and we see only one characteristic horizontal line in photo-z for BPZ. The explanation is very similar: for that specific redshift, there are no characteristic spectral features, and thus there is a degeneracy in photometric redshift. In Fig.5.10, we note the peculiar diagonal lines approximately parallel to the solid line (where $z_{photo} = z_{spec}$). This degeneracy occurs whenever it is not possible to distinguish between two (or more) pairs of broad emission lines that are separated by the same relative wavelength interval (Abramo et al. 2012), which means that there are two "local minima" in the likelihood as a function of redshift.

We also note that σ_{NMAD} for ALHAMBRA decreases by a factor of the order of 10 when compared with J-PLUS, for both codes, reflecting our understanding that an increase in the number of filters of the photometric system improves the photo-z estimation, because more spectral features can be detected.

5.3 Including priors

Here we discuss the results obtained with BPZ using three photometric systems: ALHAMBRA, an incomplete set of filters from J-PAS containing 42 filters and J-PAS with 55 filters. The study of this incomplete J-PAS system is motivated by the observing strategy of using the bluest narrow-band (NB) filters and all of the redder NB filters plus 3 broad-band filters to improve the photo-z precision for red galaxies, which will be probably adopted in the first 18 months of the survey. In fact, we had some problems while including four of the NIR filters in our simulations, so we completed the "holes" with similar filters from ALHAMBRA, but using the same central wavelengths and magnitude limits as the analogous ones from J-PAS.

For this analysis, we use the data set defined in §3.5, composed of 218 quasars and 2,719 stars.

5.3.1 ALHAMBRA

First, we define the quality cuts in diagrams of magnitude as a function of the χ^2 obtained for the sets of templates of quasars and of stars in Fig.5.11.

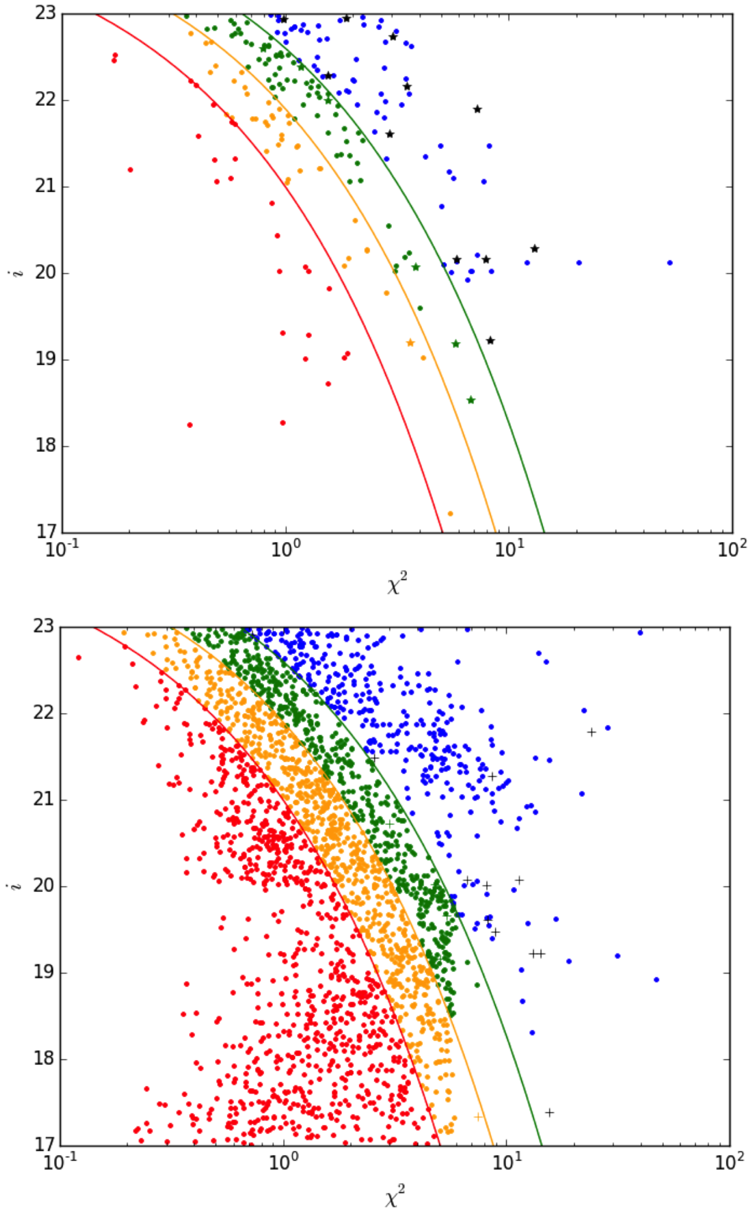


Figure 5.11: Magnitude-χ^2 diagram for the ALHAMBRA photometric system before the inclusion of priors. The colored lines represent the quality cuts: the red one (first curve from left to right) is the high, orange is the medium and green is the low; all the objects below any of these curves satisfy the respective cut. All of the objects that do not satisfy any of the cuts are represented in blue. Upper panel: Results obtained with the quasar templates. Real quasars classified as quasars are represented as dots; real quasars classified as stars are represented with a symbol of star. Lower panel: Results obtained with the star templates. Real stars classified as stars are represented as dots; real stars classified as quasars are represented with a cross.

The scatter plot of the comparison between the photometric redshifts and the spectroscopic redshifts before the priors is shown in Fig.5.12. Since some stars are classified as quasars, there is also a histogram showing at which redshift these stars are put, with the same color code as the quality cuts. We only show the real quasars that end up classified as quasars and the real stars that are also classified as quasars.

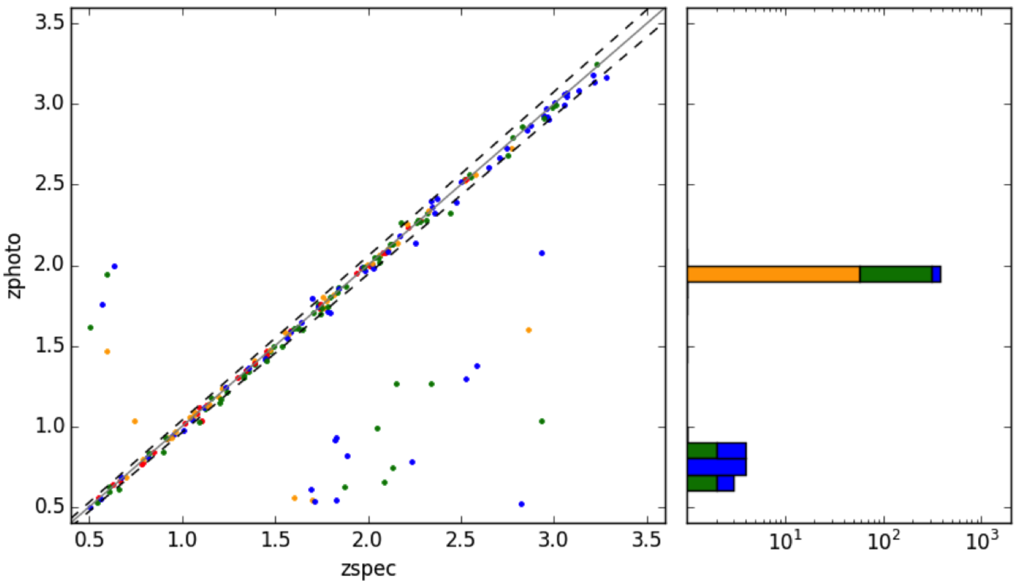


Figure 5.12: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars before the inclusion of priors for the ALHAMBRA photometric system. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts, which will be defined in the following. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts.

The precision of the photo-z estimation for the real quasars that end up classified as quasars with the ALHAMBRA photometric system is $\sigma_{\mathrm{NMAD}} = 0.009$.

Fig.5.13 compares the original stellar types of our sample and the types with which the real stars end up classified before and after the priors.

In Fig.5.14, we show a comparison between the photometric redshift and the spectroscopic redshift after the inclusion of priors.

Finally, in Fig.5.15 and 5.16 we compare the types of stars in which the quasars end up wrongly classified, and also the types for the stars that end up wrongly classified as quasars, before and after the priors.

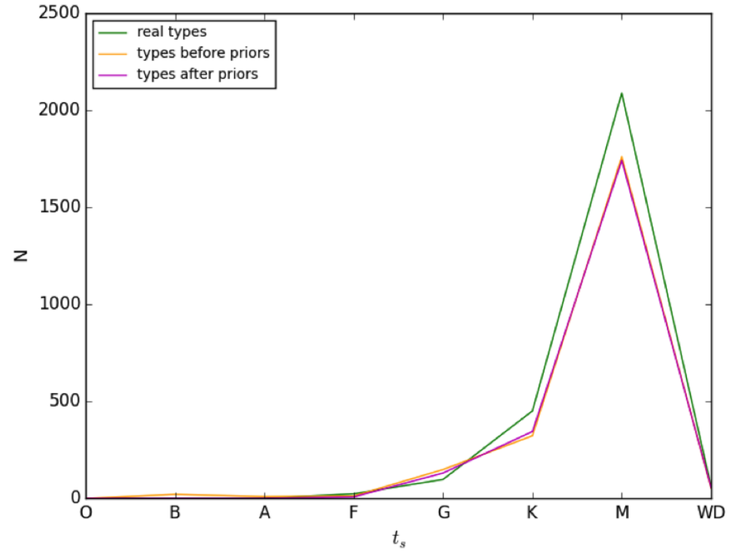


Figure 5.13: Comparison between the stellar types before and after the priors for the ALHAMBRA photometric system.

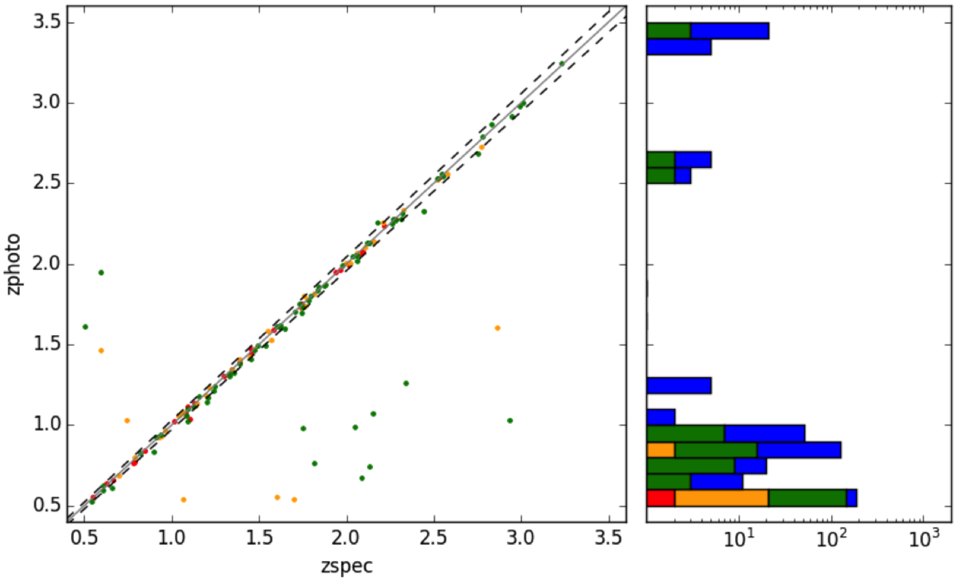


Figure 5.14: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars after the inclusion of priors for ALHAMBRA. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts. The accuracy of the photo-z estimation is $\sigma_{NMAD} = 0.007$.

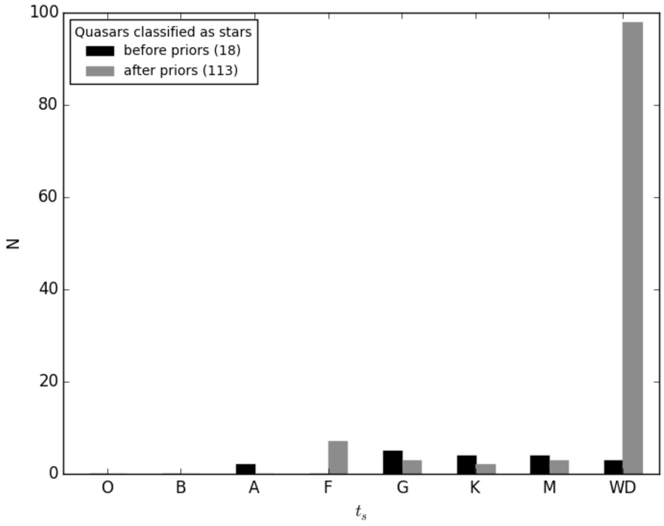


Figure 5.15: Stellar types in which the quasars end up classified before and after the inclusion of priors.

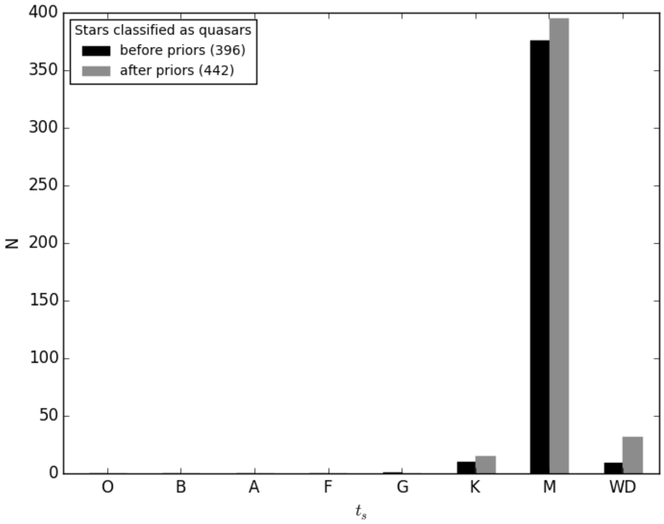


Figure 5.16: Stellar types of the real stars that end up classified as quasars before and after the priors.

In table 5.1 we summarize the numbers of objects that satisfy each quality cut, and how they are classified before and after the priors. We also define the purity and completeness of our sample for quasars and stars, before and after the inclusion of priors, for each quality cut. The purity p_k and the completeness C_k for the object of kind k (quasar or star) are given by:

$$p_k = \frac{K}{Q_k + S_k} \tag{5.1}$$

where Q_k and S_k are the number of real quasars and real stars, respectively, that end up classified as kind k;

$$C_k = \frac{K}{N_k} \tag{5.2}$$

where K is the number of real objects of kind k that are classified as kind k and N_k is the total number of objects of kind k in our sample (218 quasars and 2719 stars).

Table 5.1: Classification of our sample before and after the inclusion of priors for the ALHAMBRA photometric system.

	Before priors							
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	828	27 (27)	801 (801)	100%	12.4%	100%	29.5%
medium	1562	67 (68)	1493 (1494)	98.5%	30.7%	99.9%	54.9%
low	2130	134 (140)	1983 (1990)	95.7%	61.5%	99.6%	72.9%
	After priors						
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	828	12 (15)	798 (813)	80%	5.5%	98.2%	29.3%
medium	1562	34 (44)	1484 (1518)	77.3%	15.6%	97.8%	54.6%
low	2130	63 (119)	1933 (2011)	52.9%	28.9%	96.1%	71.1%

In Table 5.1, N is the total number of objects that satisfy the respective quality cut, the numbers in parentheses in the columns Q and S are the total numbers of objects classified as quasars and stars, respectively. The numbers that appear out of the parentheses are the real objects that were correctly classified. The indices q and s that accompany the purity p and the completeness C correspond to the characterization of the quasars and the stars, respectively.

We are also interested in objects classified at $z \geq 2.1$, which is motivated by their importance for studies in the Lyα forest. In Table 5.2 we report the numbers of quasars (at any z_{spec}) that are classified as $z \geq 2.1$ quasars (q), the true number of $z_{spec} \geq 2.1$ quasars that are classified as $z_{photo} \geq 2.1$ quasars (t) and stars that are classified as $z \geq 2.1$ quasars (s), as well as the final purity $p = t/(q+s)$, before and after the priors.

Table 5.2: Photometric redshift classification ($z \geq 2.1$) of our sample before and after the inclusion of priors for ALHAMBRA.

	before priors				after priors			
Quality cut	q	t	s	p	q	t	s	p
high	3	3	0	100%	1	1	0	100%
medium	11	11	0	100%	9	9	2	81.8%
low	19	19	0	100%	17	17	5	77.3%

5.3.2 J-PAS with 42 filters

In the following we show the results for the J-PAS photometric system with 42 filters.

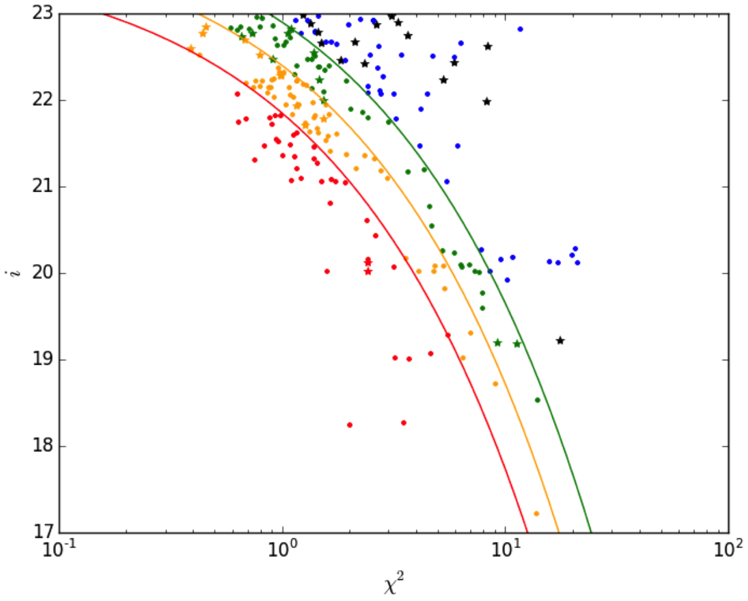


Figure 5.17: Magnitude-χ^2 diagram with the templates of quasars for the J-PAS photometric system with 42 filters before the inclusion of priors. The colored lines represent the quality cuts: the red one (first curve from left to right) is the high, orange is the medium and green is the low; all the objects below any of these curves satisfy the respective cut. All of the objects that do not satisfy any of the cuts are represented in blue. Real quasars classified as quasars are represented as dots; real quasars classified as stars are represented with a symbol of star.

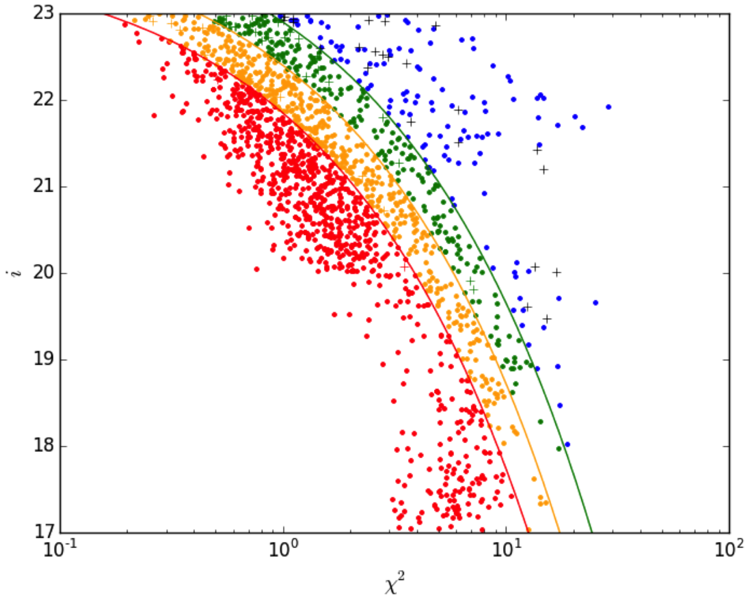


Figure 5.18: Magnitude-χ^2 diagram with the templates of stars for the J-PAS system with 42 filters before the inclusion of priors. The colored lines represent the quality cuts: the red one (first curve from left to right) is the high, orange is the medium and green is the low; all the objects below any of these curves satisfy the respective cut. All of the objects that do not satisfy any of the cuts are represented in blue. Real stars classified as stars are represented as dots; real stars classified as quasars are represented with a cross.

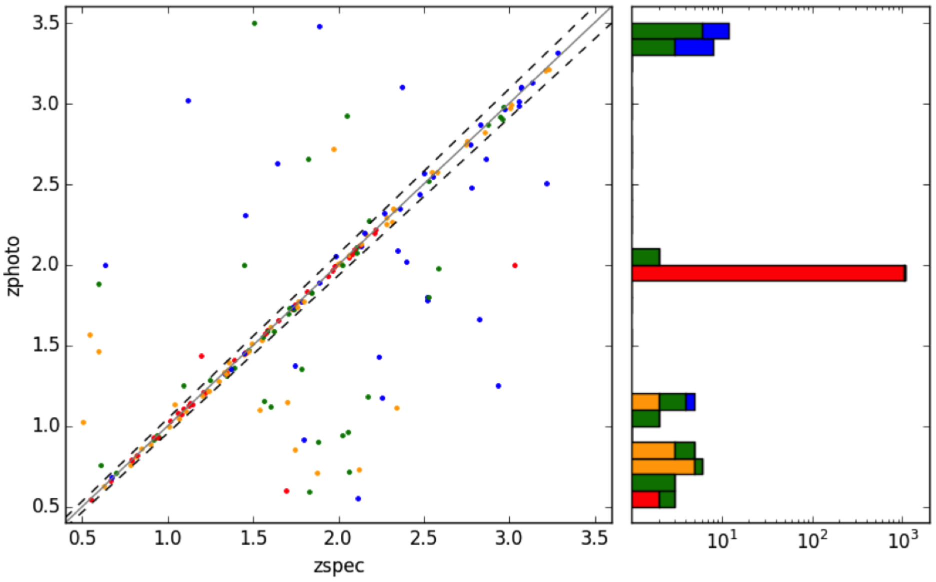


Figure 5.19: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars before the inclusion of priors for the J-PAS photometric system with 42 filters. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts.

The accuracy of the photo-z estimation for the real quasars that end up classified as quasars with the J-PAS photometric system with 42 filters is $\sigma_{\mathrm{NMAD}} = 0.011$.

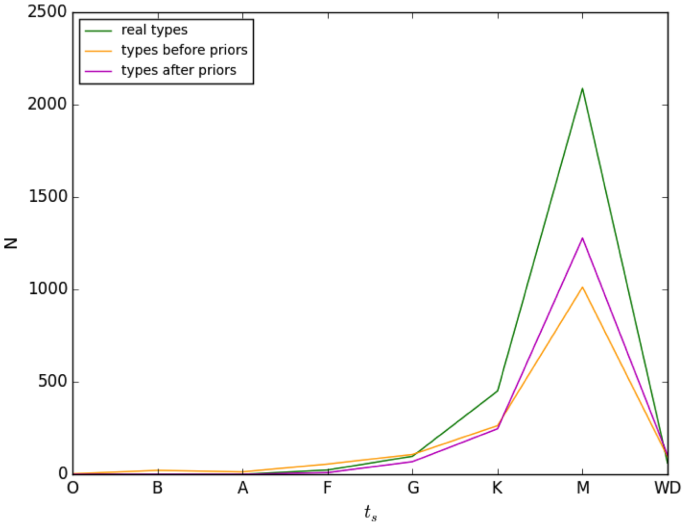


Figure 5.20: Comparison between the stellar types before and after the priors for the J-PAS photometric system with 42 filters.

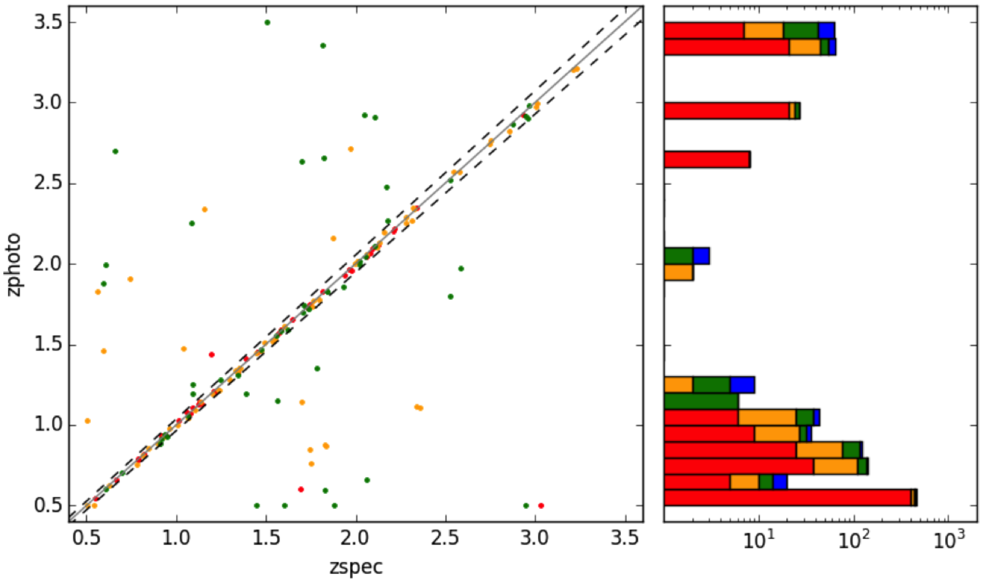


Figure 5.21: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars after the inclusion of priors for the J-PAS photometric system with 42 filters. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts. The accuracy of the photo-z estimation is $\sigma_{NMAD} = 0.009$.

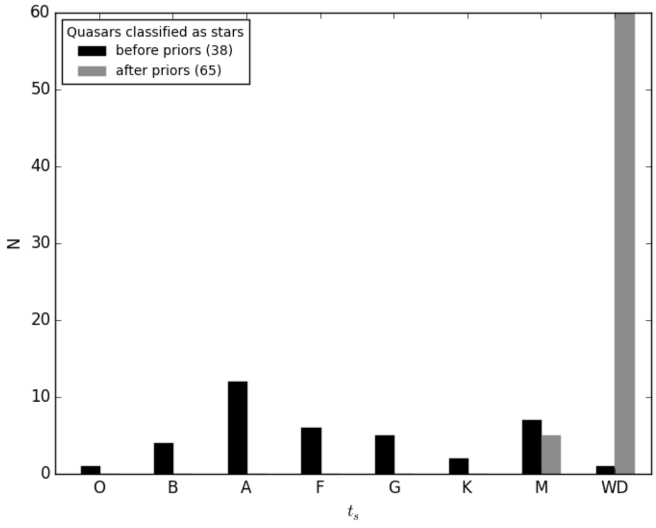


Figure 5.22: Stellar types in which the quasars end up classified before and after the inclusion of priors.

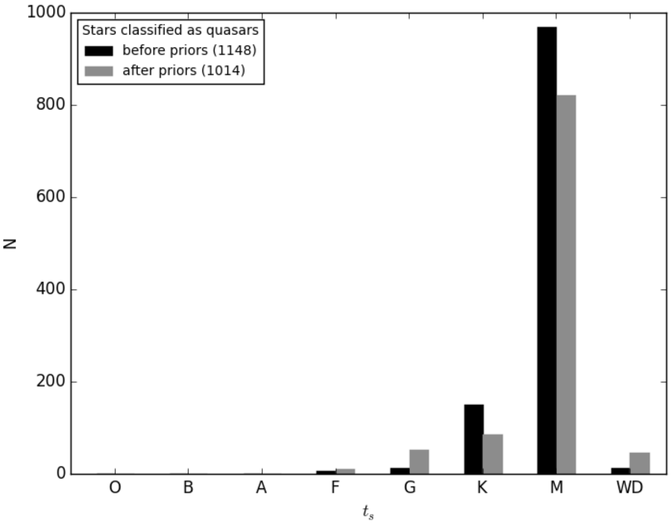
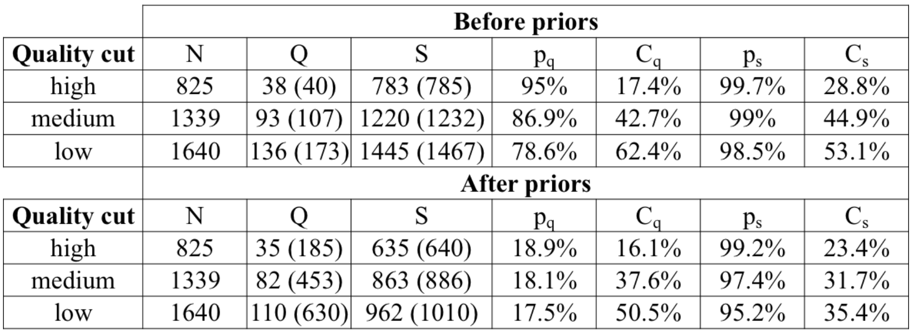


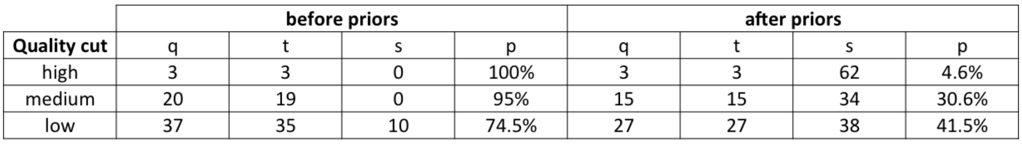
Figure 5.23: Stellar types of the real stars that end up classified as quasars before and after the priors.

Table 5.3: Classification of our sample before and after the inclusion of priors for the J-PAS photometric system with 42 filters.



	Before priors						
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	825	38 (40)	783 (785)	95%	17.4%	99.7%	28.8%
medium	1339	93 (107)	1220 (1232)	86.9%	42.7%	99%	44.9%
low	1640	136 (173)	1445 (1467)	78.6%	62.4%	98.5%	53.1%
	After priors						
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	825	35 (185)	635 (640)	18.9%	16.1%	99.2%	23.4%
medium	1339	82 (453)	863 (886)	18.1%	37.6%	97.4%	31.7%
low	1640	110 (630)	962 (1010)	17.5%	50.5%	95.2%	35.4%

Table 5.4: Photometric redshift classification ($z \geq 2.1$) of our sample before and after the inclusion of priors for J-PAS with 42 filters.



	before priors				**after priors**			
Quality cut	q	t	s	p	q	t	s	p
high	3	3	0	100%	3	3	62	4.6%
medium	20	19	0	95%	15	15	34	30.6%
low	37	35	10	74.5%	27	27	38	41.5%

5.3.3 J-PAS with 55 filters

In the following we show the results for the J-PAS photometric system with 55 filters.

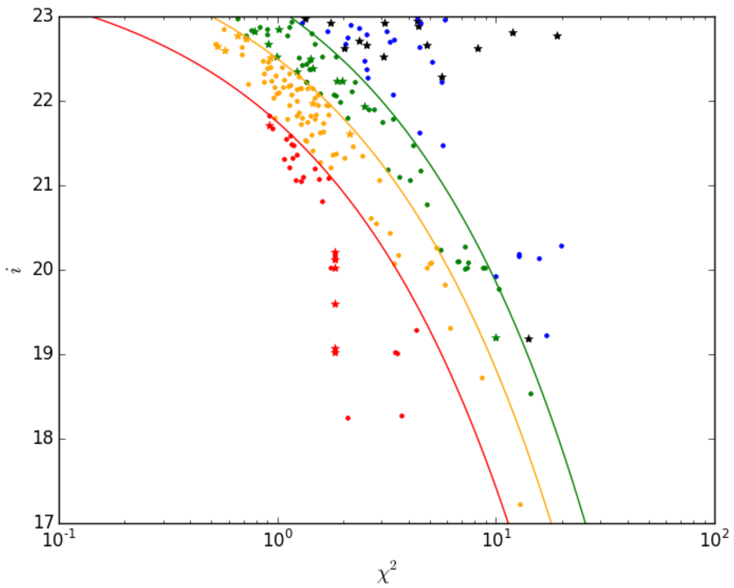


Figure 5.24: Magnitude-χ^2 diagram with the templates of quasars for the J-PAS system with 55 filters before the inclusion of priors. The colored lines represent the quality cuts: the red one (first curve from left to right) is the high, orange is the medium and green is the low; all the objects below any of these curves satisfy the respective cut. All of the objects that do not satisfy any of the cuts are represented in blue. Real quasars classified as quasars are represented as dots; real quasars classified as stars are represented with a symbol of star.

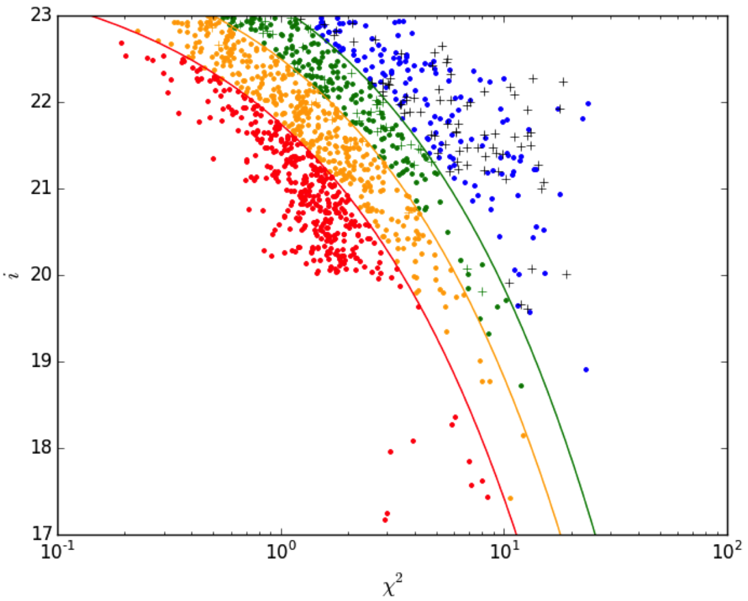


Figure 5.25: Magnitude-χ^2 diagram with the templates of stars for the J-PAS system with 55 filters before the inclusion of priors. The colored lines represent the quality cuts: the red one (first curve from left to right) is the high, orange is the medium and green is the low; all the objects below any of these curves satisfy the respective cut. All of the objects that do not satisfy any of the cuts are represented in blue. Real stars classified as stars are represented as dots; real stars classified as quasars are represented with a cross.

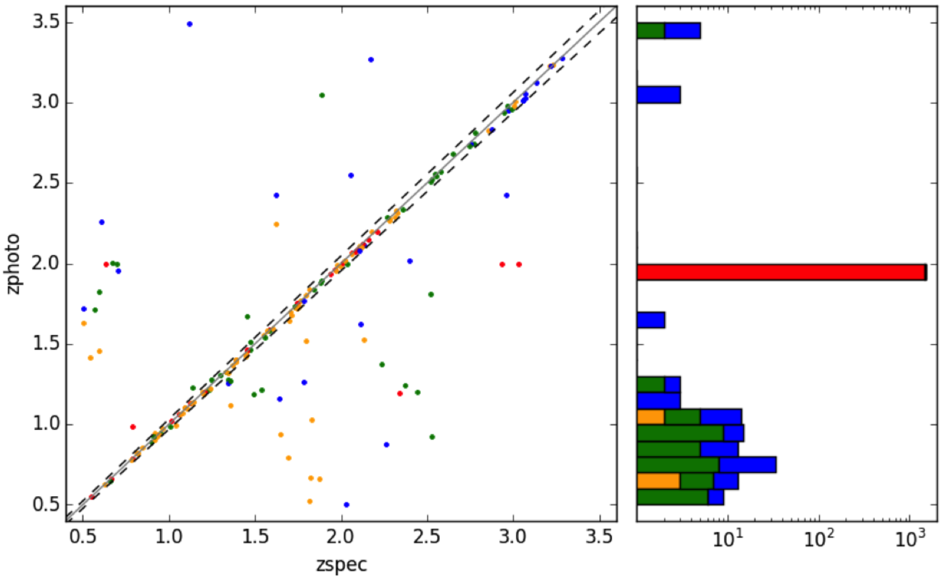


Figure 5.26: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars before the inclusion of priors for the J-PAS system with 55 filters. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts.

The accuracy of the photo-z estimation for the real quasars that end up classified as quasars with the J-PAS photometric system with 55 filters is $\sigma_{\mathsf{NMAD}} = 0.008$.

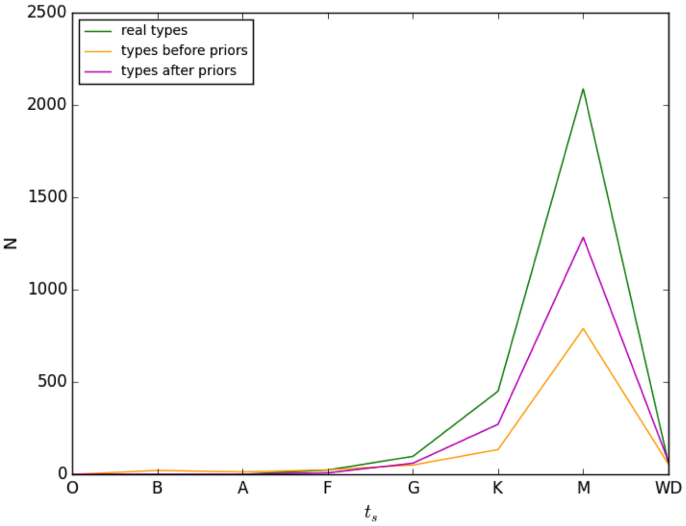


Figure 5.27: Comparison between the stellar types before and after the priors for the J-PAS with 55 filters.

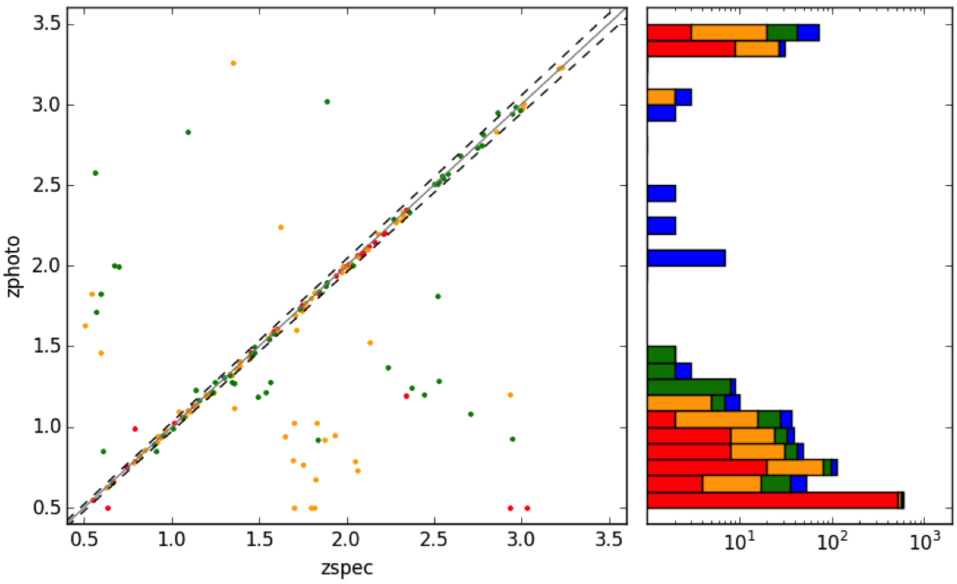


Figure 5.28: Comparison between the photometric redshift with the spectroscopic redshift for the real quasars of our sample that end up classified as quasars after the inclusion of priors for the J-PAS with 55 filters. The solid line gives $z_{photo} = z_{spec}$ and the dashed lines represent the boundary between good solutions and outliers. The colored dots represent our χ^2 quality cuts. In the right, the histogram shows at which photo-z's the real stars are classified as quasars and the color code is the same as the quality cuts. The accuracy of the photo-z estimation is $\sigma_{NMAD} = 0.007$.

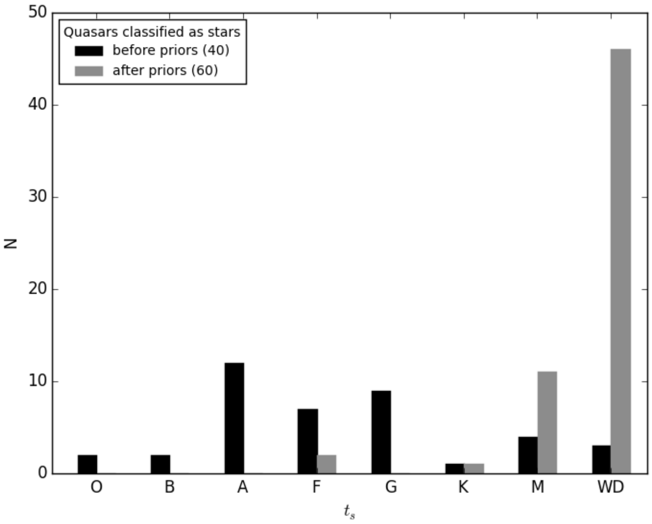


Figure 5.29: Stellar types in which the quasars end up classified before and after the inclusion of priors.

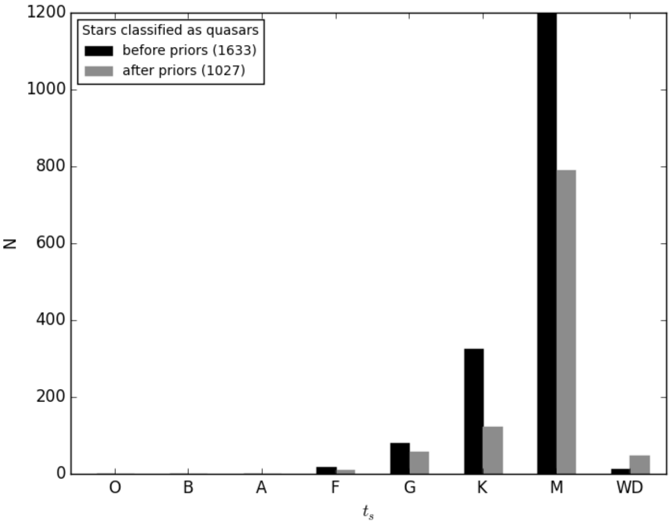
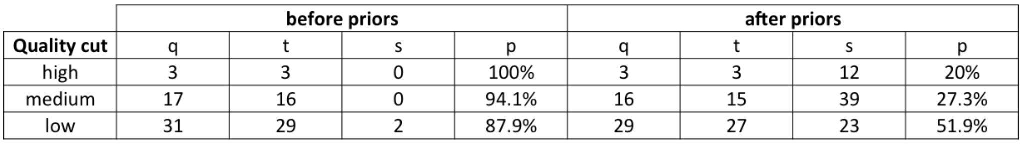


Figure 5.30: Stellar types of the real stars that end up classified as quasars before and after the priors.

Table 5.5: Classification of our sample before and after the inclusion of priors for the J-PAS photometric system with 55 filters.

	Before priors						
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	406	26 (26)	373 (380)	100%	11.9%	98.2%	13.7%
medium	875	102 (110)	751 (765)	92.7%	46.8%	98.2%	27.6%
low	1152	150 (196)	931 (956)	76.5%	68.8%	97.4%	34.2%
	After priors						
Quality cut	N	Q	S	p_q	C_q	p_s	C_s
high	406	24 (85)	312 (321)	28.2%	11%	97.2%	11.5%
medium	875	88 (338)	509 (537)	26%	40.4%	94.8%	18.7%
low	1152	127 (494)	610 (658)	25.7%	58.3%	92.7%	22.4%

Table 5.6: Photometric redshift classification ($z \geq 2.1$) of our sample before and after the inclusion of priors for J-PAS with 55 filters.



	before priors				**after priors**			
Quality cut	q	t	s	p	q	t	s	p
high	3	3	0	100%	3	3	12	20%
medium	17	16	0	94.1%	16	15	39	27.3%
low	31	29	2	87.9%	29	27	23	51.9%

5.3.4 Summary

Although we considered an ALHAMBRA filter system with only 18 filters, its performance is superior to that of J-PAS if we consider only the number of real quasars classified as quasars (200/218) and stars classified as stars (2323/2719) before the priors. As for the photo-z accuracy, J-PAS with 55 filters performs best, with $\sigma_{\mathsf{NMAD}} = 0.008$. After the priors, both J-PAS photometric systems perform better for classifying real quasars as quasars, but ALHAMBRA classifies best real stars. The photo-z uncertainty decreases for all systems after the priors. These differences can be attributed to the fact that ALHAMBRA goes deeper in magnitude, while the narrow-band filters from J-PAS were designed to detect better emission lines from quasars.

In general, we note that, for the three photometric systems, most of the quasars that end up classified as stars are classified as white dwarfs, and most of the stars that end up classified as quasars, both before and after the priors, are M-type stars. So, we identify a pattern which might be related to the template choices or might be related to intrinsic features of our objects.

The completeness relates the number of objects that satisfy the quality cuts to the total number of the correspondent species in the sample, so it reflects the way by which these cuts are defined. Although low for the three photometric systems, the completeness does not vary a lot from the case without priors to the case with priors. The purity, on the other hand, indicates if the objects end up classified with their real classification. We note in our analysis that the purity for stars is always higher than that for quasars, both before and after the priors, indicating that, comparatively, few quasars end up classified as stars. However, the contamination of the quasar population by stars is still high, and thus some refinements are necessary in our methods.

We also identify $z_{photo} = 2$ as the photometric redshift where most of the stars classified as quasars are placed. However, we see in tables 5.2, 5.4 and 5.6 that actually very few stars end up classified as $z \geq 2.1$ quasars, and this is an important result for separating quasars from stars in imaging surveys. This gives us a clear idea of which are the bins of redshift where the contamination problem occurs more frequently.

The results for all the photometric systems are summarized in table 5.7.

Table 5.7: Results for the photo-z estimation using ALHAMBRA and J-PAS.

	Before priors			**After priors**		
Filter system	quasar	star	σ_{nmad}	quasar	star	σ_{nmad}
Alhambra	200	2323	0.009	105	2277	0.007
J-PAS 42	180	1572	0.011	153	1705	0.009
J-PAS 55	178	1086	0.008	158	1692	0.007

5.3.5 Additional analysis

In this section we present a combined analysis including the results of the three filter systems, when we consider an additional uncertainty of σ_T to account for the fact that the quasar templates do not fit well the spectra of real quasars in our sample. This procedure was briefly explained in Chapter 4 (§4.2), and consists basically of adding an uncertainty σ_T in the denominator of the likelihood of the fluxes:

$$\mathscr{L}(\mathbf{f}|t,z) = \exp\left[-\sum_k \frac{(f_k - f_k^t)^2}{\sigma_k^2 + \sigma_T^2} \right] \tag{5.3}$$

where we assume σ_T to be 10% of σ_k.

Table 5.8: Classification of the objects in our sample in the case in which an error of 10% is attributed to a possible mismatch between the quasar templates and the spectra of quasars in our sample.

	Before priors			**After priors**		
Filter system	quasar	star	σ_{nmad}	quasar	star	σ_{nmad}
Alhambra	197	2326	0.01	101	2287	0.01
J-PAS 42	188	1574	0.01	142	1699	0.008
J-PAS 55	170	1087	0.006	162	1678	0.006

In table 5.8, "quasar" refers to the total number of real quasars classified as quasars; similarly, "star" refers to the total number of real stars classified as stars. These are the number of objects without considering any sort of quality cut.

The photo-z uncertainties improve a lot when we introduce an error to the templates of quasars to account for the wavelength-dependent template mismatches. However, we do not verify a significant improvement in the way the objects are classified. In other words: by including an error for the quasar templates we would expect to find less stars being classified as quasars, i.e., an increase in the number of real stars that end up classified as stars.

This means that our problem of having a mismatch between the quasar templates and the quasar spectra is not so trivial to be solved. Therefore, a more detailed study will be necessary to understand these limitations, include more representative quasar templates, and improve our results.

Chapter 6

Conclusions

There is strong evidence that quasars and stars can be confused in color-color diagrams (and also in color-magnitude diagrams), especially at intermediate redshifts ($z \sim 1.5 - 2.5$), where quasars overlap with the stellar locus.

The goal of this work was to develop an efficient technique to classify adequately quasars and stars using medium- and narrow-band filter systems. The main novelty of this approach is the use of Bayesian priors both for the angular distribution of stars of different types on the sky and for the distribution of quasars and AGNs as a function of redshift. In this context, we present the combined analysis of photometric redshift determination and characterization of a population of quasars and stars both without priors and using Bayesian priors.

We employ two photometric redshift codes in our analysis: BPZ and LePhare. Since BPZ was developed to work with galaxies, some adaptations were necessary in order to include templates for quasars and stars. In the analysis of a mixed sample of quasars and stars, we had actually to run BPZ twice and perform a joint analysis with the results. LePhare, on the other hand, has a size limitation with respect to the dimension of some vectors, which takes into account the redshift step and the number of filters, and does not provide all the redshift likelihoods for stars.

Since we use template-fitting codes, we start with a previous study of the set of quasar templates that minimizes the photometric redshift uncertainty for a sample of 63 spectroscopically confirmed quasars detected in the ALHAMBRA fields. This set contains 7 quasar templates.

The stars are categorized in 118 types and are later collected into eight "broad" types (O, B, A, F, G, K, M and white dwarf) because, in principle, the information of subclasses is not relevant for this analysis.

We compare the photometric redshift estimation for LePhare and BPZ for a sample of 8,716 quasars. This comparative study is performed for J-PLUS and

ALHAMBRA systems, because their reduced number of filters (11 and 18, respectively, in the wavelength range of our spectra) allow us to use LePhare. The results suggest that ALHAMBRA performs better than J-PLUS, and we attribute this to the higher number of filters and absence of holes in the wavelength coverage of the ALHAMBRA photometric system, which improves the photo-z estimation, since more spectral features can be detected.

Our main results are shown for a sample that contains the mean expected number of quasars and stars on 1 deg^2 given, respectively, by the luminosity function of Croom et al. (2009) and the Besançon Model for the Milky Way stellar population. We have 218 quasars selected from a SDSS catalog of about 10,000 quasars and 2,719 stars selected from the SDSS Spectra of Everything.

With the sample of (z, i_{mag}) for quasars and (t_s, i_{mag}) for stars, we simulate the fluxes we would expect in three photometric systems (ALHAMBRA with 18 filters, an incomplete set of filters from J-PAS with 42 filters and J-PAS with 55 filters). The goal of this analysis was to determine the effect of inclusion of priors, and it was performed using only BPZ.

A comparison between the results before and after the priors is presented in terms of different χ^2 and magnitude quality cuts, as well as in terms of the completeness and purity of the sample.

We can notice that, before the priors, the performance with ALHAMBRA is superior than that with J-PAS if we consider only the number of real quasars classified as quasars (200/218) and stars classified as stars (2323/2719). As for the photo-z accuracy, J-PAS with 55 filters performs best, with $\sigma_{\mathsf{NMAD}} = 0.006$. After the priors, both J-PAS photometric systems perform better for classifying real quasars as quasars, but ALHAMBRA classifies best real stars. These differences can be explained by the fact that ALHAMBRA goes deeper in magnitude, while the narrow-band filters of J-PAS are best for identifying quasars. The photo-z uncertainty decreases for all systems after the priors.

In general, few stars end up classified as $z \geq 2.1$ quasars, which is an important result for separating quasars from stars in imaging surveys. The analysis for these objects ($z \geq 2.1$) is motivated by their importance for studying the Lyα forest.

The photometric redshift uncertainties improve a lot when we introduce an error to the templates of quasars to account for the wavelength-dependent template mismatches. However, we would expect the number of real stars that are classified as quasars to decrease when we introduce this error, and this is not what we have obtained. There is no significant improvement in the way the objects end up classified, especially the stars, and thus more refinements are still necessary.

As for further developments, some improvements are necessary before applying our method to real catalogs. First, we must improve our set of quasar templates so that it can reproduce better our population of quasars. We would also like to study which is the best set of templates for stars that maximizes the classification in types and also that minimizes the number of stars that end up classified as quasars. Lastly, we would also like to properly identify a larger number of low luminous objects.

We show the results for only one realization of the combined sample of quasars and stars. In order to account for the low number of objects (especially stars, what has limited our analysis at 1 deg^2) and Poisson variations, it would also be desirable to make more realizations (ideally $>$10) of different samples.

There are still many open issues which can be analyzed with these new tools, which allow a fresh look at existing datasets. We will also, of course, study those issues with the data coming in the near future from J-PLUS and J-PAS.

First, we lack a thorough understanding of the environments where quasars are found, especially a) whether or not the bias of those objects is scale-dependent and b) what is the role of assembly bias.

Second, the numbers and redshift distributions of quasars of different magnitudes should be better understood (i.e., there are uncertainties in the quasar luminosity function at high redshifts and at the faint end), and we must be able to tell the difference between true modulations in the distribution of those objects, and gaps in our catalogs created by incompleteness or impurity of our datasets.

Finally, we also need to determine how the different abundances of stars at different regions of the sky affect the quasar catalog—especially at the faint end of our survey. We would like to apply our new methods and the improved, higher-quality datasets to test these issues.

Bibliography

[1] Abdalla, F.; Banerji, M.; Lahav, O.; Rashkov, V., 2008, arXiv: astro-ph/08123831

[2] Alam, S.; Albareti, F. D.; Prieto, C. A., et al., 2015, ApJS, 219, 27

[3] Amendola, L.; Appleby, S.; Bacon, D., et al., 2012, arXiv: 1206.1225

[4] Arnouts, S.; Cristiani, S.; Moscardini, L.; Matarrese, S.; Lucchin, F., et al., 1999, MNRAS, 310, 540

[5] Benítez, N. 2000, Ap.J., 536, 571

[6] Benítez, N.; Gaztañaga, E.; Miquel, R., et al., 2009, ApJ, 691, 241

[7] Benítez, N.; Dupke, R.; Moles, M., et al., 2014, arXiv: 1403.5237

[8] Blake, C. & Bridle, S., 2005, MNRAS , 363, 1329

[9] Bolzonella, M.; Mirallers, J-M.; Pelló, R., 2000, A&A, 363, 476

[10] Boyle, B.J.; Shanks, T. et al., 2000, MNRAS, 317, 1014

[11] Brammer, G.B.; van Dokkum, P. G.; Coppi, P., 2008, ApJ, 686, 1503

[12] Carliles, S.; Budavári, T.; Heinis, S.; Priebe, C.; Szalay, S., 2010, ApJ, 712, 511

[13] Colless, M., 2000, Astron. Soc. Aust., 17, 215

[14] Colless, M.; Dalton, G.; Maddox, S., et al., 2001, MNRAS, 328, 1039

[15] Colless, M.; Peterson, B. A.; Jackson, C., et al., 2003, arXiv: astro-ph/0306581

[16] Collister, A. A. & Lahav, O., 2004, PASP, 116, 345

[17] Connolly, A. J.; Csabai, I.; Szalay, A. S.; Koo, D. C.; Kron, R. G.; Munn, J. A., 1995, AJ, 110, 2655b

[18] Connolly, A. J.; Szalay, A. S.; Dickinson, M.; SubbaRao, M. U.; Brunner, R.J., et al., 1997, ApJ, 486, L11

[19] Croom, S. M.; Richards, G. T.; Shanks, T., et al., 2009, MNRAS, 399, 1755

[20] Dahlen T., Mobasher, B., Dickinson, M., et al., 2010, ApJ, 724, 425

[21] Dahlen, T.; Mobasher, B.; Faber, S. M., et al., 2013, arXiv: 1308.5353

[22] Dawson, K. S.; Schlegel, D. J.; Ahn, C. P., et al., 2012, arXiv: 1208.0022v3

[23] Dodelson, S. Modern Cosmology, Academic Press, USA, 2003

[24] Feldmann, R.; Carollo, C. M.; Porciani, C., et al., 2006, MNRAS, 372, 565

[25] Feroz, F. (2008). *Bayesian methods for astrophysics and particle physics*. PhD thesis. Cavendish Astrophysics and Clare Hall College, Cambridge

[26] Gao, S.; Just, A.; Grebel, E. K., 2012, arXiv: 1207.4551

[27] Gerdes, D. W.; Sypniewski, A. J.; McKay, T. A., et al., 2010, ApJ, 715, 823

[28] Hogg, D. W., 2000, arXiv: 9905116

[29] Hopkins, P. F.; Hernquist, L.; Cox, T. J., et al., 2005, ApJ, 630, 705

[30] Ilbert, O.; Arnouts, S.; McCracken, H. J.; Bolzonella, M.; Bertin, E., et al., 2006, A&A, 457, 841

[31] Ivezic, Z.; Tyson, J. A.; Abel, B., et al., 2008, arXiv: 0805.2366v4

[32] Kato, Y., 2006, Astronomische Nachrichten, 327, 450

[33] Kormendy, J. & Richstone, D., 1995, ARA&A, 33, 581

[34] Laureijs, R.; Amiaux, J.; Arduini, S., et al., 2011, arXiv: 1110.3193

[35] Lewis, I.J.; Amiaux, J.; Arduini, S., et al., 2002, MNRAS, 333, 279

[36] Loh, E. & Spillar, E., 1986, ApJ, 303, 154

[37] Matute, I.; Márquez, I.; Masegosa, J., et al., 2012, A&A, 542, A20

[38] Molino, A.; Benítez, N.; Moles, M., et al., 2014, MNRAS, 441, 2891

[39] Moles, M.; Benítez, N.; Aguerri, J. A. L., et al., 2008, arXiv: 0806.3021

[40] Mayer, L.; Kazantzidis, S.; Escala, A. & Callegari, S., 2010, Nature, 466, 1082

[41] Nenkova, M.; Sirocky, M. M.; Ivezic, Z., et al., 2008, ApJ, 685, 147

[42] Nenkova, M.; Sirocky, M. M.; Nikutta, R., et al., 2008, ApJ, 685, 160

[43] Penzias, A. A. & Wilson, R. W., 1965, AJ, 142, 419

[44] Planck Collaboration, 2015, arXiv: 1502.01589

[45] Richards, G. T.; Strauss, M. A.; Fan, X., et al., 2006, AJ, 131, 2766

[46] Richards, G. T.; Myers, A. D.; Gray, A. G., et al., 2009, ApJS, 180, 67

[47] Robin, A. C.; Reylé, C.; Derrière, S.; Picaud, S., 2003, A&A, 409, 523 (Besançon Model available at `http://model.obs-besancon.fr`; erratum: 2004, A&A, 416, 157)

[48] Schlegel, D. J.; Blanton, M.; Eisenstein, D., et al., 2007, in Bulletin of the American Astronomical Society, Vol. 39, p. 966, American Astronomical Society Meeting Abstracts, 132.29

[49] Schneider, P. Extragalactic Astronomy and Cosmology, Springer Berlin Heidelberg New York, New York, 2006

[50] Shankar, F.; Dai, X.; Sivakoff, G. R., 2008, ApJ, 687, 859

[51] Singh, C. B.; de Gouveia dal Pino, E. M.; Kadowaki, L. H. S., 2015, ApJL, 799, L20

[52] SDSS Collaboration: `http://www.sdss.org/instruments/camera`

[53] Smoot, G. F.; Bennett, C. L.; Kogut, A., et al., 1992, ApJ, 396, L1

[54] Spectra of Everything available at http://classic.sdss.org/dr6/products/spectra/special.html

[55] Takada, M.; Ellis, R.; Chiba, M., et al., 2012, arXiv: 1206.0737

[56] The Dark Energy Survey Collaboration, 2005, arXiv: astro-ph/0510346

[57] Torres, D. F. & Anchordoqui, L. A., 2004, Rep. Prog. Phys., 67, 1663

[58] Vanden Berk, D. E.; Richards, G. T.; Bauer, A., et al., 2001, AJ, 122, 549

[59] VanderPlas, J., 2014, arXiv: 1411.5018

[60] Vanzella, E.; Cristiani, S.; Fontana, A., et al., 2004, A&A, 423, 761

[61] Weinberg, S. Cosmology, Oxford University Press, Oxford, 2008

[62] White, S. D. M., 1996, arXiv: astro-ph/9602054

[63] Wolf, C.; Meisenheimer, K.; Kleinheinrich, M., et al., 2004, A&A, 421, 913

[64] York, D. G.; Adelman, J.; Anderson, J. E., et al., 2000, AJ, 120, 1579

[65] Yuksel, H.; Stanev, T.; Kistler, M. D.; Kronberg, P. P., 2012, arXiv: 1203.3197

[66] Zakamska, N. L.; Strauss, M. A.; Krolik, J. H., et al., 2003, arXiv: 0309551

[67] 2QZ team: http://www.2dfquasar.org