

Universidade de São Paulo
Instituto de Física

Redshifts fotométricos de quasares em levantamentos de bandas estreitas

Carolina Queiroz de Abreu Silva

Orientador: Prof. Dr. Luis Raul Weber Abramo

Tese de doutorado apresentada ao Instituto de Física
da Universidade de São Paulo, como requisito parcial
para a obtenção do título de Doutora em Ciências.

Banca Examinadora:

Prof. Dr. Luis Raul Weber Abramo (IF/USP)

Prof. Dr. Marcos V. Borges Teixeira Lima (IF/USP)

Prof. Dr. Martín Makler (CBPF)

Prof. Dr. Olivier Ilbert (LAM, França)

Prof. Dr. Rogério Riffel (UFRGS)

São Paulo
2021

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

Silva, Carolina Queiroz de Abreu

Redshifts fotométricos de quasares em levantamentos de bandas estreitas. São Paulo, 2021.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Física. Departamento de Física Matemática.

Orientador: Prof. Dr. Luis Raul Weber Abramo

Área de Concentração: Quasares

Unitermos: 1. Quasares; 2. Redshifts fotométricos; 3. Levantamentos fotométricos de bandas estreitas; 4. Estrutura em larga-escala.

USP/IF/SBI-002/2021

University of São Paulo
Institute of Physics

Photometric redshifts of quasars in narrow-band surveys

Carolina Queiroz de Abreu Silva

Advisor: Prof. Dr. Luis Raul Weber Abramo

Thesis submitted to the Physics Institute of the
University of São Paulo in partial fulfillment of the
requirements for the degree of Doctor of Science.

Examining Committee:

Prof. Dr. Luis Raul Weber Abramo (IF/USP)

Prof. Dr. Marcos V. Borges Teixeira Lima (IF/USP)

Prof. Dr. Martín Makler (CBPF)

Prof. Dr. Olivier Ilbert (LAM, France)

Prof. Dr. Rogério Riffel (UFRGS)

São Paulo
2021

Declaração de responsabilidade

As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade da autora e não necessariamente refletem a visão da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Àquela que me inspira a reencontrar
a luz em meio às noites mais escuras:
minha amada mãe.

Agradecimentos

Fato: nunca pensei que esse dia chegaria tão rápido. Confesso que não me sinto muito diferente do que era ontem, quando ainda era apenas uma garota, repleta de expectativas em relação ao mundo, afinal esse é só o fim de mais uma etapa que me permitirá alcançar voos ainda mais altos. E, no entanto, tudo mudou. Porque, ao longo dessa jornada, os desafios foram muito além do que simplesmente atingir objetivos acadêmicos e os ganhos são para toda a vida. Essa foi uma jornada sobre aprender a confiar no meu potencial, sobre me maravilhar com a possibilidade de contagiar outras pessoas com a minha paixão em desvendar os mistérios do Universo, sobre refletir na importância do meu papel como divulgadora de ciência.

E, em meio a um ano tão conturbado, a tantas incertezas quanto ao futuro da educação e da ciência, não só no nosso país como no mundo todo, acho que consigo resumir esse misto de sentimentos em uma única palavra: gratidão. Gratidão por ter encontrado forças para não desistir em meio a tantos obstáculos; gratidão por ter tido a oportunidade de conviver com pessoas tão sábias, dentro e fora do mundo acadêmico, e que me inspiraram a seguir em frente; gratidão por também poder dar a minha contribuição (ainda que singela) a esse oceano de conhecimento. Sou muito grata a todos vocês que me apoiaram e me aconselharam ao longo desses anos, permitindo, assim, a realização de mais esse sonho.

Primeiramente, agradeço o Prof. Raul pela paciência, confiança e orientação. Seu entusiasmo e vasto conhecimento em cosmologia me cativaram e me motivam desde o mestrado.

Agradeço o suporte financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), através dos processos 2015/11442-0 e 2019/06766-1.

Agradeço o trabalho de todos os membros dos projetos J-PAS e S-PLUS envolvidos na obtenção e calibração dos dados utilizados nesse trabalho.

Agradeço também a Universidade de São Paulo e o Instituto de Física, assim como

o Laboratório de Astrofísica de Marselha (LAM), por proporcionarem a infraestrutura física e computacional necessárias para a realização desse doutorado. E, em especial, o suporte fornecido pelos funcionários de biblioteca, limpeza e secretaria. Meu muito obrigada à Mme. Françoise Maxant por ter me ajudado com os documentos necessários para o estágio no LAM.

A todos os professores que contribuíram para a minha formação; em particular, os professores Laerte Sodré Junior e Roderik Overzier por compartilharem seus conhecimentos em astronomia, o Prof. Matthew Pieri por ter me recebido e orientado durante meu estágio de pesquisa em Marselha, o Prof. Olivier Ilbert por ter disponibilizado a nova versão do código LePhare, e a Profa. Elisa Ferreira por ter me convidado para conhecer o Instituto de Astrofísica do Max Planck. Deixo um obrigada especial para a Profa. Claudia Mendes de Oliveira, pela disposição e amizade.

Estendo meu obrigada a todos os meus colegas e amigos que participaram dessa jornada, pelo ambiente agradável de trabalho, pelas discussões frutíferas, risadas, conselhos, ou simplesmente pelo ombro amigo. Em especial: Henrique Fabrelli, Alberto Molino, Rafael Christ, Natalia Villa Nova, Fabio Cafardo, Maria Luisa Buzzo, Lilianne Nakazono, Arianna Cortesi, Stela Adduci, Rodrigo Voivodic, Caroline Guandalin, Renan Boschetti, Antonio Montero-Dorta, André Alencar, Hugo Camacho, Natalí de Santi, Sean Morrison e Ignasi Pérez i Ràfols.

Por fim, gostaria de expressar minha gratidão aos meus maiores apoiadores: Rafaelly (por quem eu tenho um carinho especial), minhas tias Waldete e Vanir, meu tio Beto (mais conhecido como *Uberto*), meu avô Paulo (que deixou saudades), meu pai Sergio por sempre ter priorizado a minha educação e, não menos importante, Floquinho, por me lembrar todos os dias de que a vida de gato é muito mais fácil. Em especial, meu muito obrigada para a minha maior fã, minha mãe Vilma, pelo apoio incondicional e por me encorajar a seguir os meus sonhos.

“I want to be all that I am capable of
becoming.”

Katherine Mansfield

Resumo

Quasares são núcleos ativos de galáxias alimentados pela acreção de matéria em um buraco negro supermassivo encontrado em seus centros. Eles não apenas são os traçadores de estruturas em larga escala mais brilhantes e com maiores valores de bias, como também acredita-se que eles coevoluem com suas galáxias hospedeiras, tornando-os valiosos tanto para a cosmologia quanto para estudos de evolução de galáxias. Seus espectros no UV-óptico são caracterizados por um contínuo não-térmico e uma série de linhas de emissão largas, que podem ser bem resolvidas com filtros de bandas estreitas. Nesta tese, nós apresentamos um método para estimar redshifts fotométricos (“photo-zs”) precisos para quasares detectados em levantamentos fotométricos de bandas estreitas. Nosso método (apelidado de QPz) encontra, em primeira aproximação, o melhor ajuste para o fotoespectro de quasares em termos dos chamados autoespectros – os modos mais relevantes de variação de quasares de linhas largas e obtidos a partir de uma seleção de espectros do SDSS. Em cada redshift, o fotoespectro é modelado como uma combinação linear de quatro autoespectros “redshiftados” e combinados com uma lei de extinção do tipo lei de potência. Neste trabalho, testamos os quasares confirmados espectroscopicamente que foram detectados no catálogo primário do miniJPAS e nos primeiros dados divulgados com o S-PLUS, assim como construímos e adotamos catálogos de fluxos simulados para demonstrar o potencial que o J-PAS terá no futuro de fornecer photo-zs precisos para quasares. Validamos nosso método comparando sua performance com uma nova versão do código LePhare (ainda não disponível publicamente), que constitui um método padrão de ajuste de modelos espectrais empregado na literatura.

Palavras-chave: Quasares; redshifts fotométricos; levantamentos fotométricos de bandas estreitas; estrutura em larga-escala

Abstract

Quasars are active galactic nuclei powered by the accretion of matter onto a supermassive black hole lying at their centers. They are not only the brightest and most highly biased tracers of large-scale structure, but they are also believed to co-evolve with their host galaxies, making them invaluable sources to both cosmology and galaxy evolution studies. Their UV-optical spectra are characterized by a non-thermal continuum and a series of broad emission lines, which can be well-resolved with narrow-band filters. In this thesis, we present a method for estimating precise photometric redshifts (photo-zs) for quasars detected with narrow-band photometric surveys. Our method (dubbed QPz) finds a first-approximation best-fit model for the quasar photospectrum in terms of the so-called eigenspectra — the most relevant modes of variation of broad-line quasars, computed using as a basis a selection of SDSS spectra. At each redshift the photospectrum is modelled as a linear combination of four redshifted eigenspectra combined with an extinction power law. In this work, we utilized spectroscopically confirmed quasars detected with the miniJPAS primary catalog and the first data release of S-PLUS as test cases, and we also constructed and employed mock catalogues to demonstrate the potential that the future J-PAS will have to deliver precise photo-zs for quasars. We validate our method by comparing its performance with a new version (not publicly available yet) of the LePhare code, which is a standard template-fitting approach employed in the literature.

Keywords: Quasars; photometric redshifts; narrow-band filter surveys; large-scale structure

Contents

1	Introduction	1
2	Large-scale structure in the Universe	7
2.1	Galaxy counts	12
2.2	Quasar target selection	17
2.2.1	SDSS-I/II	18
2.2.2	SDSS-III	20
2.2.3	SDSS-IV	21
2.2.4	Superset	21
2.2.5	Final remarks on quasar selection	22
2.3	Quasar clustering	23
2.3.1	Quasar clustering in eBOSS	26
2.4	Multi-tracer technique	27
2.4.1	Simulating a quasar and ELG photometric survey	31
3	“Twinkle, twinkle, quasi-star”	39
3.1	Black holes as central engines	40
3.1.1	A black hole in our backyard	42
3.1.2	First observation of gravitational waves	43
3.1.3	First image of a black hole	44
3.1.4	Eddington luminosity	45
3.1.5	Black hole accretion	46
3.1.6	A remarkable relation	49
3.1.7	Black hole seeds	53
3.2	The unified model of AGNs	55
3.2.1	Quasars	62
3.3	Searching for quasars	64

3.3.1	Quasar luminosity function	68
3.3.2	Classification using machine learning	73
3.4	Final remarks on quasar classification	86
4	Photometric redshift estimation	89
4.1	Redshifts	94
4.2	Photometric redshift methods	95
4.3	LePhare++	97
4.4	Quasar Photometric Redshifts (QPz)	99
4.4.1	Global eigenspectra	100
4.4.2	Number of eigenspectra	104
4.4.3	Luminosity-redshift dependency	106
4.4.4	Not only photospectra	108
4.4.5	Type-II quasars	108
4.5	Photo-z quality	109
4.5.1	Normalized Median Absolute Deviation (NMAD)	109
4.5.2	Bias	110
4.5.3	Outlier fraction	110
4.5.4	Odds	111
4.5.5	Redshift-magnitude priors	111
4.5.6	Kolmogorov-Smirnov (K-S) statistics	111
4.6	Final remarks on photo-z estimation	112
5	Mock catalogs	113
5.1	Photometry	115
5.2	Synthetic fluxes	116
5.3	Galactic extinction correction	118
5.4	Training sets for classification	119
5.5	Non-observations and non-detections	121
5.6	Final remarks on mock catalogs	122
6	The miniJPAS survey	125
6.1	miniJPAS	127
6.1.1	Aperture choice	130
6.1.2	Aperture corrections	132

6.2	Stellarity	133
6.3	Results	134
6.3.1	Effect of different filter configurations	138
6.3.2	Priors from the luminosity function	140
6.3.3	Variability	140
6.3.4	Luminosity-redshift bin eigenspectra	140
6.3.5	Completeness of the quasar sample	140
6.3.6	Validating the results with mocks	141
6.4	Final remarks on miniJPAS quasars	143
7	S-PLUS	145
7.1	Results	148
7.2	Final remarks on S-PLUS	150
8	Conclusions and perspectives	151
	Appendix A Tables	171

List of Figures

2.1	Scheme of interactions between the different components of the Universe.	8
2.2	Three-dimensional distribution of galaxies obtained with eBOSS.	10
2.3	Theoretical power spectrum and correlation function.	12
2.4	Evolution of quasar linear bias as a function of redshift for different surveys.	14
2.5	Correlation function as a function of the radial and transverse components obtained by 2dFGRS Collaboration.	16
2.6	SDSS photometric system.	19
2.7	Color-color distribution of stars and quasars selected from the SDSS Early Data Release.	19
2.8	Quasar mean occupation functions.	25
2.9	Correlation function for eBOSS quasars.	26
2.10	Posterior distribution for AP parameters, $f\sigma_8$ and linear bias for eBOSS quasars.	27
2.11	Observable Universe.	28
2.12	Polar projection of the footprint for a simulated survey of quasars and ELGs.	32
2.13	Radial selection functions of quasars and ELGs as a function of redshift.	33
2.14	Projections in the $x - y$ plane of seven redshift slices of the first box of a simulated QSO-ELG photometric survey.	34
2.15	Comparison of the estimated monopoles and quadrupoles for quasars and ELGs on the redshift range $1.4 < z < 1.8$.	36
2.16	Comparison of the estimated monopoles and quadrupoles for quasars and ELGs on the redshift range $1.8 < z < 2.2$.	37
2.17	Comparison of the estimated monopoles and quadrupoles for quasars and ELGs on the redshift range $2.2 < z < 2.6$.	38

3.1	Illustration with the different features of an active galactic nucleus (AGN) and their corresponding scales.	40
3.2	Orbits of stars around Sagittarius A*.	43
3.3	Elliptical galaxy M87 depicting a jet and best-model image of M87* shadow.	45
3.4	Radiative and jet modes.	49
3.5	M- σ relation.	50
3.6	M- σ relation for elliptical and spiral galaxies.	51
3.7	Redshift evolution of the black hole accretion rate compared with the star formation rate distribution.	52
3.8	Schematic diagram of the relationship between fuel supply, galaxy growth and black hole growth.	52
3.9	Black hole masses for high redshift ($z > 6$) quasars.	54
3.10	Unified model of AGNs.	56
3.11	Comparison of optical spectra of 7 AGNs and a normal galaxy.	57
3.12	SED fitting for type-I AGNs.	59
3.13	Example of a BPT diagram.	61
3.14	Example of a changing look quasar.	62
3.15	Example of a BAL quasar.	63
3.16	First quasar discovered by Pan-STARRS 1 using the technique of i -dropout.	66
3.17	Schechter luminosity function.	69
3.18	Quasar luminosity function.	72
3.19	Quasar luminosity function at high redshifts.	73
3.20	Example of an artificial neural network architecture.	76
3.21	Star/quasar classification with ANNz2.	78
3.22	Mean classification with ANNz2 for 9 stars and quasars.	79
3.23	Mean classification over all MLMs with ANNz2.	79
3.24	Probability density distribution for 9 quasars and stars classified with ANNz2.	81
3.25	Classification of S-PLUS quasars with random forest.	83
3.26	New representation of the miniJPAS photospectra of a quasar, galaxy and star used as input for the CNN.	84
3.27	Confusion matrices for miniJPAS using CNN.	86
4.1	Optimal filter systems for photo- z estimation.	93

4.2	Magnitude depth and redshift precision as a function of number of filters.	93
4.3	Color-redshift distribution for the templates used in LePhare.	98
4.4	First four global quasar eigenspectra	102
4.5	Comparison of the performance for different numbers of eigenspectra. . .	105
4.6	Quasar eigenspectra in different redshift-luminosity bins.	107
4.7	QPz redshift estimation for quasar spectra.	108
4.8	QPz redshift estimation for type-II quasars.	109
5.1	Example of synthetic Jspectra for four quasars at different redshifts. . . .	117
5.2	S/N distribution in the mocks.	118
5.3	Number of observed filters in miniJPAS as a function of the tile.	122
5.4	Histograms with the number of filters with non-observations and non-detections as a function of the r -band magnitude.	123
6.1	J-PAS photometric system.	126
6.2	Quasars detected with miniJPAS in the AEGIS fields.	128
6.3	Magnitude-redshift distribution of quasars in miniJPAS.	129
6.4	Estimated depths in all filters for miniJPAS.	129
6.5	Some Jspectra for quasars.	130
6.6	Comparison of photometric performance with QPz for different apertures.	131
6.7	Aperture corrections applied to the miniJPAS data as a function of the tile.	133
6.8	Stellarity parameter for miniJPAS quasars as a function of the tile.	134
6.9	QPz model fit for quasars.	135
6.10	Comparison of the photo- z performance of QPz and LePhare.	136
6.11	Comparison of the spectroscopic with the photometric redshift distribution for QPz and LePhare.	136
6.12	Examples of PDFs obtained with QPz.	137
6.13	Cumulative distribution of photometric errors as a function of magnitude and redshift.	138
6.14	Cumulative distribution of outlier fraction as a function of magnitude and redshift.	139
6.15	Quasar completeness.	141
6.16	Photometric redshift performance with a mock quasar catalog.	142
7.1	S-PLUS photometric system.	145

7.2	Main quasar emission lines at different redshifts in the S-PLUS filter system.	146
7.3	S-PLUS footprint.	147
7.4	Examples of S-PLUS photospectra for different sources.	147
7.5	Magnitude-redshift distribution of quasars in S-PLUS.	148
7.6	Scatter plot obtained with QPz for S-PLUS quasars after quality cut.	149
7.7	χ^2 distributions for QPz and LePhare obtained for the S-PLUS quasars.	149
7.8	Scatter plot for S-PLUS quasars.	150

List of Tables

2.1	Visual inspection classifications from the Superset catalog.	22
3.1	Purity and completeness for S-PLUS quasars classified with ANNz2. . . .	80
3.2	Purity and completeness for S-PLUS stars classified with ANNz2.	80
4.1	Summary of four multi-fiber facilities under construction.	90
4.2	Eigenvalues of the global quasar eigenspectra.	101
4.3	Photometric results with 4, 5, 10 and 50 eigenspectra.	106
5.1	F-score values of the combined performances of different machine learning methods in the classification of (miniJPAS) sources from the test sample of the mocks.	121
6.1	miniJPAS pointings.	127
6.2	Photometric results for miniJPAS quasars using different apertures.	131
A.1	Best-fit parameters for quasar mean occupation functions.	171
A.2	Summary of the masses of high redshift quasars.	172
A.3	Main emission lines of quasars.	173
A.4	Central wavelengths of J-PAS filters.	174

CHAPTER 1

INTRODUCTION

“Those of us engaged in scientific research generally do it because we can’t help it - because nature is the biggest and most complicated jumbo holiday crossword puzzle you have ever seen.”

Edward Hinds, *New Scientist* Sept. 1997

Modern cosmology has witnessed an enormous growth in data acquisition, accompanied by dazzling advancements in its ability to provide answers to fundamental questions about the composition and evolution of the Universe. Yet, many challenges remain, notably those concerning the relations between galaxy formation and the evolution of the Universe on large scales.

Galaxy surveys are the primary tools that help us shed light on some of these puzzles, allowing us to map the large-scale structure (LSS), to measure the matter power spectrum, and to estimate how galaxies and other tracers are related to the underlying dark matter density. The next generation of surveys will allow us to peer wide (i.e. map large areas and volumes) and deep (detect faint, distant sources) into space with increasing resolution, probing the sky through electromagnetic waves, gravitational waves, cosmic rays, and neutrinos – just to mention a few. This will open the opportunity for synergies between the various experiments, where the combination of different messengers and different tracers will improve our ability to understand more completely these astrophysical phenomena.

Surveys of the large-scale structure can be divided into two main types, according to the techniques used to map the sources: spectroscopy and photometry. The former relies on a spectrograph to acquire the spectral energy distribution of sources, and usually operates in the optical and near-infrared¹ part of the electromagnetic spectrum. Spectro-

¹Here we give two references of IR surveys: Spitzer (a space telescope dedicated to imaging and spectroscopy; [Werner et al. 2004](#)) and 2MASS (imaging; [Skrutskie et al. 2006](#)).

scopic surveys provide the 3D positions of galaxies and are able to identify key features in their spectral energy distributions, estimating with high accuracy their spectral types and redshifts. However, they are limited both by the need for a campaign of target acquisition and by the combination of throughput and sensitivity of the spectrograph, which usually require large exposure times. Therefore, this technique tends to demand a high level of financial resources in order for the survey to acquire useful spectra for a large number of sources. Photometric surveys, on the other hand, overcome these constraints by employing band filters to count all the photons inside certain windows of the electromagnetic spectrum. This allows the detection of larger numbers of objects, as well as fainter sources, but it also means that the light of the source is only observed through a handful of filters (resulting in the so-called “photospectra”), compromising the efficacy to distinguish between point-like sources and degrading the estimation of their redshifts.

These issues can be partly resolved by employing multi-band narrow filters (by “narrow” here we mean filters with widths of $\sim 100 \text{ \AA}$). Obviously, on one hand, using a large number of narrow filters, at a fixed exposure time, decreases significantly the photometric depth (Benítez et al. 2009) given the time constraints for running the survey. However, on the other hand, it also brings an important gain in the power of breaking color-redshift degeneracies, improving the accuracy with which the photometric redshifts can be estimated, and also the ability to classify sources using photometric data alone.

Many projects have made use of a large number of narrow-bands, e.g. COMBO-17 (Wolf et al. 2003), COSMOS (Scoville et al. 2007a), ALHAMBRA (Moles et al. 2008), SHARDS (Pérez-González et al. 2013), and PAU (Eriksen et al. 2019). In particular, the J-PLUS (Cenarro et al. 2019) and S-PLUS (Mendes de Oliveira et al. 2019) surveys are mapping large areas of the Northern and Southern hemispheres with 12 filters, including 7 narrow-bands centered on prominent stellar spectral features. Even more enthralling is the photometric system of the *Javalambre-Physics of the Accelerating Universe Astrophysical Survey* (J-PAS; Benitez et al. 2014), which consists of a unique set of 54 contiguous narrow-band filters. The project is about to start and will deliver low-resolution spectra for millions of galaxies and quasars over thousands of square degrees, achieving the highest redshift precisions of any photometric survey so far.

Active galactic nuclei (AGNs) – or, in a broader sense, quasars – are of high interest to astronomy: they are not only the brightest and most highly biased tracers of large-scale structure, but their astrophysical properties (energy powered by the accretion of matter onto a supermassive black hole) make them a key ingredient in galaxy evolution models. Quasars also work as lighthouses, serving as background light sources to map the intervening neutral hydrogen gas through Gunn-Peterson effect ([Gunn & Peterson 1965](#)), resulting in the so-called Lyman- α forest.

Quasar candidates typically appear as point-like sources in photometric imaging and, thus, they can be easily confused with stars and even unresolved galaxies. This issue on population mixtures was first identified in the context of the use of broad-band imaging to pre-select spectroscopic targets for the Sloan Digital Sky Survey ([Richards et al. 2009](#); [Leistedt et al. 2013](#); [Leistedt & Peiris 2014](#)). However, this pre-selection is unable to avoid contamination by other sources in color-magnitude and color-color diagrams, leading to sub-optimal source classification and lower target success rate, with the effect that fibers end up allocated only to the brightest, most clearly distinguished quasars.

The J-PAS photometric system, on the contrary, will be able to resolve the broad emission lines of type-I quasars (as well as most broad absorption line objects), and to detect the narrow-lines of many type-IIIs. [Abramo et al. \(2012\)](#) showed that J-PAS will have the potential to observe nearly ~ 240 quasars per square degree for a limiting magnitude of $g < 23$, which means that a J-PAS survey of quasars could be the largest and most complete of its kind, having a significant impact in terms of cosmological applications.

Prior to the arrival of the final scientific instrument (JPCam; [Taylor et al. 2014](#); [Marín-Franch et al. 2017](#)), the J-PAS telescope was equipped with a single CCD camera, called JPAS-Pathfinder, with which it was possible to carry out the first observations and test the performance of the optical system. This survey, dubbed miniJPAS ([Bonoli et al. 2020](#)), imaged nearly 1 deg^2 in the Extended Groth Strip (AEGIS) fields using 60 filters, and was ideal to forecast the photometric precision that we expect to obtain with J-PAS once it becomes fully operational.

In this thesis we search for quasars in the miniJPAS and S-PLUS (DR1) data sets²,

²All magnitudes here are presented in the AB system. Throughout this work, we have adopted the standard Λ CDM cosmology provided by [Planck Collaboration et al. \(2020a\)](#) with parameters $(h_0, \Omega_m, \Omega_\Lambda, \Omega_k) = (0.674, 0.314, 0.686, 0.00)$.

and characterize these photometric samples in terms of some techniques available to us, such as machine learning algorithms for classification and identification of quasar candidates, optimized codes for photometric redshift estimation, Bayesian approaches to help break degeneracies between pairs of emission lines, and methods to analyse multi-tracer catalogs. The availability of spectroscopic redshifts for galaxies and quasars in the mini-JPAS and S-PLUS footprints makes those data sets ideal test cases to characterize the performance of photometric redshift (photo-z) estimates for different magnitudes and redshift ranges.

Our ultimate goal is to combine the byproducts of those techniques to assemble fully probabilistic catalogs, with probabilities for each object of being of a certain type (e.g. star, galaxy or quasar)³, and estimate the photo-zs for those identified with a high probability of being quasars. In particular, in this work we focused on the estimation of photometric redshifts for quasars that had been previously classified by spectroscopic surveys. The classification of point-like sources and consequent identification of quasar candidates constitutes an ongoing project that has been currently treated in a joint-effort with other members of the J-PAS and S-PLUS Collaborations, and hence is outside the scope of this thesis.

For the quasar photo-z estimation, we followed [Yip et al. \(2004\)](#) and implemented a technique that models the quasar fluxes through a linear combination of amplitudes of the principal components of quasar spectra, including an extinction power law to allow for the slope of the eigenspectra to adjust to and fit more reddened quasars. In the end, our method finds a first-approximation best-fit model for the quasar photospectrum in a 5-dimensional space, while providing full probability distribution functions for their redshifts.

In order to validate the photo-z results obtained with our code (dubbed QPz; [Queiroz et al. 2021a, in prep.](#)), we compared its performance with an improved version of the standard template-fitting code LePhare ([Ilbert et al. 2006](#); [Arnouts et al. 1999](#)). Our analyses indicate that the precision achieved with QPz is very similar to that obtained with LePhare; the performance of the former being, however, superior in terms of the

³In the case of stars, we consider main sequence, white dwarfs, carbon stars and cataclysmic variables (CV). Note, however, that we are not distinguishing the different stellar types nor separating galaxies between red and blue.

outlier fraction.

By using mock catalogs (Queiroz et al. 2021b, in prep.), we also forecast the impact of the science that will result from large-area quasar catalogs produced on the basis of J-PAS data. These mocks are produced by convolving SDSS DR16 spectra with the J-PAS photometric system, and employing the same signal-to-noise ratio expected for point-like sources detected on the AEGIS fields with miniJPAS (for that purpose we employ the 3''-aperture magnitudes). Those mock quasar catalogs have been employed in many other projects within the J-PAS collaboration, and in particular for simulating lensed quasars. Finally, this same machinery has been used to produce mock catalogs for galaxies and stars as well, which are an essential part for the training and test sets used for quasar classification with machine learning.

In addition to the science that J-PAS will be able to conduct using the quasars it identifies, the collaboration will join efforts with the WEAVE survey (Dalton 2016), which is a multi-object spectrograph that will start observing in 2021. Part of the WEAVE strategy is to follow-up high-redshift ($z > 2.1$) quasars in order to conduct a Lyman- α forest and metal line absorption survey (Pieri et al. 2016). The targets for this WEAVE-QSO survey will be provided mainly by J-PAS, which is the only instrument in the world capable of identifying quasars in numbers and down to the depths needed by WEAVE to do its science. For this reason, an additional question that we want to answer regards the critical combinations of filters able to detect the main quasar features. In order to provide quasar targets with nearly 100% completeness and purity for the WEAVE-QSO survey, J-PAS will probably observe with only half of the narrow-bands during the first years of observations. Therefore, determining how the photo- z precision degrades as a reduced number of filters is observed was also very pressing, since it affects the survey strategy as a whole.

Finally, our preliminary results (Abramo, Queiroz et al. 2021 in prep.) show that the inclusion of photometric errors for quasars in multi-tracer catalogs produce some interesting effects on the power spectrum on large scales, which could be detected with J-PAS. The impact of the science that will result from quasar catalogs produced on the basis of multi-band photometric data includes the measurement of baryon acoustic oscillations

and redshift-space distortions, a better understanding of the environments where quasars are found (e.g., by determining the halos they inhabit, or whether their bias is scale-dependent), and information about both the density and luminosity evolution of quasars, especially at high redshifts and at the faint end. All this potential points to an exciting future with quasar research.

As a suggestion of reading: Chapters 2 and 3 contain a detailed introduction on cosmology with quasars and quasar physics, respectively; while Chapters 4, 5 and 6 contain the main contributions and results of this thesis.

CHAPTER 2

LARGE-SCALE STRUCTURE IN THE UNIVERSE

The past decade has witnessed some astonishing advances in cosmology, underpinned by the quantity and precision of cosmic structure observations, and stimulated by the possibility of having unique insights about the physics of the Universe. We are now exploring the cosmic frontier, and 2019 Nobel prize to James Peebles “for contributions to our understanding of the evolution of the Universe”¹ has drawn even more notability to the field.

The birth of modern cosmology dates back to around the 1920s-1930s when most of its key elements were forged: the General Theory of Relativity of Albert Einstein; the first results of Friedman-Lemaître-Robertson-Walker (FLRW) metric describing the dynamics of a homogeneous and isotropic distribution of matter; and Edwin Hubble’s findings on the first observational evidence of a correlation between distance and recession velocity of galaxies – i.e., the expansion of the Universe.

Thenceforth, several observational discoveries were made, being worth mentioning:

- the detection of the cosmic microwave background (CMB) radiation ([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 3.8×10^5 years old;
- the overwhelming indirect evidence for the existence of an exotic type of “dark” matter, that does not interact via the electromagnetic force, after numerous indica-

¹Copywrite ©The Royal Swedish Academy of Sciences (<https://www.nobelprize.org/>)

tions that ordinary matter alone was not sufficient to account for the gravitational effects detected, for instance, in the rotation curves of galaxies, in the velocity distributions of galaxies in clusters, in image distortions by gravitational lensing, and in extended X-ray emission from clusters of galaxies; and

- the accelerated expansion of the Universe from the measurement of light curves of distant Type-1A supernovae (Riess et al. 1998; Perlmutter et al. 1999).

These discoveries have contributed to consolidate and shape the current paradigm of cosmology, the so-called Λ CDM model, that states that we live in an expanding Universe, homogeneous and isotropic at large scales, which today contains three major components: a cosmological constant (Λ), that appears as a minimalistic modification to Einstein's equations and constitutes the simplest type of dark energy; cold dark matter (CDM); and ordinary (baryonic) matter. Fig. 2.1 schematizes the various ways in which the different components of the Universe interact with each other, as well as with other components which were more dominant at earlier times. The gravitational sector is governed by the Einstein equations, and the remaining interactions between the particles are described by the Standard Model of particles and fields in what concerns the Electroweak and Nuclear forces. This complex ensemble of interacting particles and fields is then described by Boltzmann equations.

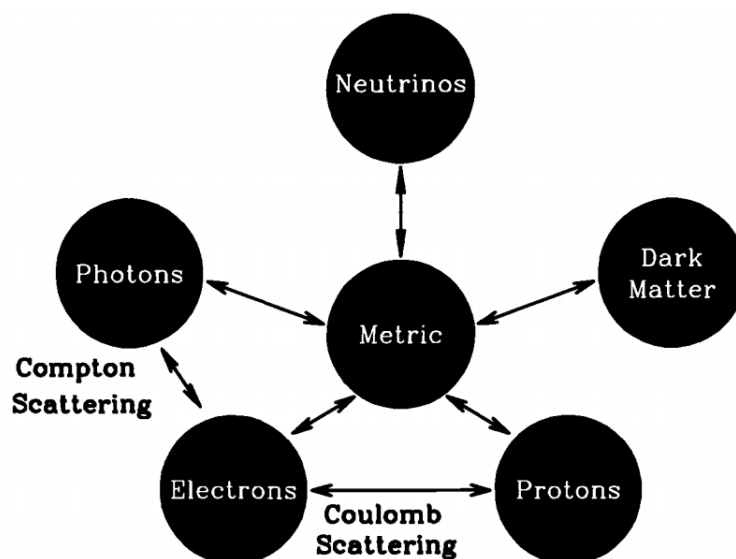


Figure 2.1: Scheme of interactions between the different components of the Universe. Figure extracted from Dodelson (2003).

Understanding the nature of dark matter and dark energy (which presently account

for more than $\sim 95\%$ of the content of the Universe), the dynamics of structure formation on large-scales, and the interplay between the dark sector and the accelerating expansion of the Universe [in which the distances between the galaxies are described in terms of a function of time – the scale factor $a(t)$] remain as major challenges in modern cosmology, motivating numerous investigations in the last decades.

Most of these investigations were only possible thanks to the great advances in some important complementary areas, such as astronomical instrumentation, highly sensitive detectors, data reduction software, dynamic data storage, advances in numerical simulation techniques, and high-speed processors.

The complexity of cosmological phenomena is also manifested on small and intermediate scales, where the Universe is far from homogeneous. Galaxy surveys have revealed the existence of filaments and clusters of luminous matter, surrounded by enormous voids, showing that the galaxies are clearly not randomly distributed: instead, their positions are correlated in space. Hence, by combining the images of the distribution of galaxies or clusters of galaxies with their spectroscopic observations it is possible to learn how these systems formed and evolved in time, shedding light on the mechanisms that led to the cosmic web of structures that we see today (illustrated in Fig. 2.2).

Within this framework, galaxy surveys actually have access to only a particular window of the sky, composed of an angular mask of the observed area and a radial distribution of the (biased) tracers of the underlying matter density field, $\rho(\mathbf{r}, t)$. By mapping the distributions of these tracers it is possible to catalog their positions, and observations may also provide us with properties such as spectral types, colors, stellar masses and shapes. Finally, in order to have an appropriate statistical description of their spatial distributions, one can either use the correlation function, or its Fourier transform – the power spectrum. Although these two-point functions contain most of the cosmological information, it has recently become more popular to also explore higher-order correlations such as the 3- and 4-point functions – see, e.g., [Planck Collaboration et al. \(2020a\)](#) in the context of the CMB; [Planck Collaboration et al. \(2020b\)](#) for constraints on primordial non-Gaussianity.

The anisotropies observed in the CMB spectrum, consisting of small temperature

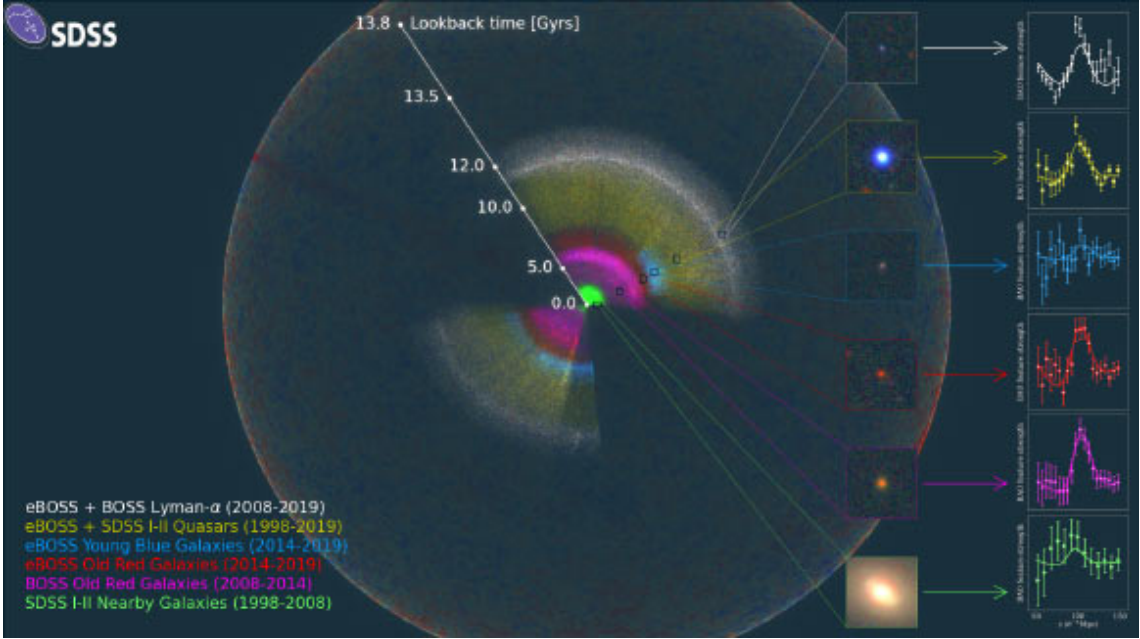


Figure 2.2: Three-dimensional distribution of galaxies depicted by different colors within the observable Universe (outer sphere, showing fluctuations in the cosmic microwave background) from the Extended Baryon Oscillation Spectroscopic Survey (eBOSS), an SDSS Collaboration. The observer is located at the center of this map, and as we look out in distance, we look back in time. Denser regions depict the clumpy, filamentary “cosmic web”. Galaxies at larger distances are fainter and hence harder to detect. The two slices correspond to nearly opposite directions in the sky; yet, the statistical fluctuations seem to be very similar on each strip. The inset for each color-coded section of the map includes an image of a typical galaxy or quasar from that section, and also the signal of the (BAO) pattern that the eBOSS team measures there. Credit: Anand Raichoor (EPFL); Ashley Ross (Ohio State University); SDSS Collaboration (2020).

fluctuations of $\delta T/T \sim 10^{-5}$, suggest that the structures that we see today started from small inhomogeneities. In order to study how the density fluctuations evolve in space and in time, it is customary to define the relative density contrast (or overdensity) field:

$$\delta(\mathbf{r}, t) = \frac{\rho(\mathbf{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (2.1)$$

where $\bar{\rho}(t)$ is the mean density at a given time t .

The density contrast is believed to be the outcome of some random process in the early Universe, and it grows in amplitude due to gravitational interactions: it is positive in overdense regions and negative in underdense regions. Its minimum possible value is $\delta = -1$, and, in principle, there is no upper limit on δ .

Alternatively, we can also express the density fluctuation field $\delta(\mathbf{r})$ within a large comoving box of comoving volume V , in terms of the individual Fourier components $\delta_m(\mathbf{k})$ of the density contrast:

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \delta_m(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k. \quad (2.2)$$

The properties of the field $\delta(\mathbf{r})$ are determined by specific processes occurring during the inflationary phase in the very early Universe and subsequent evolutionary processes preceding recombination. If the phases of the different Fourier components are uncorrelated, or almost uncorrelated, and if the amplitudes of the modes $\delta_{\mathbf{k}}$ are nearly Gaussian (which is a good approximation on large scales), then we say that $\delta(\mathbf{r})$ is a Gaussian random field.

In terms of the density contrast, we can define the two-point correlation function:

$$\langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle = \xi(\mathbf{r}), \quad (2.3)$$

and the Fourier counterpart of this two-point function, the matter power spectrum $P_m(k)$ (Peebles 1980; Peacock 1999):

$$\langle \delta_m(\mathbf{k}, z)\delta_m^*(\mathbf{k}', z) \rangle = (2\pi)^3 P_m(k, z)\delta_D(\mathbf{k} - \mathbf{k}') \quad (2.4)$$

where $\delta_m^*(\mathbf{k}) = \delta_m(-\mathbf{k})$, δ_D is the Dirac delta function, and $\langle \dots \rangle$ denotes an ensemble average².

The matter power spectrum is one of the main products of galaxy catalogs, encompassing a lot of information with regard to the formation of structures over a range of scales. A key feature is the imprint of “wiggles” (see left panel of Fig. 2.3), which correspond to the acoustic scale of baryon-photon decoupling during the recombination epoch. These are the so-called baryon acoustic oscillations (BAOs). Such scale can be measured as a slightly bigger probability of finding galaxies separated by this characteristic distance, which can therefore serve as a statistical standard-ruler. Today, its value is approximately 148 Mpc (illustrated on the right panel of Fig. 2.3).

In practice, in order to compute the correlation function what galaxy surveys really

²Note that, although the theory specifies an ensemble average, observationally we have access to only one realization of the random process (i.e. we only have access to one Universe). Nevertheless, on large scales the density field is well approximated by a Gaussian distribution, having the power spectrum as the second moment (i.e. variance) of the distribution. Then, by the Ergodic hypothesis, the ensemble average becomes equivalent to a spatial average over one realization of the random field.

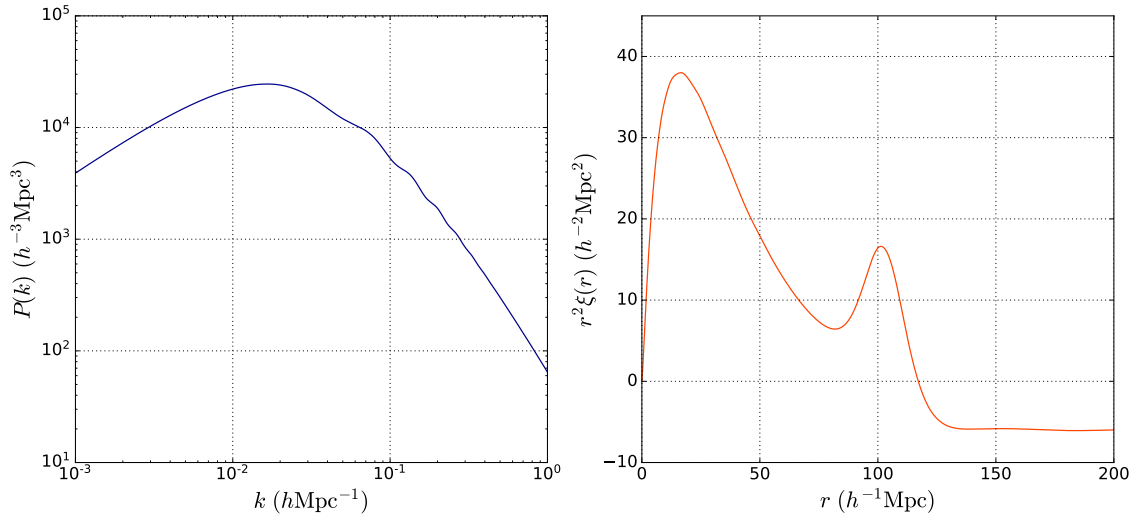


Figure 2.3: Left panel: theoretical power spectrum at $z = 0$ for the Λ CDM model generated with the Code for Anisotropies in the Microwave Background (CAMB, Lewis et al. 2000). Right panel: corresponding two-point correlation function computed with the COLOSSUS Cosmology toolkit (Diemer 2018). The second peak corresponds to the BAO scale.

do is to estimate how much the galaxy sample deviates from a homogeneous distribution when compared to a control (“random”) set of objects. A popular estimator was introduced by Landy & Szalay 1993, and it is still very employed nowadays.

2.1 Galaxy counts

The current standard hierarchical structure formation scenario states that objects in the Universe form through the gravitational collapse of small (primordial) fluctuations in the density field (Press & Schechter 1974; White & Rees 1978). These collapsed structures form clumps dominated by dark matter, and grow in mass via the accretion of matter from their surroundings or through mergers with other halos.

Since dark matter only interacts gravitationally and, as the name suggests, it does not emit light, it is not straightforward to probe the total matter distribution. In fact, galaxy surveys only detect tracers of the underlying matter density field, such as galaxies, quasars, and clusters. This relationship was clarified by Bardeen et al. (1986), who showed that the density peaks of a Gaussian field are related to the density fields of the matter halos through a “halo bias”. This means that those gravitational collapsed structures are biased in relation to the matter distribution – i.e., their correlation functions and power spectra are biased with respect to the matter power spectrum, in such a way that

halo bias grows steeply with halo mass.

Although the connection between galaxy positions and the three-dimensional power spectrum for dark matter is not direct (since only luminous matter can be observed), galaxies, in turn, also work as biased tracers of peaks of the underlying density field. In other words, the dark matter density contrast is related to the tracer's density field through a bias b :

$$\delta_g = b \delta_{DM} \quad (2.5)$$

where a value of $b \neq 1$ means that the luminous matter does not follow the same distribution as the total matter. Note that we are assuming here the approximation of a linear, scale-independent bias (which should be valid on large scales). Nevertheless, in principle the bias is also a function of the redshift and luminosity: more luminous, and therefore rarer, objects seem to be more highly biased.

The bias can, thus, be interpreted as a measure of the degree of clustering, since tracers with higher values of bias are more likely to be found near the highest density peaks in the mass distribution. Or, equivalently, we say that more massive galaxies are more strongly clustered.

This seems to be the case for quasars. For instance, the two-point clustering of confirmed quasars from the final sample of the Baryon Oscillation Spectroscopic Survey (BOSS) has provided a precise measurement of $b_q = 3.54 \pm 0.10$ for the quasar bias over the redshift range $2.2 \leq z \leq 2.8$ (Eftekharzadeh et al. 2015). In Fig. 2.4 we show the values of quasar linear bias as a function of redshift obtained by different cosmological surveys.

At this stage, one may also find useful to define the galaxy density contrast in terms of the galaxy counts as a function of position:

$$\delta_g(\mathbf{r}) = \frac{N_g(\mathbf{r}) - \bar{N}_g(\mathbf{r})}{\bar{N}_g(\mathbf{r})}. \quad (2.6)$$

By computing the galaxy correlation function:

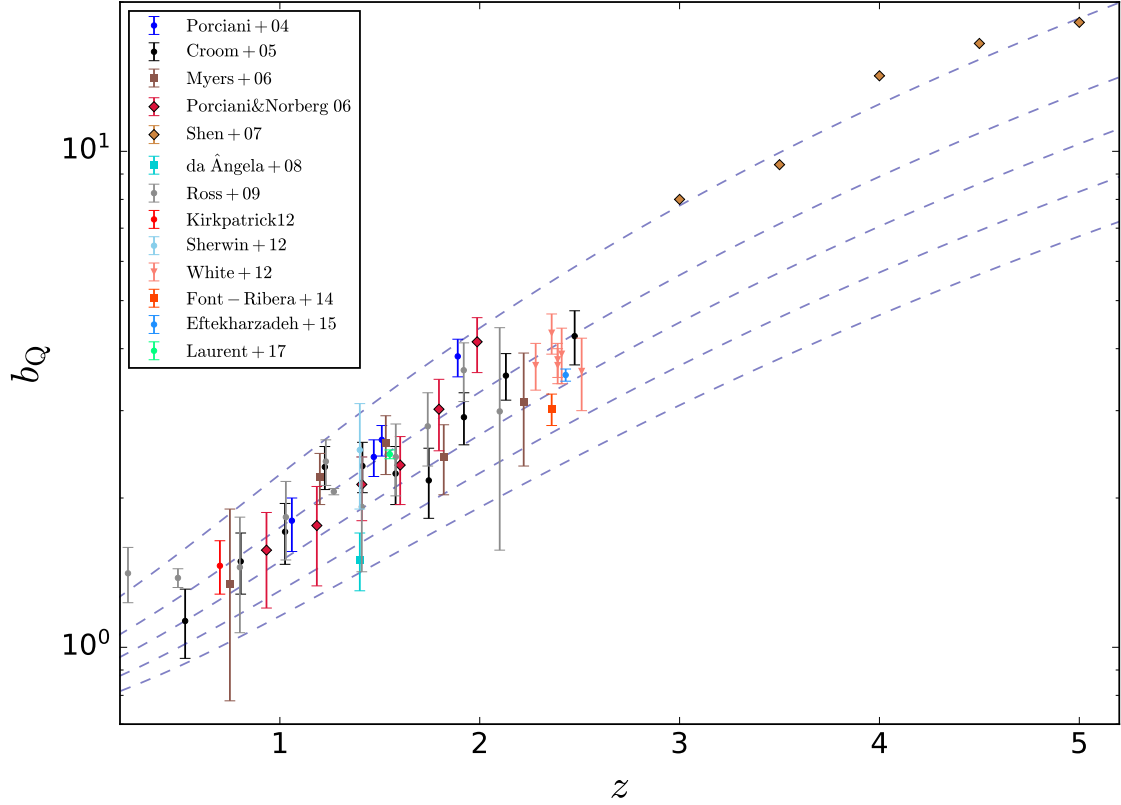


Figure 2.4: Evolution of quasar linear bias as a function of redshift for different surveys. Results compiled from: Porciani et al. 2004, Croom et al. 2005, Myers et al. 2006, Porciani & Norberg 2006, Shen et al. 2007, da Ângela et al. 2008, Ross et al. 2009, Kirkpatrick 2012, Sherwin et al. 2012, White et al. 2012 Font-Ribera et al. 2014, Eftekharzadeh et al. 2015, and Laurent et al. 2017. The bias values were not recalibrated and may depict slightly different cosmologies. The dashed lines show the linear halo bias evolution using the formalism of Tinker et al. 2010 and for the following values of mass (from bottom to top): $\log h^{-1}M_{\odot} = 11.7, 12, 12.3, 12.6,$ and 13 .

$$\xi_{gg}(\mathbf{r}) = \langle \delta_g(\mathbf{x})\delta_g(\mathbf{x} + \mathbf{r}) \rangle = b_g^2 \xi(\mathbf{r}) + \frac{1}{\bar{n}_g(\mathbf{r})} \delta_D(\mathbf{r}) \quad (2.7)$$

we obtain the correlation function of matter $\xi(\mathbf{r})$ in the first term, and the shot-noise in the last term, which appears because we are describing counts of galaxies (i.e., a discrete distribution) in terms of a point process sampling (which we assume to be Poissonian) from a continuous density field.

The two-point correlation function is the most popular tool for large-scale clustering analysis. The quantity $\xi_{gg}(\mathbf{r})$ can be described in terms of the probability of finding a galaxy within a volume V lying at a distance \mathbf{r} from an arbitrarily chosen galaxy:

$$dP = \bar{n}_g [1 + \xi_{gg}(\mathbf{r})] dV. \quad (2.8)$$

Note that a completely random distribution yields $\xi_{gg}(\mathbf{r}) = 0$.

Equivalently, we can write the galaxy power spectrum as:

$$P_{gg}(k) = b_g^2 P_m(k) + \frac{1}{\bar{n}_g}. \quad (2.9)$$

In order to properly characterize the distribution of galaxies we need information in 3D, which can be provided by galaxy redshift surveys. However, when galaxy distances are measured in redshift space (which is the case for essentially all surveys nowadays), their peculiar velocities distort the pattern of galaxy clustering by displacing the galaxy positions along the line-of-sight. These are the redshift space distortions (RSDs), which have two limits:

- on large scales, the peculiar velocities reflect the (linear) infall motions towards overdensities, causing structures in real space to appear “squashed” along the direction of the line-of-sight in redshift space. This is often called the Kaiser effect ([Kaiser 1987](#));
- on small scales, peculiar velocities reflect the (non-linear) virialized motions of galaxies inside their host halos, causing those structures to appear “stretched” along the line-of-sight in redshift space. This is often called the Fingers-of-God effect.

Figure 2.5 illustrates the effects of redshift space distortions in the correlation function for the 2dFGRS.

Consider, for instance, a pair of two galaxies separated by comoving distances s_1 and s_2 , and with an angular separation θ . This allows us to express the two-point correlation function $\xi^s(\pi, \sigma)$ in terms of the pair’s comoving radial ($\pi = |s_1 - s_2|$) and transverse ($\sigma = (s_1 + s_2)\theta/2$) separations. Now, since the effects of RSDs are limited to the line-of-sight direction, we can integrate out the effects of peculiar velocities and redshift errors by integrating $\xi^s(\sigma, \pi)$ over π . This results in the projected correlation function:

$$w_p(\sigma) = \int_0^{\pi_{cut}} \xi^s(\sigma, \pi) d\pi \quad (2.10)$$

where π_{cut} is the scale at which the redshift space distortions can be neglected.

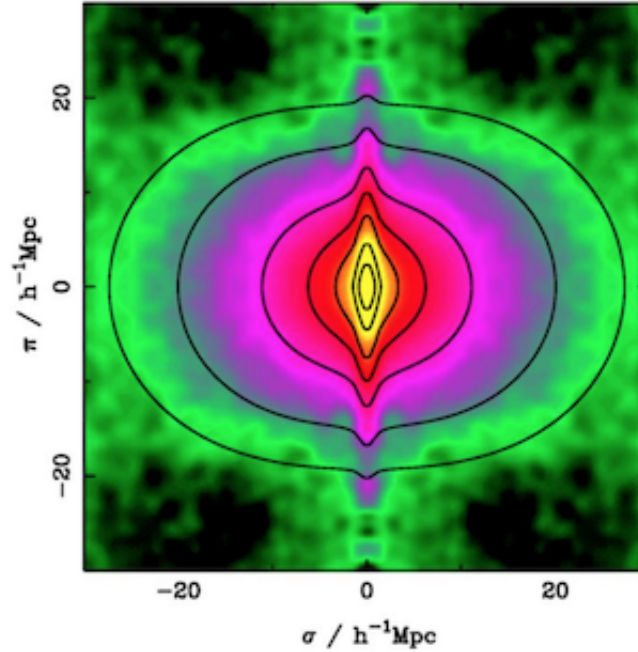


Figure 2.5: Correlation function as a function of the radial (π) and transverse (σ) components obtained by 2dFGRS Collaboration. Note the effects caused by the redshift space distortions: the “finger-of-God” radial stretching at small separations, and the apparent flattening of structures at large separations. Credit: [Peacock et al. 2001](#).

These distortions have the advantage of carrying information about the dynamics of the galaxies inside the large-scale density field, and they can also be used to measure the growth of structures. Therefore, the effects can be employed as a test to distinguish, for instance, between dark energy and modified gravity models ([Joyce et al. 2016](#); [Aparicio Resco et al. 2020](#)). The RSDs can be quantified in terms of the matter growth rate f , defined as

$$f = \frac{d \ln(G)}{d \ln(a)}, \quad (2.11)$$

where G is the growth function and a is the scale factor. The growth rate f is a measurement of how fast galaxies are approaching each other or, in other words, it characterizes the strength of the gravitational interactions that drive structure formation. In practice, G is usually parametrized in terms of $f\sigma_8$, where σ_8 is the variance of the density field smoothed on a scale of $8 h^{-1} \text{Mpc}$ (or, in other words, the amplitude of perturbations). The gravitational growth factor can also assume the following parametrization: $f = [\Omega_m(z)]^\gamma$, with $\gamma \approx 0.55$ for ΛCDM .

In the regime of linear theory, [Kaiser 1987](#) showed that the redshift space power

spectrum $P^s(\mathbf{k})$ can be related to the real space power spectrum $P(k)$ through an effective bias $b \rightarrow b + f\mu^2$, which in principle can be a general function of redshift, scale and angle:

$$P^s(k, \mu) = (b + f\mu_k^2) P_m(k) \quad (2.12)$$

where $\mu_k = \hat{k} \cdot \hat{r}$ is the cosine of the angle between the Fourier mode and the line-of-sight.

Since in an isotropic Universe the density spectrum cannot contain a preferred direction, the anisotropy in $P^s(k, \mu)$ must be symmetric under rotations around the line-of-sight, as well as with regards to inversions in μ , i.e. $P^s(k, \mu) = P^s(k, -\mu)$, and thus the distortions will depend only on even powers of μ .

In order to measure the extent of the deviation from isotropy, one can decompose the power spectrum into multipole moments using the Legendre polynomials $L_\ell(\mu)$:

$$P_\ell(k) = \frac{2\ell + 1}{2} \int_{-1}^1 d\mu P^s(k, \mu) L_\ell(\mu). \quad (2.13)$$

In the linear regime, the monopole ($\ell = 0$), quadrupole ($\ell = 2$) and hexadecapole ($\ell = 4$) are sufficient to completely characterize $P_g^s(k)$ (Percival & White 2009):

$$P_{g,\ell=0}(k) = \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) P_g(k) \quad (2.14)$$

$$P_{g,\ell=2}(k) = \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) P_g(k) \quad (2.15)$$

$$P_{g,\ell=4}(k) = \frac{8}{35}\beta^2 P_g(k) \quad (2.16)$$

where $\beta \equiv \frac{f}{b}$ is the redshift space distortion parameter. Note, however, that the estimates of $P_\ell(k)$ rapidly become noisy for multipoles with $\ell > 2$.

2.2 Quasar target selection

This thesis focuses on surveys of photometric quasars; more specifically, on quasar candidates identified with narrow-band filter surveys, such as J-PAS, J-PLUS and S-

PLUS. The main advantage of this type of survey is their capacity to identify a large number of quasars up to redshift $z \sim 5$, thus probing large cosmological volumes with a higher efficiency compared with spectroscopic surveys, and improving significantly the statistics.

However, selecting quasar targets is a more challenging task compared to selecting galaxies. The main reason is that the broad-band photometric colors of quasars are very similar to the colors of stars, with the result that a non-negligible fraction of these photometric candidates are actually found to be stars. Hence, it is difficult to distinguish quasars from stars, and to estimate their redshifts using only broad-band photometric information.

The clustering of quasars has been studied mostly in optical frequencies, using both spectroscopic (e.g. [Ross et al. 2009](#); [Shen et al. 2007](#)) and photometric (e.g. [Myers et al. 2006](#)) catalogs drawn mainly from early SDSS data, and were used to constrain numerous cosmological and astrophysical quantities of interest, such as the quasar bias, primordial non-Gaussianities, and the quasar luminosity function ([Richards et al. 2006b](#); [Myers et al. 2007](#); [Serber et al. 2006](#); [Slosar et al. 2008](#); [Strand et al. 2008](#)).

The Sloan Digital Sky Survey is one of the most successful astronomical surveys to date, imaging the sky in five broad-bands *ugriz* (shown in Fig. 2.6) specially designed for the survey ([Fukugita et al. 1996](#)). In this section we give a brief historical overview of the strategies employed by SDSS in the task of targeting quasars³ for spectroscopy.

2.2.1 SDSS-I/II

The quasar target selection algorithm employed in SDSS-I/II ([Richards et al. 2002](#)) basically explores all known regions of color space occupied by quasars, avoiding those regions where the quasar density is much lower than the density of contaminants. It primarily identifies quasars as outliers (i.e. sources lying more than 4σ) from the stellar locus in color-color diagrams (i.e. *ugri* cubes: $u - g$, $g - r$, $r - i$, or *griz* cubes: $g - r$, $r - i$, $i - z$) and by matching (within $2''$) unresolved sources to the FIRST catalog of radio sources ([Becker et al. 1995](#)) – independently of their colors. The contamination between

³Here we focus on the quasar target selection. As for the redshift measurement, SDSS employs automated pipelines based mainly on fitting templates derived from a principal component analysis (PCA), and the identification of emission lines in the spectra. The SDSS pipeline redshifts are generally accurate in the cases where the ZWARNING flag is 0.

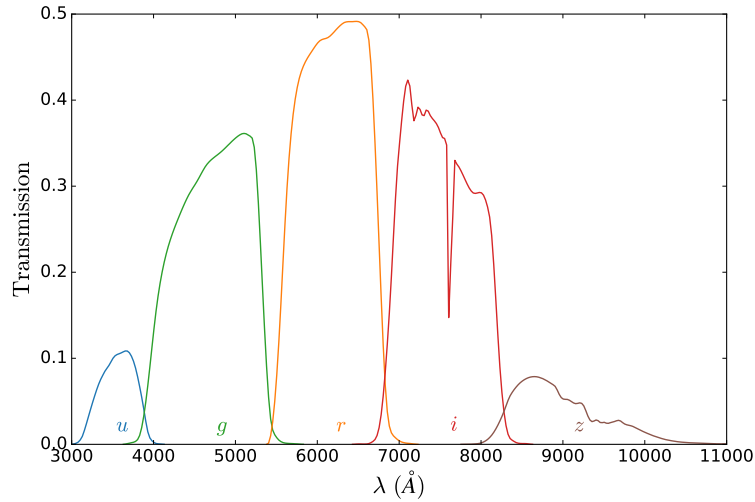


Figure 2.6: SDSS photometric system composed of five broad-bands *ugriz*.

quasars and stars in these color-color diagrams is illustrated in Fig. 2.7.

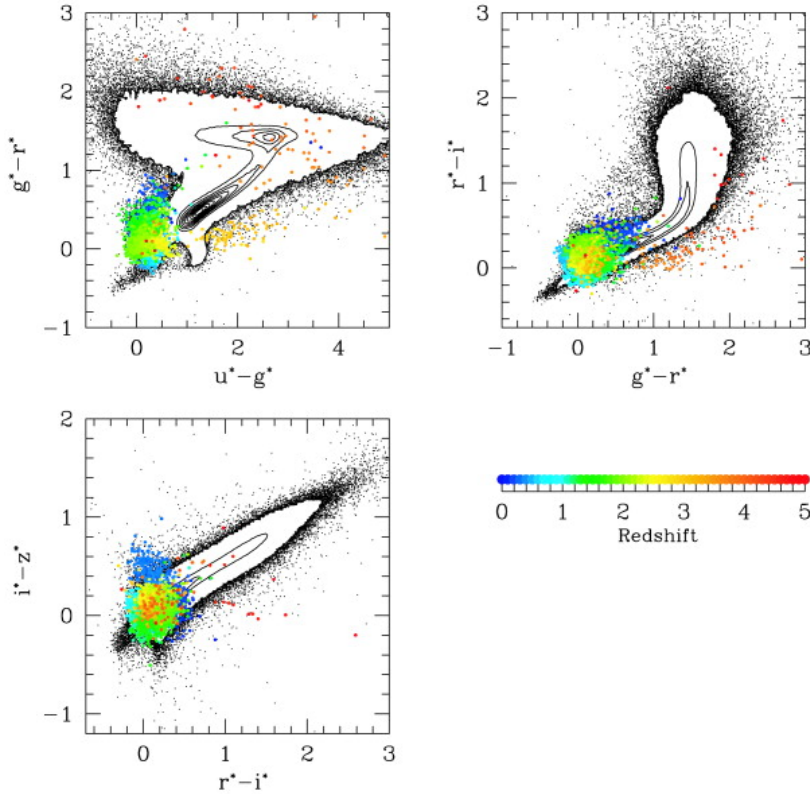


Figure 2.7: Color-color distribution of stars (black contours) and quasars (color-coded by redshift) selected from the SDSS Early Data Release. Credit: [Richards et al. \(2002\)](#).

The algorithm is sensitive to quasars at $z \lesssim 5.8$, and all measurements are performed using PSF magnitudes, with corrections for Galactic extinction following [Schlegel et al. 1998](#). The rejection criteria include objects flagged as having “fatal” or “nonfatal” photometric errors (such as those with nearby bright and/or saturated sources, objects too close to the edge, blended objects or having any interpolation issues), or lying in any of

three color-defined exclusion regions dominated by white dwarfs, A stars, and M star plus white dwarf pairs. In order to exclude very bright sources, which could contaminate the spectra from adjacent fibers, a cut of $i_{\text{PSF}} > 15.0$ is further applied. Still, the fibers were only allocated to the brightest, most clearly distinguished quasars, implying limiting magnitudes of $i_{\text{PSF}} < 19.1$ and $i_{\text{PSF}} < 20.2$ for the main quasar (UV-excess objects) sample and the high-redshift quasar sample, respectively.

Inevitably, this strategy leads to an irregular redshift distribution. The overall completeness for the SDSS-I/II quasar samples is expected to be over 90%, with an overall efficiency (i.e. fraction of quasar targets that end up being real quasars) above 65%.

2.2.2 SDSS-III

Unlike its predecessors, the SDSS-III survey (dubbed Baryon Oscillation Spectroscopic Survey, BOSS), was designed to map the large-scale structure using not only galaxies but also the Lyman- α forest, employing quasars at $2.15 < z < 3.5$. This is a challenging task, because the quasar locus crosses that of main sequence stars at $z \sim 2.7$. In order to overcome this, BOSS quasar targeting (Ross et al. 2012) goes to fainter magnitudes, $g < 22.0$ or $r < 21.8$, increasing significantly the surface density of candidates.

During the first year of BOSS, four different methods were developed to improve quasar selection using photometric data:

- a “Kernel Density Estimation” (Richards et al. 2009), which measures the densities of quasars and stars in color–color space from training sets and uses these to select high probability targets;
- an artificial neural network (Yèche et al. 2010), which takes as input the SDSS photometry and errors from a training set in order to run a classification scheme (star versus quasar) and generate a photometric redshift estimate;
- a “Likelihood” approach (Kirkpatrick et al. 2011) which determines the likelihood of each object being a quasar, given its photometry and models for the stellar and quasar loci;

- an “Extreme Deconvolution” (XDQSO) (Bovy et al. 2011) selection, which performs a density estimation of stars and quasars by incorporating photometric uncertainties. This “radical deconvolution” technique amounts to giving up on any hope of understanding the systematics of the dataset, and marginalizing against any and all sorts of systematics – known and unknown.

Finally, radio confirmed quasars from FIRST and previously known quasars with $z > 2.15$ are also targeted. In the end, many objects are selected by more than one of those algorithms. BOSS science goals require ~ 20 quasars per square degree (at $z > 2.2$). In order to meet other quasar science goals, such as clustering and luminosity function measurements, a uniformly selected sample (called CORE) was defined, based on the XDQSO routine. A BONUS sample with additional information from near-infrared photometry from the UKIDSS Large Area Survey (Lawrence et al. 2007) is also available.

2.2.3 SDSS-IV

The SDSS-IV corresponds to the last generation of SDSS, and encapsulates the Extended Baryon Oscillation Spectroscopic Survey (eBOSS). The main (CORE) eBOSS quasar sample (Myers et al. 2015) targets quasars at $z > 0.9$ over the entire eBOSS footprint, involving a homogeneous selection with the likelihood-based routine XDQSOz, additional mid-infrared-optical color cuts (from WISE; Wright et al. 2010), and applying the same limiting magnitudes as BOSS (with the caveat of only considering targets fainter than $i > 17$). eBOSS does not retarget known (previously well-classified) quasars, though.

eBOSS also encompasses a specific program to select quasars based on variability (Palanque-Delabrouille et al. 2016) and using time-domain data from the Palomar Transient Factory (PTF; Rau et al. 2009). This sample is targeted to supplement Ly- α quasars in the CORE sample.

2.2.4 Superset

In addition to the quasar catalogs that accompany SDSS data releases, a superset of observations (Superset hereafter) has been created since DR12 to select quasar spectra

based on the targeting bits assigned for quasar programs. To ensure great completeness and purity, a subsample of spectra was visually inspected (since, e.g., in the case of the last data release, DR16Q, the sample is too large to be fully inspected), followed by the confirmation or correction of their classification and redshifts by the QuasarNet tool (Busca & Balland 2018).

The visual inspection and highly-specialized targeting algorithms guarantee a secure identification of the sources, as well as the inclusion of all quasar targets (including stars and galaxy spectra whose broad-band colors are consistent with those of quasars), which makes this catalog ideal to identify spectroscopically confirmed quasars, providing a census of the main contaminants to the quasar sample.

In this thesis we use the latest public release of the Superset (DR16Q-Superset)⁴, which contains 1,440,615 quasar targets. Additional cuts can be applied to the entries `Z_CONF` (where values < 3 ensure large confidence rating for the visually inspected redshift) and `CLASS_PERSON` (which corresponds to the object classification from visual inspection and is shown in Table 2.1).

Table 2.1: Visual inspection classifications from the Superset catalog.

Value	Object type	Value	Object type
0	not-inspected	1	star
3	quasar	4	galaxy
30	BAL-quasar	50	possible blazar

2.2.5 Final remarks on quasar selection

Because quasars are selected by their optical colors, regions of the sky in which the SDSS photometry is poor are unlikely to have complete quasar targeting. Inhomogeneities in the quasar target selection, particularly those related to extinction and depth of the imaging data used for targeting, are the main sources of systematics in the evaluation

⁴Superset catalog available at:
https://data.sdss.org/sas/dr16/eboss/qso/DR16Q/DR16Q_Superset_v3.fits

of the correlation function. As we will show later, J-PAS will be less affected by these pre-selection effects.

2.3 Quasar clustering

Hopkins et al. (2008), Cattaneo et al. (2009), Kormendy & Ho (2013), Dubois et al. (2013), Heckman & Best (2014), just to cite a few, have shown that the epoch of quasar activity is related to the star formation activity in elliptical galaxies. That, combined with the evidence that supermassive black holes are believed to inhabit the centers of massive galaxies (Kormendy & Richstone 1995, Miyoshi et al. 1995, Ferrarese & Merritt 2000, Gebhardt et al. 2000, Ghez et al. 2008, Genzel et al. 2010), suggests that black holes and galaxy spheroids share a common evolutionary physical process. Moreover, since galaxies reside in dark matter halos, a connection between black holes and their dark matter halos is naturally expected. Observationally, quasars are very luminous objects and are believed to inhabit the centers of very massive halos (with typical masses of $10^{12} - 10^{13} h^{-1} M_{\odot}$), which means that they have a great potential to probe structure formation in the Universe.

With the advent of the 2dF Quasi-Stellar Object Redshift Survey (2QZ; Croom et al. 2005) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), the first high precision measurements of the two-point correlation function with quasars were possible. However, measuring the small-scale quasar clustering in fiber-based spectroscopic surveys is usually hindered by fiber collisions. For instance, the SDSS fibers cannot be placed at less than $\sim 55''$ of each other on a single spectroscopic plate – this distance corresponds to a comoving separation of $1 h^{-1}$ Mpc at a redshift of 1.5 (Richardson et al. 2012).

This problem can be partly overcome by measuring the angular clustering of photometrically selected quasars (Myers et al. 2006; Myers et al. 2007; Ivashchenko et al. 2011; Ho et al. 2015). In this regard, Abramo et al. (2012) pointed out that the J-PAS survey could deliver nearly 240 type-I quasars per square degree up to a magnitude limit of $g \sim 23$, with a photometric redshift precision of $\sim 0.002(1+z)$, which could clearly have a large impact not only in the study of quasar clustering, but also in cosmology

applications as a whole.

In order to provide a detailed description of the clustering of quasars we employ Halo Occupation Distributions (HODs) to describe the types of dark matter halos that host quasars. HODs (see, e.g. [Seljak 2000](#); [Peacock & Smith 2000](#); [Berlind & Weinberg 2002](#); [Zheng et al. 2007](#)) constitute an empirical and intuitive approach to describe how luminous tracers occupy the dark matter halos. The HOD description consists of: (i) a one-halo term, relating the (small-scale) galaxy-dark matter statistics (via a probability $P(N|M_h)$ that a halo of a given mass M_h hosts N galaxies of a certain type); and (ii) a two-halo term, which characterizes how halos of a given mass M cluster around each other (giving, thus, insights about the large-scale clustering).

The conditional probability $P(N|M_h)$ is usually assumed to be a Poisson distribution with mean $\langle N(M_h) \rangle$, which can be modeled in various ways. In general, it is decomposed into central, $\langle N_{cen}(M_h) \rangle$, and satellite, $\langle N_{sat}(M_h) \rangle$, contributions:

$$\langle N(M_h) \rangle = \langle N_{cen}(M_h) \rangle + \langle N_{sat}(M_h) \rangle. \quad (2.17)$$

For quasars, the halo occupations of central and satellite quasars are often assumed to be independent, meaning that the occupation fraction of satellite quasars does not depend on whether there is a central quasar in the halo. In the literature there are two favored parametrizations for the quasar HOD ([Richardson et al. 2012](#); [Shen et al. 2013](#)). In the first parametrization, the mean occupation function of central quasars is given as a softened step function with characteristic mass scale M_{min} (corresponding to the scale at which, on average, half of the halos host one quasar) and transition width $\sigma_{\log M}$:

$$\langle N_{cen}(M_h) \rangle = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\log M_h - \log M_{min}}{\sigma_{\log M}} \right) \right]. \quad (2.18)$$

This form is similar to the galaxy HOD ([Zheng et al. 2005, 2007](#)), and it is loosely motivated by cosmological hydrodynamical simulations of AGNs ([Di Matteo et al. 2008](#); [Chatterjee et al. 2012](#)).

The satellite component, in turn, is parametrized as a rolling-off power-law:

$$\langle N_{sat}(M_h) \rangle = \exp \left(-\frac{M_{cut}}{M_h} \right) \left(\frac{M_h}{M_1} \right)^\alpha \quad (2.19)$$

where M_{cut} is the mass scale below which the satellite mean occupation decays exponentially, M_1 is the approximate scale at which halos host, at least, one satellite, and α is a power index.

The second parametrization admits six free parameters (as opposed to the five parameters of the previous HOD), where the satellite component assumes the same form as Eq. 2.19, but the central component assumes a log-normal form given by

$$\langle N_{cen}(M_h) \rangle = f_{cen} \exp \left[-\frac{(\log M_h - \log M_{cen})^2}{2\sigma_M^2} \right]. \quad (2.20)$$

Compared with the five-parameter model, this form reduces the number of central quasars in massive halos.

In Fig. 2.8 we illustrate the mean (total) occupation functions for quasars selected from the SDSS DR7 at an effective redshift of $\bar{z} = 0.5$. The best-fit parameters were derived in Shen et al. (2013) and are shown in Table A.1.

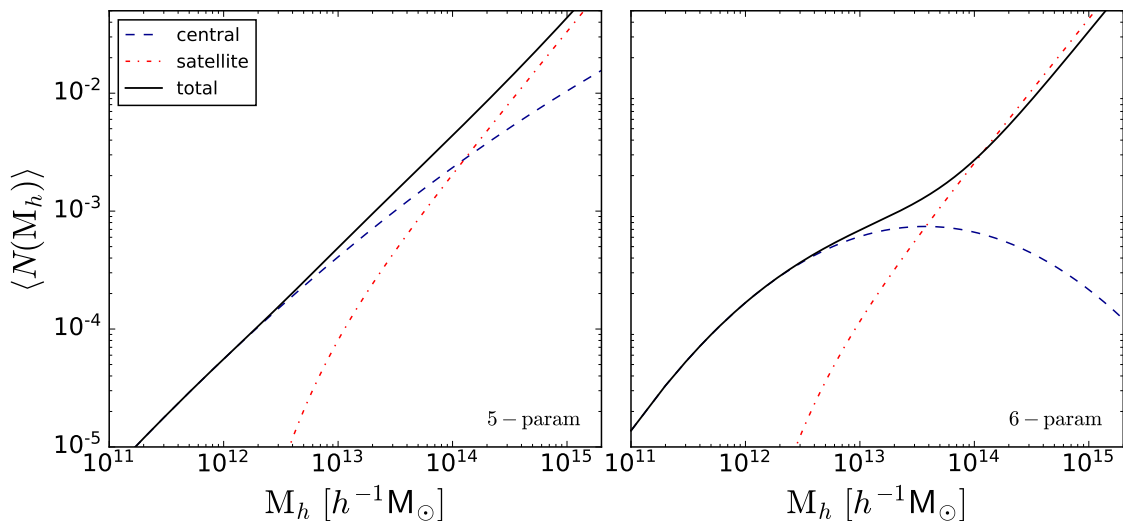


Figure 2.8: 5-parameter (on the left) and 6-parameter (on the right) mean occupation functions for SDSS DR7 quasars at $\bar{z} = 0.5$. The black solid line corresponds to the total occupation function, while the dashed blue line and dash-dotted red line correspond to the central and satellite contributions, respectively.

It is worth mentioning that the HOD approach assumes a dependency on the halo mass alone, not taking into account possible effects of assembly bias (i.e. the degree to which the formation history and physical properties of a halo or its large-scale environment affects its bias). In this regard, an alternative approach is the method of sub-halo abundance matching (SHAM; Conroy et al. 2006; Vale & Ostriker 2006; Hearin et al.

2015).

2.3.1 Quasar clustering in eBOSS

Recently, SDSS released the largest 3-D map of the Universe ever created. Here we present some of the results obtained by Hou et al. (2020) in their analysis of anisotropic clustering of quasars from eBOSS (DR16). The quasar sample spans the redshift range $0.8 < z < 2.2$ and has an effective redshift of $z_{\text{eff}} = 1.48$.

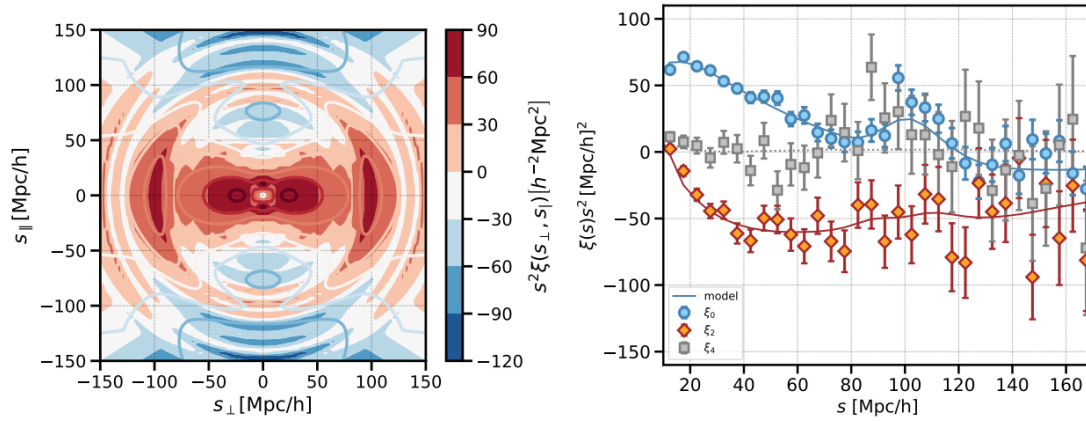


Figure 2.9: Left: 2D correlation function $\xi(s_{\parallel}, s_{\perp})$ measured from the DR16 quasar sample. The solid contour corresponds to the theory prediction. Right: The measured correlation function for monopole ($\ell = 0$, blue), quadrupole ($\ell = 2$, red) and hexadecapole ($\ell = 4$, gray). The solid lines show the best fitting full-shape model. Credit: Hou et al. (2020).

Figure 2.10 shows a corner plot with the posterior distribution of the Alcock-Paczynski (AP) parameters (Alcock & Paczynski 1979), the growth rate $f\sigma_8$, and the linear bias b_1 for the North Galactic Cap (NGC), South Galactic Cap (SGC), and the combination of both. The AP parameters are also known as the geometric distortion parameters, decomposed into perpendicular and parallel distances to the line-of-sight, and relating the BAO scale to the comoving sound horizon at the drag epoch (r_{drag} – which, in turn, depends on the ratio of baryon to radiation density). Their results yield $b_1 \sim 2.4$ at $z_{\text{eff}} = 1.48$.

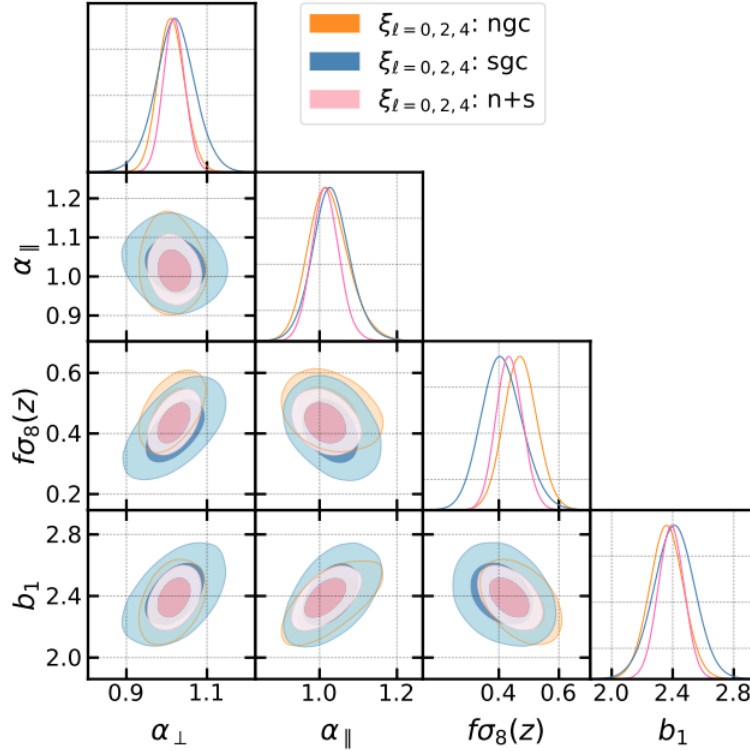


Figure 2.10: Posterior distribution for AP parameters, $f\sigma_8$ and linear bias b_1 for NGC (orange), SGC (blue), and combined (pink). Credit: [Hou et al. \(2020\)](#).

2.4 Multi-tracer technique

A critical gap between observations and science applications is the optimal extraction of information from the catalogs. In the early days, prior to the development of CCD devices, mapping the large-scale structure was done by counting galaxies on photographic emulsions. This means that one was limited by the number of galaxies that could be detected by these early telescopes and sub-optimal photon counting devices. In this regime, we say that the counts of galaxies measured at the detector output are dominated by Poisson “shot-noise”. Nowadays, telescopes have become much larger and the detection techniques have improved, so that we are much less affected by low counts of galaxies. Nevertheless, sampling increasing numbers of a single type of tracer can be very costly and eventually we are constrained by the total number of galaxies that exist, as well as the finite volume that can be probed in our past light cone (as illustrated in Fig. 2.11). As a result, the uncertainties in our measurements will have contributions coming from both shot-noise (typically dominating at small scales) and “cosmic variance”

(usually most relevant at large scales).

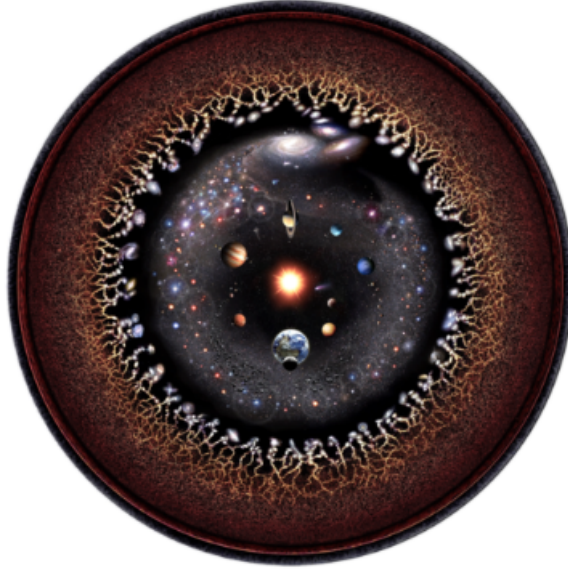


Figure 2.11: Observable Universe on a logarithmic representation. This illustration was created using Photoshop with real images from NASA, and features the solar system at the very center, surrounded by stars, nearby galaxies, the cosmic web, the CMB, and the Big Bang’s plasma on the edge. Astrophysical objects were enlarged for a better visualization. Credit: Pablo Carlos Budassi (2012).

Given this profile for the noise in galaxy clustering, we might naively infer that, in order to improve our constraints, it would be sufficient to simply increase the surveyed volume. However, up to any given redshift there is only a finite volume that we can observe, and hence a finite volume inside which one can estimate the amplitudes and phases of the (Gaussian) random modes of the density field.

So, at this point, we might ask ourselves: how to balance the shot-noise in light of cosmic variance, in such a way that we are able to recover the maximal amount of information from our catalogs? Or, in other words: how to estimate the power spectrum in a way that minimizes its covariance?

For a single type of tracer, [Feldman et al. 1994](#) (FKP) demonstrated that there is an optimal weight which minimizes the variance of the amplitude of the power spectrum averaged over some volume (bin) V_k in Fourier space. Assuming some fiducial models for the matter power spectrum $P_m = P(k; z)$, the average number of tracers in the catalog $\bar{n}(\hat{r}; z)$, and the bias of the tracer $b(z)$, FKP showed that the uncertainty in the power spectrum can be written as the Fisher information matrix for that Fourier bin:

$$\left[\frac{P(k, z)}{\sigma_P} \right]^2 = \frac{V_k}{2} \int d^3r \left[\frac{\bar{n}(\hat{r}, z) b^2(z) P(k, z)}{1 + \bar{n}(\hat{r}, z) b^2(z) P(k, z)} \right]^2. \quad (2.21)$$

This Fisher matrix was derived by Tegmark (1997), Tegmark et al. (1998), and it follows from the covariance of galaxy counts. The terms inside the integral are most easily expressed as:

$$\mathcal{F}_\alpha(k, z) = \frac{1}{2} \left(\frac{\mathcal{P}_\alpha}{1 + \mathcal{P}_\alpha} \right)^2, \quad (2.22)$$

where index α refers to a given tracer species, and

$$\mathcal{P}_\alpha = \bar{n}_\alpha(\hat{r}) [b_\alpha + f(z)\mu_k^2 + \Delta b_{\text{NG}}]^2 P(k, z) \quad (2.23)$$

is the effective redshift-space power spectrum of that tracer in units of its shot-noise, and we are considering that the bias is a general function of redshift, scale and angle, which may also include a bias correction due to primordial non-Gaussianities (NGs). We can also rewrite this result in a more compact way as $\mathcal{P}_\alpha = B_\alpha^2 P(k, z)$, where now B_α is an effective bias. In this notation, cross-correlations are represented by $\mathcal{P}_{\alpha\beta} = B_\alpha B_\beta P(k, z)$.

The result in Eq. 2.22 is telling us that there is an upper limit to the information that can be extracted from a single tracer on a finite volume. All is not lost, though. Instead of using all galaxies as a single tracer, one can gain additional information by mapping various tracers of the underlying density field at the same time. The gain from multi-tracer analyses then comes from probing the same volume more than once, each time with a different galaxy bias, reducing this way the cosmic variance for quantities that are related to the galaxy bias.

Percival, Verde & Peacock (PVP, Percival et al. 2004) proposed the first multi-tracer approach, by providing optimal weights for a minimum variance estimator of the matter power spectrum when combining several different biased tracers. However, their estimator did not provide a minimum-variance estimation of the redshift-space power spectra of the tracers.

Abramo et al. (2016) proposed another multi-tracer optimal estimator (MTPK hereafter), based on the Fisher matrix for multiple tracer species (Abramo et al. 2012;

Abramo & Leonard 2013):

$$\mathcal{F} [\log \mathcal{P}_\alpha, \log \mathcal{P}_\beta] = \frac{1}{4} \left[\frac{\mathcal{P}_\alpha \mathcal{P}_\beta (1 - \mathcal{P})}{(1 + \mathcal{P})^2} + \delta_{\alpha\beta} \frac{\mathcal{P}_\alpha \mathcal{P}}{1 + \mathcal{P}} \right] \quad (2.24)$$

where $\mathcal{P} = \sum \mathcal{P}_\alpha$ is the total effective clustering strength.

The MTPK method is a generalization of the standard weighting scheme of FKP and, when compared to PVP, it has the advantage of being optimal both in terms of the estimation of the combined power spectrum (cross-covariances were not fully developed in PVP) and of the redshift space auto-power spectrum of each individual species. In fact, by combining the power spectrum estimators of individual tracers, the results provided by MTPK are identical to those obtained with the PVP method.

The idea behind the multi-tracer technique is that cosmic variance is only inherited through the spectrum. Thus, by comparing the clustering of different tracers of large-scale structure, which can be different types of galaxies (with different types of luminosities), quasars, or even dark matter halos themselves – it is in principle possible to measure with arbitrary accuracy (i.e., without being constrained by cosmic variance) the physical parameters that determine the different clustering strengths \mathcal{P}_α (Eq. 2.23). This means that, for instance, by taking the ratio of two tracers, the Gaussian fluctuations in $P(k, z)$ are cancelled out and, thus, cosmic variance is beaten:

$$\frac{\mathcal{P}_1}{\mathcal{P}_2} = \frac{n_1 (b_1 + f\mu_k^2)^2}{n_2 (b_2 + f\mu_k^2)^2}. \quad (2.25)$$

The cancellation of cosmic variance can help us improve the constraints on some important cosmological parameters, such as the redshift-space distortion parameter β and the non-Gaussianity parameters f_{NL} and g_{NL} . Note, however, that other complications arise, coming from the covariance of the biases, non-linearities and modelling, assembly bias, and cross-covariances. This means that there is still a lot of room for improvements.

The current and next generation of multi-tracer surveys will be far superior to previous single-tracer surveys of large-scale structure. In particular, the J-PAS survey will cover an immense volume ($\sim 8500 \text{ deg}^2$ up to $z \sim 5$), providing superb datasets and census for different types of tracers (i.e., the volume sample variance will remain very small). Nevertheless, the expected density for quasars will peak around $10^{-4} h^3 \text{ Mpc}^{-3}$,

and since the amplitude of the power spectrum on $100 h^{-1}\text{Mpc}$ scales is around 10^3 (see Fig. 2.3), this implies that Poisson shot-noise will be an important nuisance. On the other hand, the quasar catalog will partly overlap with the J-PAS luminous red galaxy (LRG) and emission line galaxy (ELG) surveys, allowing us to employ different tracers of LSS to significantly improve the cosmological constraints whenever these species overlap.

In the next section we show some preliminary results obtained for an application of the MTPK method using a sample of ELGs and quasars.

2.4.1 Simulating a quasar and ELG photometric survey

When analysing photometric catalogues in redshift bins, the theoretical power spectrum predictions require precise estimates of the redshift distributions, which are in general compromised by large uncertainties in the photometric redshifts. For most surveys, this issue is even more critical for photometric quasars, since their redshift estimates are significantly more uncertain than for other types of galaxies, and include a significant fraction of catastrophic failures. As we will see in Chapter 6, this will not be case for quasars detected with the J-PAS photometric system, which will deliver very precise photo-zs for these objects.

So, we want to investigate how the redshift error distribution obtained on the basis of the J-PAS observations would affect our ability to extract information from the combined power spectrum of multiple tracers – in the presence of quasars. Here we show some preliminary results (Abramo, Queiroz et al. in prep. 2021) using the multi-tracer optimal estimator derived in Abramo et al. (2016) to analyse a set of 200 log-normal mocks containing only two types of tracers: emission line galaxies (ELGs) and quasars (QSOs), spanning the redshift range $1.4 < z < 2.6$ divided into three slices: $1.4 < z < 1.8$, $1.8 < z < 2.2$, and $2.2 < z < 2.6$. The motivation for this simulation is that the quasars would be detected by J-PAS, while the ELGs would be observed by surveys such as DESI (DESI Collaboration et al. 2016) or Euclid (Laureijs et al. 2011, Amendola et al. 2018).

The log-normal simulations produce a grid of positions for the tracers inside a box, and model the distribution of tracers as a log-normal random field:

$$1 + \delta_{LN}(\mathbf{r}) = e^{b\delta_G(\mathbf{r}) - b^2\sigma_G^2(\mathbf{r})/2} \quad (2.26)$$

where σ_G^2 is the variance of the Gaussian field inside a cell of the volume. For a detailed description on the generation of log-normal maps, see [Coles & Jones \(1991\)](#).

The simulation was implemented in the following way.

First, we imposed a mask corresponding to the range $126^\circ < \text{RA} < 160^\circ$, and $17^\circ < \text{dec} < 46^\circ$, with circular holes which simulate exclusion zones caused by objects such as bright stars. These holes were inserted randomly in the mask, but with a gradient towards the lower range of RAs, in order to emulate the increased number of bright stars as we point closer to the halo of the Milky Way. The final area, shown in [Fig. 2.12](#), spans approximately 3000 deg^2 – which is roughly what we expect to have in 3-4 years of operation of J-PAS (i.e., by 2024 or 2025).

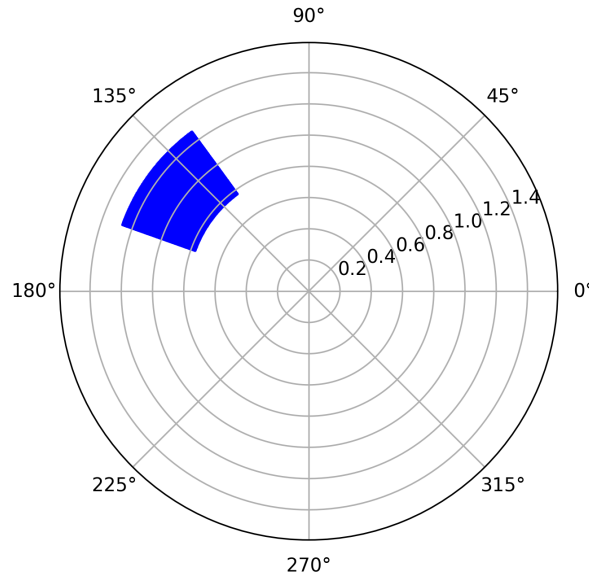


Figure 2.12: Polar projection of the simulated survey footprint (blue region). Here, RA is indicated as the polar angle in the polar plot, and dec (in radians) is denoted by the concentric rings.

Second, we employed radial selection functions that emulate the distributions of QSOs and ELGs in that redshift range. This QSO radial selection function was drawn from the QSO luminosity function (LF; [Palanque-Delabrouille et al. 2016](#)), with a completeness of 80% (meaning that we took the numbers of objects from the LF and multiplied them by 0.8). In the case of ELGs, we took their numbers from the average redshift distribution expected by the DESI and Euclid surveys. Those radial selection functions

are shown in Fig. 2.13.

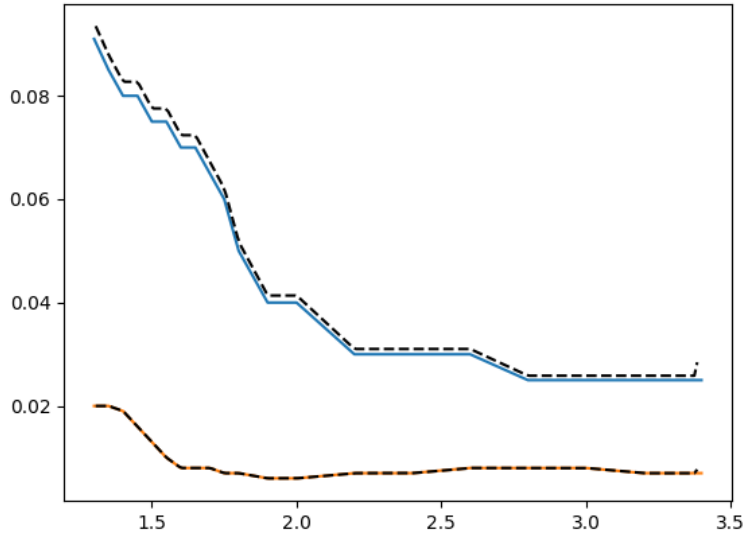


Figure 2.13: Radial selection functions, dN/dz , of quasars (red solid line) and ELGs (blue solid line) as a function of redshift. The dashed lines correspond to the normalized counts in our random maps.

We used the mask and the radial selection functions to produce a random catalog comprised of 10^9 points, which then determine how the objects are distributed inside our 3D volume. These randoms were rotated in such a way that the z -axis corresponds roughly to redshift, and then inserted in the 3D rectangular volumes corresponding to the three redshift slices defined above, and gridded into cells of side $10 h^{-1}\text{Mpc}$. These boxes have dimensions of $\{n_x, n_y, n_z\} = \{190, 172, 67\}$, $\{213, 193, 62\}$ and $\{232, 210, 58\}$ for the first, second, and third redshift slices, respectively. The filling factor of these boxes (the ratios of the cells with any number of objects divided by the total number of cells of the boxes) is approximately $\sim 50\text{-}60\%$. In Fig. 2.14 we show seven plots in $x - y$ corresponding to the slices $n_z = 5, 15, \dots, 65$ – all for the first box. One can see how the cone defined by the redshift range defines a volume that appears first as the corners of the lower Cartesian z slices, and then as a cap for the top Cartesian z slices. One can also see how the mask defines not only a range in RA and dec, but also how the exclusion zones increase as we reach the lower limit in RA (which appears as the upper part of the plots).

The actual simulations with large-scale structure were implemented in the following way. First, on each of the boxes defined above we created a log-normal mock of halos

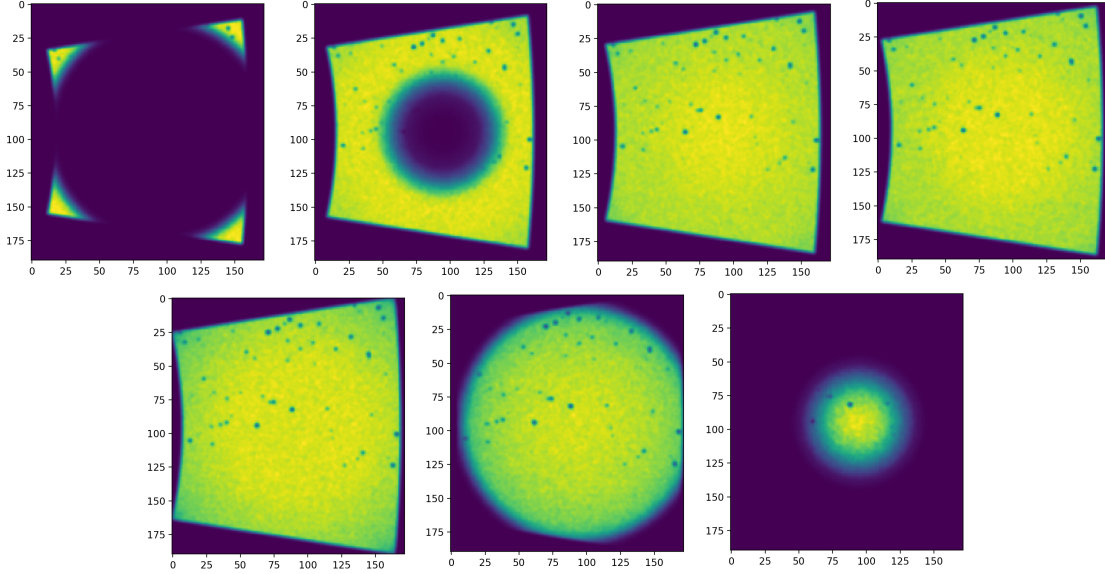


Figure 2.14: Projections in the $x - y$ plane of the mask and selection function of the first redshift slice, for seven slices of $n_z = 5, 15, \dots, 65$. The volume defined by the cone appears first as the corners of the lower Cartesian- z slice (upper left panel), and then as a cap for the top z slice (bottom right panel). Note that the holes, corresponding to the random exclusion zones, have a gradient towards lower values of RA (upper parts of y -axis).

with masses in the range $\log_{10}(M/h^{-1}M_{\odot}) = 11.5 - 14.5$. The underlying spatial matter distribution of these log-normal mocks follow a matter power spectrum using a Λ CDM cosmology with the same parameters as defined by Planck (Planck Collaboration et al. 2020a), with halo mass functions and halo biases determined by the fit of Tinker et al. (2010).

The redshift-space distortions were implemented using the Kaiser (linear) model (see Eq. 2.12). We should also remark that these log-normal mocks fill the entire boxes – the mask and selection functions for QSOs and ELGs were imposed at another stage.

Once the halo log-normal mocks in redshift space were completed, we “painted” the galaxies using a halo occupation distribution (Berlind & Weinberg 2002) integrated inside the six halo mass bins defined above. This means that, for each halo of a given mass, there is a certain probability of having a QSO or an ELG – and we use a Poisson statistics to randomly place those objects in the places of the halos according to their HODs. We have adjusted the HODs in such a way to obtain number densities similar to the radial selection functions above, as well as biases of $b_{\text{QSO}} = 2.3, 3.0, 3.5$ and $b_{\text{ELG}} = 1.3, 1.4, 1.6$ for the three redshift slices, respectively.

Finally, the nominal number of QSOs and ELGs in each box was restricted by the

angular mask and redshift dependence of those objects, resulting in 3D maps that reflect the spatial distribution of the underlying density field, at the same time that it preserves the footprint and radial selection functions of the putative survey.

However, the original simulated maps of QSOs and ELGs assume that the redshifts are perfect, when this is not always true for a photometric survey – even for a narrow-band survey such as J-PAS. In a photometric survey we have, among other issues, photometric redshifts which introduce errors which may be small or large (the so-called “outliers”). Hence, in order to simulate a photometric survey we take the original maps of QSOs and ELGs and “shuffle” the redshifts of those objects according to the realistic distributions (PDFs) of those photometric redshifts. For the quasars, we made a forecast for J-PAS using mock quasar catalogs and derived realistic errors with a model fit for quasar photo-spectra based on eigenspectra (see §4.4, 5, and 6 for more details). In the case of ELGs, we assume a Gaussian PDF with $\sigma_z = 0.003$. The final results are maps corresponding to catalogs obtained using photometric redshifts for those objects.

Finally, we compute the spectra of QSOs and ELGs of the original (“spectroscopic”) maps, as well as those of the shuffled (“photometric”) maps. In Fig. 2.15 we show the monopoles (upper panels) and quadrupoles (lower panels) of the QSO auto-spectrum (red), ELG auto-spectrum (blue), and QSO-ELG cross-spectrum (cyan), for the original (left panels) and for the shuffled maps (right panels), for the first redshift slice ($1.4 < z < 1.8$).

Quasars are more highly biased tracers of the matter density field than ELGs, and hence their spectra typically have more power. Note, however, that the clustering strength for the monopole lowers on small scales (large k) after the inclusion of photometric errors (right upper panel of Fig. 2.15). The signal-to-noise ratio for the QSO-ELG cross-spectrum has more power than the spectrum for ELGs alone, but also degrades on the small scales after the inclusion of photometric redshifts.

On the other hand, estimates for the quadrupole tend to be noisier and, thus, have naturally less power than the monopole (left bottom panel of Fig. 2.15). In this case, the inclusion of photometric errors significantly degrades the signal of the quadrupole – as we can see on the right bottom panel.

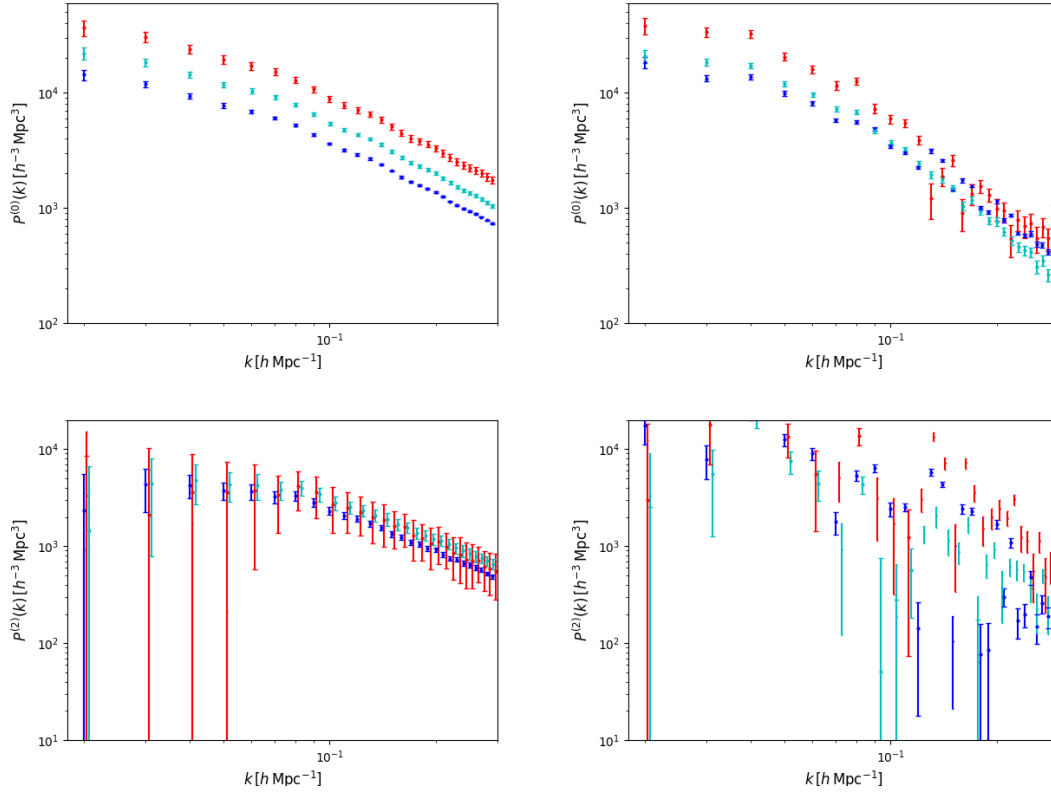


Figure 2.15: Upper panels: Comparison of the estimated monopoles for quasars (red dots) and ELGs (dark blue dots) for the original (left panel) and shuffle (right panel) maps on the first redshift slice ($1.4 < z < 1.8$). The cyan dots represent the cross-spectrum of quasars and ELGs. The symbols and error bars correspond, respectively, to the mean and variance of 200 realizations. Bottom panels: Same for the quadrupoles.

This trend is sustained in the second and third redshift slices, Fig. 2.16 and Fig. 2.17.

Note that this suppression effect was already expected from the case of simply having Gaussian photo- z errors (which is exactly the case for ELGs – dark blue dots in Fig. 2.15) and it can be explained as follows. The photo- z error distribution acts as a radial window function, in which the power spectrum $P(k)$ is damped by a Gaussian term of the order $\exp\left[-\frac{\sigma_z^2 k_{\parallel}^2}{2H^2(z)}\right]$. This means that in the presence of photo- z errors the window function suppresses modes along the line-of-sight with wavelengths $\lambda_{\parallel} < \sigma_z/H$ (e.g. Lima & Hu 2007; Chaves-Montero et al. 2018); equivalently, no significant suppression effects appear at large scales (i.e. small k). A similar effect is expected for neutrinos, for which the clustering power is suppressed on scales smaller than the free-streaming scale.

This is an ongoing analysis, but it already provides a foretaste of the potential that we will have to measure large-scale structure with QSOs at high redshifts, with photo-

metric redshifts that preserve, to a large extent, the clustering signal. The monopoles of the power spectrum are largely preserved, and there is still some signal in the quadrupole (even though it is largely washed out by the photometric redshifts). This means that we will be able to measure not only the BAO scale using these high-redshift maps, but also redshift-space distortions.

An overview of the quasar phenomenon is presented in the next chapter.

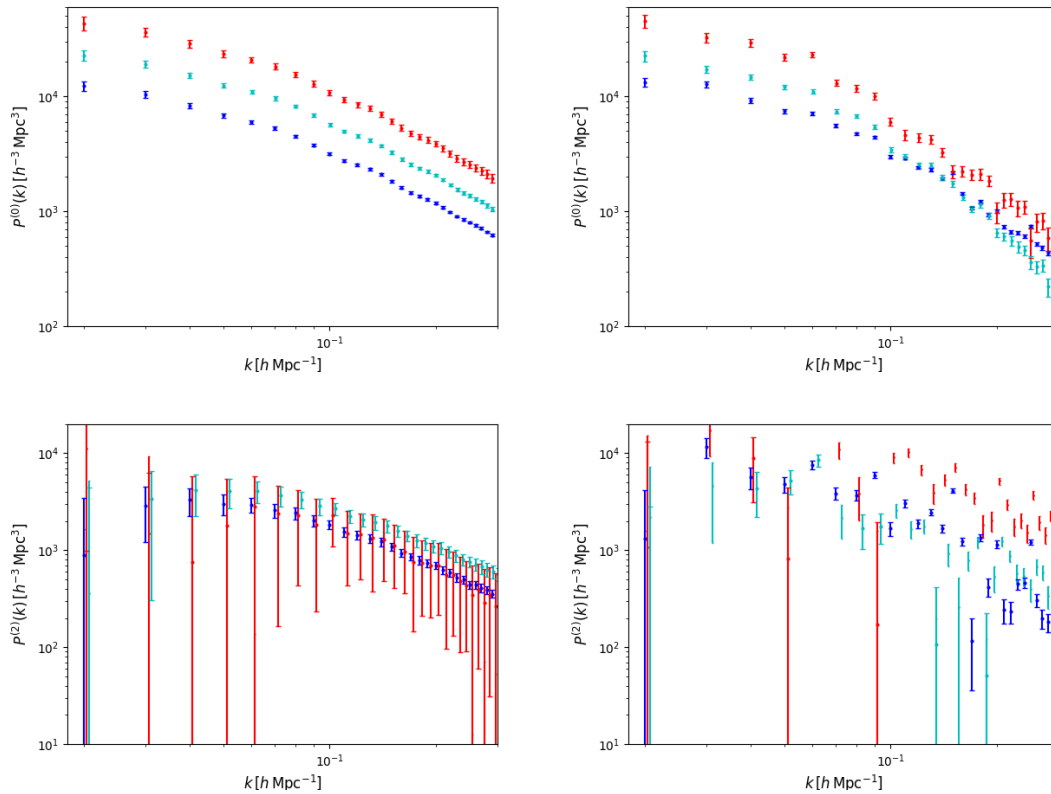


Figure 2.16: Monopoles (upper panels) and quadrupoles (lower panels) of the QSO auto-spectrum (red), ELG auto-spectrum (dark blue), and QSO-ELG cross-spectrum (cyan), for the original (left panels) and shuffled maps (right panels), for the second redshift slice ($1.8 < z < 2.2$). The symbols and error bars correspond, respectively, to the mean and variance of 200 realizations.

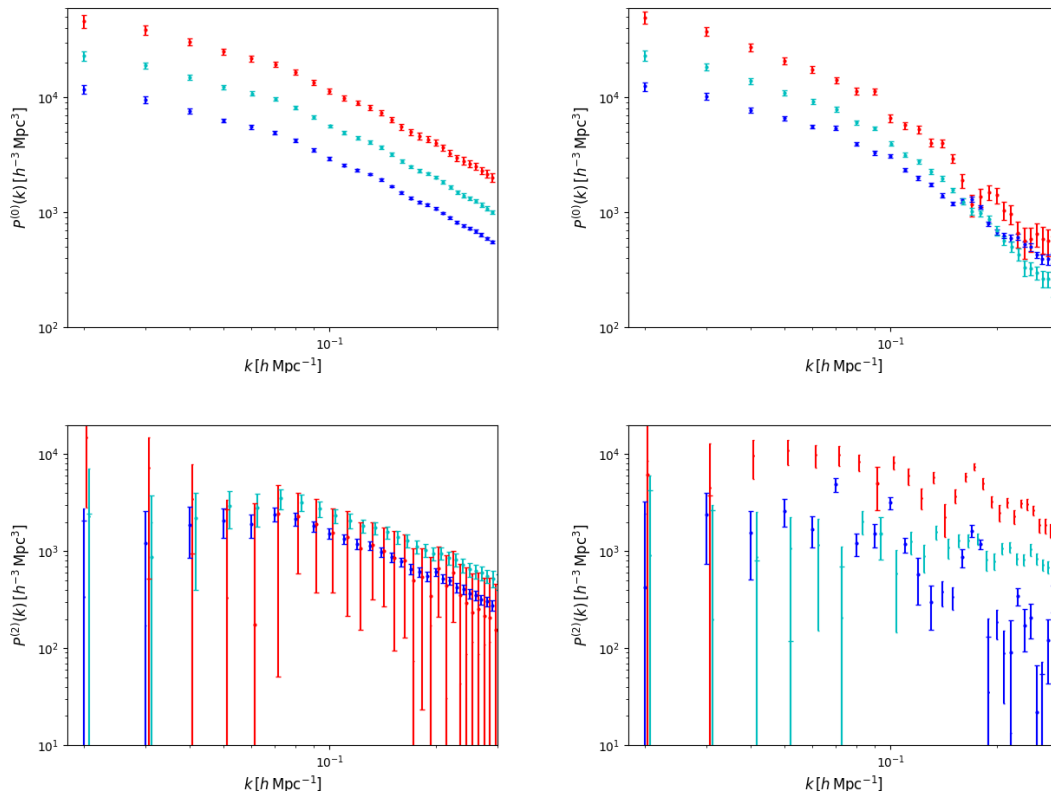


Figure 2.17: Monopoles (upper panels) and quadrupoles (lower panels) of the QSO auto-spectrum (red), ELG auto-spectrum (dark blue), and QSO-ELG cross-spectrum (cyan), for the original (left panels) and shuffled maps (right panels), for the third redshift slice ($2.2 < z < 2.6$). The symbols and error bars correspond, respectively, to the mean and variance of 200 realizations.

CHAPTER 3

“TWINKLE, TWINKLE, QUASI-STAR”

“Normal” galaxies come in different shapes and sizes; overall, the light that they emit can be attributed to the sum of the emission of individual stars in hydrodynamical equilibrium that compose them, with small contributions from gas and dust. However, there is a peculiar group of galaxies which are extremely luminous, presenting energy distributions that span the whole electromagnetic spectrum and intense emission lines in the optical. This energy release cannot be explained by thermal processes and seems to originate from a small central and compact region inside the galaxy. The flux emanating from the core suffers from variability on short timescales, and its brightness outshines its host. So, what is happening in these galaxies to produce such energetic outputs?

The central activity in these galaxies is ascribed to the presence of a supermassive accreting black hole ([Ferrarese & Merritt 2000](#); [Gebhardt et al. 2000](#); [Kormendy & Ho 2013](#)), and as the name suggests they form the class of active galactic nuclei (AGNs). As the material from the galaxy’s dense central region falls in toward the black hole, angular momentum will cause it to spiral in and settle into a flattened rotating disk. This disk, called accretion disk, heats up due to the gravitational and frictional forces at work, and glows. A corona of hot material forms above the accretion disc and can inverse-Compton scatter photons up to X-ray energies. A torus of gas and dust surrounding the accretion disk completes the picture (see [Fig. 3.1](#)). Moreover, in about 1% of the AGNs, the black hole and accretion disk produce narrow beams (“jets”) of energetic particles and eject them outward in opposite directions away from the disk.

Radio galaxies, Seyferts, quasars, LINERs, blazars: they are all active galaxies viewed by different angles. It is beyond the scope of this thesis to give a thorough analysis about black holes and each member of the AGN family. Our intention is to have a

better understanding about the main characteristics of the AGN phenomenon (in particular, focusing on quasars) and discuss their importance for galaxy evolution. For this, we divided this chapter in three parts. The first half is dedicated to present a brief overview of the central black hole responsible for the powering of the AGN activity. Then, in the second part, we present the main properties of AGNs and quasars. Finally, the third part is focused on the identification of quasars.

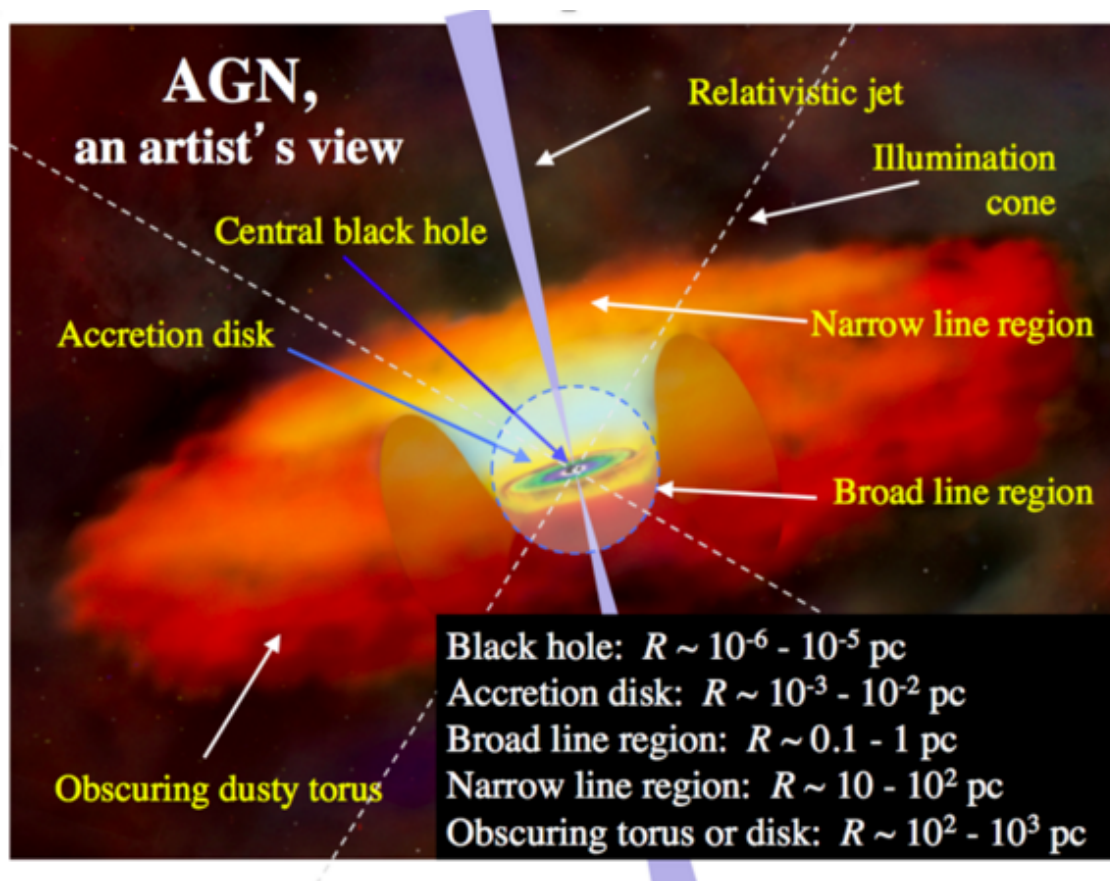


Figure 3.1: Illustration with the different features of an active galactic nucleus (AGN) and their corresponding scales. The extreme luminosity is powered by accretion onto a supermassive black, and it may present a jet. A pc is a parsec, equivalent to 3.26 light-years or 10^{16} m. Credit: A.C. Fabian, University of Cambridge.

3.1 Black holes as central engines

Quoting Luis Ho: “our confidence that supermassive black holes must power AGNs largely rests on the implausibility of alternative explanations”. In order to better understand this argument, let’s first take a look at some observational evidences on AGNs being powered by SMBHs¹.

¹The arguments presented here are based on the discussion by [Schneider \(2006\)](#).

1. Some radio galaxies present radio structures that extend to distances larger than 0.7 Mpc. This clearly imposes a minimum lifetime of $\tau \gtrsim 10^7$ yr (see Eq. 3.6), if we assume that the radio emission expands outwards at the speed of light.
2. Luminous quasars may reach luminosities of $L_{\text{bol}} \sim 10^{47}$ erg s⁻¹, imposing an energy release of $E_{\text{tot}} \sim 10^{61}$ erg during their minimum age.
3. The X-ray radiation emitted by some quasars is observed to vary in timescales of order of minutes, which means that, for the major part of the emitting region to be in causal contact, the spatial extent of the source (diameter) must be very small – of order of light-minutes ($\sim 10^{15}$ cm).

A “classical” method of producing energy is the nuclear fusion (e.g. taking place in stars). However, its efficiency ($\eta_{\text{max}} \lesssim 0.7\%$) is too low to be considered as the primary energy source in AGNs. Here is the reason²: in order to generate an energy of $E = 3 \times 10^{61}$ erg through the burning of hydrogen, a total mass of $M = E/\eta c^2 \sim 2 \times 10^9 M_{\odot}$ would be necessary, implying that all of the matter that was not converted into fuel (i.e. $(1 - \eta)m \simeq m$) must be present in the core of the AGN. However, the Schwarzschild radius for this mass is $r_S = 2GM/c^2 \sim 6 \times 10^{14}$ cm, i.e. same order as the estimated extent of the central source. Therefore, the assumption of energy production by thermonuclear processes seems inappropriate, because such a compact core could not emit any sort of radiation that it (eventually) produced.

In that case, “what darkness lurks in the hearts of galaxies?” (Djorgovski et al. 2008). The only alternative mechanism known for yielding large enough values of efficiency to produce such energetic outputs is mass accretion onto a compact object (e.g. Salpeter 1964; Zel’dovich 1964; Lynden-Bell 1969; Rees 1984), which refers to stellar remnants such as white dwarfs, neutron stars and black holes. Based on the previous arguments of compactness and mass, our best guess is that the central engine must be a black hole – and a very massive one.

Black holes (BHs) of all sizes are very common in the Universe: there are numerous known stellar-sized BHs (with masses $\sim 3 - 30 M_{\odot}$) in our Galactic neighborhood, as

²Another way of seeing this is that in order to produce this amount of energy the star would have to convert all of its hydrogen into helium in less than 10^7 yr (which corresponds to the black hole duty cycle). This implies that all of the stars would have masses $> 10 M_{\odot}$.

well as in several nearby galaxies. Since the discovery of the Seyfert galaxies in the 1940s and the first quasars in the 1960s, there has been increasing evidence for the existence of supermassive black holes (SMBHs, with masses of order $10^6 - 10^{10} M_{\odot}$) at the center of many – if not all, massive galaxies (M87: [Harms et al. 1994](#); [Kormendy & Richstone 1995](#); NGC4258: [Miyoshi et al. 1995](#); NGC3115: [Kormendy et al. 1996](#); [Emsellem et al. 1999](#); SgrA*: [Genzel et al. 1996](#), [Ghez et al. 1998](#), [Ghez et al. 2000](#), [Genzel et al. 2010](#); M31: [Statler et al. 1999](#), [Bacon et al. 2001](#); [Kormendy & Ho 2013](#) – just to cite a few). These observational evidences emerged from optical/infrared imaging and spectroscopy on the Hubble Space Telescope (HST) and large ground-based telescopes, as well as from very long baseline radio interferometry (VLBI).

The list of BHs is now long enough, and accretion onto SMBHs is well accepted as the energy source powering AGNs. From the theory of accretion onto black holes, assuming the infall of matter onto a non-rotating (Schwarzschild) black hole, potential energy can be converted into kinetic energy with a maximum efficiency of $\eta = \frac{GM_{BH}}{rc^2} = \frac{r_S}{2r} \lesssim 6\%$, for $r > r_S$. For rotating black holes, this efficiency can be as high as 42%.

3.1.1 A black hole in our backyard

Sagittarius A*, located at the center of the Milky Way, is the closest supermassive black hole to us. At a distance of about 8 kpc from Earth and with a mass of about $4 \times 10^6 M_{\odot}$, Sgr A* gained more notoriety in 2020 with half of the Physics Nobel Prize being awarded to Andrea Ghez and Reinhard Genzel due to their discoveries of this invisible and compact radio source through the analysis of the orbits of stars near the Galactic center (see Fig. 3.2), as well as measurements of the size and motion of the central source ([Genzel et al. 1996](#); [Ghez et al. 1998](#); [Ghez et al. 2000](#); [Genzel et al. 2010](#); [Ghez et al. 2008](#); [Gillessen et al. 2009a](#); [Gillessen et al. 2009b](#)).

Sgr A* is very bright in the radio and is highly obscured by interstellar dust. Although being dormant today, it must have experienced multiple gas-accretion episodes in the past to grow to its current mass. This picture is supported by a series of findings, such as the young stellar disk around Sgr A* – a remnant of a dense accretion disk ([Levin & Beloborodov 2003](#)), the Fermi bubbles ([Su et al. 2010](#)), and the bright emission of H α

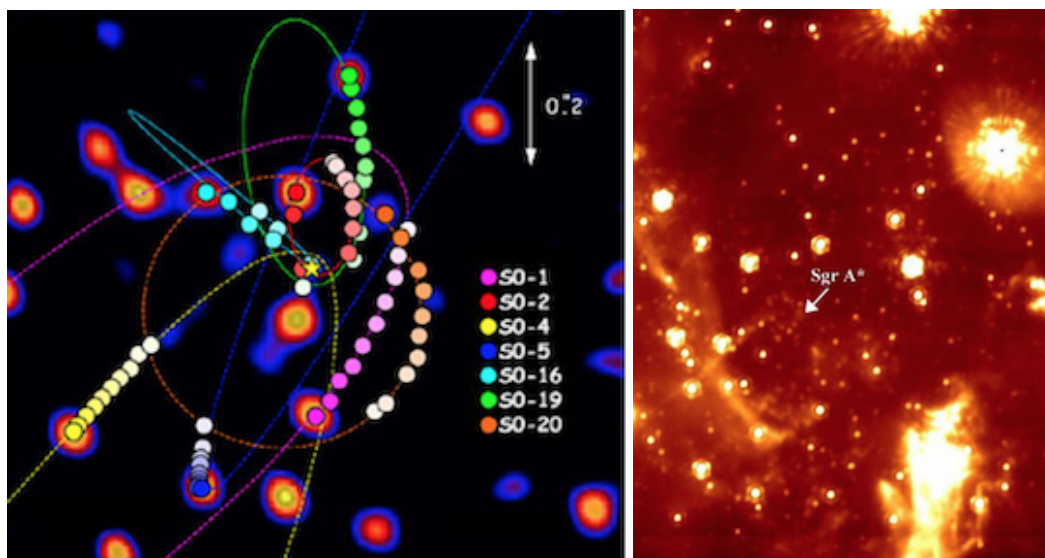


Figure 3.2: Left panel: Orbits of stars within the Milky Way’s central tenth of a light-year loop around the known location of Sagittarius A* (yellow star symbol). The fuzzy blobs are diffraction-limited star images in an infrared adaptive-optics frame taken by a 10-m Keck telescope in 2004. The frame is $1''$ (0.13 light-year) square. The annual average positions for these stars are plotted as colored dots (which have increasing color saturation with time). Also plotted are the best-fitting orbit solutions. Credit: UCLA Galactic Center Group/Keck Observatory. Right panel: Narrow-field image of the Galactic center, taken at an infrared wavelength of $3.8 \mu\text{m}$, resolving stars around the position of the central black hole. The frame is $10''$ tall, with a resolution of 82 mas . Credit: Alan MacRobert.

lines in the Magellanic Stream (Bland-Hawthorn et al. 2013).

3.1.2 First observation of gravitational waves

Another striking accomplishment was the detection of gravitational waves from the coalescence of two black holes (GW150914) by the twin detectors of the Laser Interferometer Gravitational-Wave Observatory (LIGO; Abbott et al. 2016). This transient gravitational wave signal detected by the LIGO and VIRGO Collaborations is hailed as a milestone, because it demonstrated the existence of binary stellar-mass black hole systems, demonstrating that such mergers could occur within the current age of the universe, and, more importantly, it corroborated the general relativity predictions of space-time distortions in the context of large-scale cosmic events.

In 2017, gravitational waves originated by a kilonova event (i.e. merger of two neutron stars) were detected (GW170817) and were accompanied by a short gamma-ray burst, which was monitored across the entire electromagnetic spectrum (e.g. Abbott et al. 2017a; Abbott et al. 2017b; Coulter et al. 2017; Díaz et al. 2017; Goldstein et al. 2017).

This marked the beginning of multi-messenger astronomy, and the Physics Nobel Prize of that year was awarded to Rainer Weiss, Barry C. Barish and Kip S. Thorne for their contributions to the observation of gravitational waves.

Since then, several events that released gravitational waves have been observed. Recently, in an updated catalog of gravitational wave transients, [Abbott et al. \(2020\)](#) have shown that the black hole merger rate appears to have peaked around eight billion years ago, following a period in which stars were forming – some of which were later turned into black holes, at a particularly high rate.

3.1.3 First image of a black hole

An ultimate proof for the existence of a supermassive black in the center of a galaxy came in 2019 with the historic image of the shadow of the black hole in the center of the elliptical galaxy Messier 87 (M87, shown in Fig. 3.3) captured by an international network of eight ground-based radio telescopes called the Event Horizon Telescope (EHT; Event Horizon Collaboration et al. [2019a](#), [2019b](#), [2019c](#), [2019d](#), [2019e](#), [2019f](#)).

Since a black hole is an object whose gravity is so strong that not even light can escape, it is impossible to capture direct images from it. Nevertheless, the hot accretion disk encircling the black hole shines bright, and against this bright disk, the black hole seems to cast a shadow. This dark shadow is attributed to gravitational light bending and photon capture at the region of non-return (a.k.a. event horizon). The EHT observed the central radio source (dubbed M87*) at a wavelength of 1.3 mm, and resolved an asymmetric bright emission ring with a diameter of $42 \pm 3 \mu\text{as}$. The asymmetric ring (which appears as a brightness excess at the bottom of right panel of Fig. 3.3) is produced by a combination of strong gravitational lensing and relativistic beaming of the emission from plasma rotating (in the clockwise direction, as seen by the observer) close to the speed of light around the black hole. By comparing their images to general-relativistic magnetohydrodynamic (GRMHD) simulations of black holes, a mass of $6.5 \times 10^9 M_{\odot}$ was retrieved.

Overall, the image obtained by the EHT collaboration is consistent with expectations for the shadow of a Kerr black hole as predicted by general relativity, and has opened

up a new window of observations for the study of supermassive black holes.

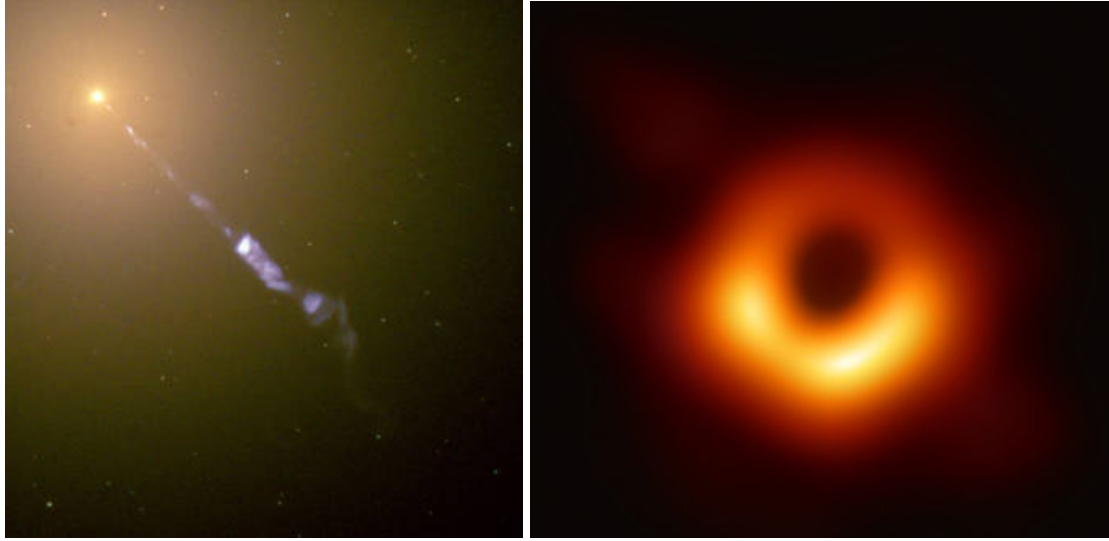


Figure 3.3: Left panel: Elliptical galaxy M87 depicting a jet. This Hubble image is a composite of individual observations in visible and infrared light. Credits: NASA and the Hubble Heritage Team (STScI/AURA). Right panel: Best-model image of M87* shadow. The asymmetric ring indicates that the bottom part of the emission region is moving towards the observer. Credit: EHT Collaboration.

3.1.4 Eddington luminosity

An important source of radiation in astrophysics results from release of binding gravitational energy due to the accretion onto massive objects. This process depends on the geometry of the system and can proceed via different paths.

The Eddington luminosity (L_{Edd} , named after the astronomer Sir Arthur Eddington) determines the maximum luminosity that a compact (spherical) object in hydrostatic equilibrium can produce. It determines, thus, the condition for accretion to occur: if the luminosity is higher than this limit, then the radiation pressure becomes larger than the gravitational force, and the object can no longer remain bound. Suppose, for example, that we have a central point source with mass M , total luminosity L , monochromatic luminosity L_ν , and a fully ionized gas blob at a distance r from the source. Then, the radiation force acting upon a gas particle is given by

$$f_{\text{rad}} = \frac{N_e \sigma_{\text{T}}}{4\pi r^2 c} \int_0^\infty d\nu L_\nu = \frac{N_e \sigma_{\text{T}}}{4\pi r^2 c} L \quad (3.1)$$

where N_e is the electron density, c the speed of light, and σ_{T} the Thomson scattering

cross-section (or, equivalently, the effective area of an electron when it is illuminated by radiation). The gravitational force per particle can be written as

$$f_g = \frac{GM\mu m_p N_e}{r^2} \quad (3.2)$$

where G is the gravitational constant, m_p the proton mass, and μ the mean molecular weight (i.e., mean number of protons and neutrons per electron). Then, the limiting requirement for accretion can be estimated assuming an equilibrium between the gravitational force on the gas and the radiation pressure force:

$$L_{\text{Edd}} = \frac{4\pi GMm_p c}{\sigma_T} \approx 1.3 \times 10^{38} \left(\frac{M}{M_\odot} \right) \text{ erg s}^{-1} \quad (3.3)$$

assuming pure hydrogen gas and an isotropic emission of radiation. Note that the Eddington limit does not depend on the distance from the compact object. Moreover, it defines the maximum luminosity allowed for objects being powered by a steady-state accretion flow, over a long period of time. Of course, the luminosity can exceed L_{Edd} for a short period (e.g., immediately after an outburst). Finally, for our assumption of a fully ionized plasma, the dominant source of opacity is Compton scattering. More realistic situations may involve partly neutral gas, though, leading to higher opacity and, consequently, a smaller effective L_{Edd} .

Expressing Eq. 3.3 in units of the luminosity of the Sun, we obtain

$$L_{\text{Edd}} \simeq 3 \times 10^4 \left(\frac{M}{M_\odot} \right) L_\odot. \quad (3.4)$$

Typical quasars ($M \sim 10^8 M_\odot$) have Eddington luminosities of order of $10^{46} \text{ erg s}^{-1} \simeq 3 \times 10^{12} L_\odot$.

3.1.5 Black hole accretion

If the conversion of infalling mass into energy takes place with an efficiency η , the accretion rate can be defined as $\dot{M} = L/\eta c^2 \equiv (L/L_{\text{Edd}})\dot{M}_{\text{Edd}}$. Setting the accretion luminosity equal to the Eddington limit gives us the maximum rate at which a black hole can accrete gas:

$$\dot{M}_{\text{Edd}} = \frac{L_{\text{Edd}}}{\eta c^2} \approx 2M_8 \left(\frac{\eta}{0.1} \right)^{-1} M_{\odot} \text{yr}^{-1} \quad (3.5)$$

where M_8 is the central mass in units of $10^8 M_{\odot}$ and the maximum mass-to-luminosity conversion efficiency is of order $\eta \sim 0.1$. Note that the black hole growth rate is $\dot{M}(1-\eta)$.

We can also estimate a characteristic time in which the mass of the SMBH will increase significantly:

$$t_{\text{evol}} = \frac{M}{\dot{M}} \simeq \eta \left(\frac{L}{L_{\text{Edd}}} \right)^{-1} 5 \times 10^8 \text{ yr}. \quad (3.6)$$

It is worth mentioning that, when considering accretion via a disk, the simplest possible case is a spherically symmetric accretion; this means that, in order to be accreted, the gas blob must fall into a “sphere of influence” called the Bondi radius (Bondi 1952), defined as the distance at which the escape velocity equals the sound speed c_s :

$$r_A = \frac{2GM}{c_s^2} \quad (3.7)$$

where $c_s = \sqrt{\gamma k_B T / \mu m_p}$, γ is the adiabatic index of the gas, k_B is the Boltzmann constant, and T is the temperature of the gas.

Following Netzer (2013), we can express a spherical accretion rate of hot gas, with constant temperature and no radiation pressure force, as

$$\dot{M}_{\text{Bondi}} = 4\pi \lambda c_s \rho r_A^2 \quad (3.8)$$

where λ is a correction factor of order 0.1 that depends on γ .

Obviously, the Bondi accretion model is very simple, because it assumes spherical symmetry (while realistic accretion flows have angular momentum), and a constant accretion rate (while in the presence of wind, the accretion rate decreases from the Bondi radius towards the black hole). Still, the assumption for the formation of accretion disks is that the viscosity in the disk is sufficient to provide the necessary mechanism to transfer outward the angular momentum of the gas and allow it to spiral into the center. On the way, the gas loses a considerable fraction of its gravitational energy, which can be efficiently ($\eta \sim 4 - 42\%$) converted into electromagnetic radiation (that peaks in the UV).

It can also be converted to kinetic energy, in which case the gas is blown away from the disk, or the gas can be heated to high temperatures, causing much of the energy to be advected into the BH. In addition, accretion disks are probably endowed with a magnetic field.

The accretion rate determines the accretion mode of the black hole. There are basically three different regimes: (*i*) when the accretion rate is lower than $0.01\dot{M}_{\text{Edd}}$, the black hole accretes gas by hot accretion flow, resulting in a geometrically thick disk (Yuan & Narayan 2014); (*ii*) when it is in the range $0.01\dot{M}_{\text{Edd}} \lesssim \dot{M} \lesssim \dot{M}_{\text{Edd}}$, the black hole accretes gas by cold accretion flow, yielding the standard thin disk (Shakura & Sunyaev 1973); (*iii*) when it is higher than the Eddington rate, the black hole accretes gas by slim disk (Abramowicz et al. 1988). Each one of these disks can be optically thin or thick, depending on the column density or the level of ionization of the gas. Regarding the accretion mode, it can be of two types (shown in Fig. 3.4; see Heckman & Best 2014 for more details), which imply in quite different spectra. In the radiative mode, also known as the accretion mode, the SMBH is surrounded by a geometrically-thin, optically-thick accretion disk through which a cold inflow occurs. As the name suggests, AGNs in this regime are radiatively efficient. The accretion disk is encircled by a hot corona, which Compton-upscatters photons from the disk into the X-ray regime. The ionizing radiation from the disk (and the corona) heats and photo-ionizes dense gas clouds located further away from the SMBH, producing UV, optical and near-IR permitted emission-lines, which have successfully explained many luminous AGNs. An obscuring structure (i.e. region of dusty molecular gas in a toroidal shape) involving the SMBH and accretion on larger scales completes this picture. Depending on the line-of-sight, we observe a type-I or II AGN, which might have a jet or not.

The second category is the radio (kinetic, or even jet) mode, associated with low accretion rates (i.e. radiatively inefficient) and moderate radio luminosities. The geometrically thin disk is absent or truncated in the inner regions, being replaced by a geometrically thick structure, with the formation of a hot accretion flow. These are the so-called advection-dominated or radiatively inefficient accretion flows (ADAFs/RIAFs; Narayan & Yi 1994; Ho 2008), which can launch the characteristic two-sided jets. The ADAF

solution was invoked to explain the observed features in low-luminosity AGNs, which contain very massive BHs that are not increasing significantly in mass. This is the case of Sgr A* in our Galactic center.

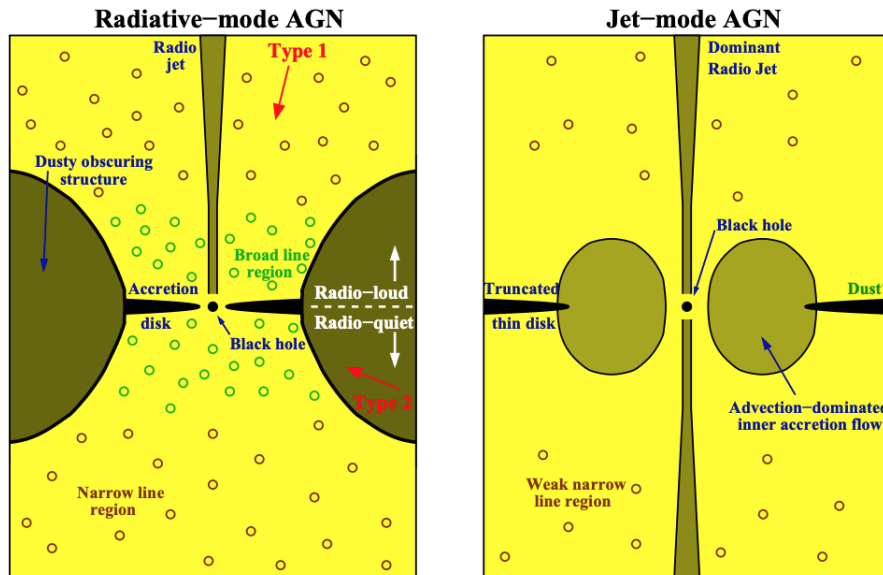


Figure 3.4: Illustration with the two types of accretion modes: radiative (left panel) and jet (right panel). Radiative-mode AGNs have a geometrically-thin, optically-thick accretion disk, surrounded by an obscuring structure of dusty molecular gas. Jet-mode AGNs have a geometrically-thick advection-dominated accretion flow, with the possibility of a transition to an outer (truncated) thin disk. Credit: Heckman & Best (2014).

Lastly, when the accretion rate exceeds the Eddington limit, the accreting gas becomes optically too thick to radiate all the dissipated energy. As a result, radiation is then trapped and advected inward. This model looks similar to ADAFs in the sense that “advection” plays an important role and the radiative efficiency is very low, but in the case of high accretion rates AGNs have a so-called slim disk. This model has been applied to narrow-line Seyfert galaxies and ultra-luminous X-ray sources.

In general, the activity in AGNs can be triggered by gas-rich mergers, recycled material from internal galactic processes and accretion of gas from intergalactic material (Harrison 2017).

3.1.6 A remarkable relation

The growing realization that all massive galaxies must host a central supermassive black hole was followed by a number of remarkable observations showing that black

hole masses³ are actually tightly correlated with some host-galaxy properties, despite a difference of several orders of magnitude in physical size scales (Kormendy & Ho 2013). The galaxy formation and the accretion onto the SMBH are closely related by two relations: the correlation between BH mass and the luminosity of the “bulge” part of the host galaxy; and the $M - \sigma$ relation, which represents a relationship between the cumulative past star-formation and black hole accretion histories of individual galaxies. These scaling relations are illustrated in Fig. 3.5 for dynamical black hole detections obtained for 37 galaxies observed with the Hubble Space Telescope (Kormendy & Gebhardt 2001). Despite the large scatter in the $M_{\bullet} - M_{B,\text{bulge}}$ on the left panel, the correlation is robust and yields a BH mass of $M_{\bullet} \propto L_{B,\text{bulge}}^{1.08}$. Since $M/L \propto L^{0.2}$, this implies that the BH mass is proportional to the bulge mass: $M_{\bullet} \propto M_{\text{bulge}}^{0.90}$. In contrast, the scatter in the $M_{\bullet} - \sigma_e$ correlation is small and implies $M_{\bullet} \propto \sigma_e^{3.65}$.

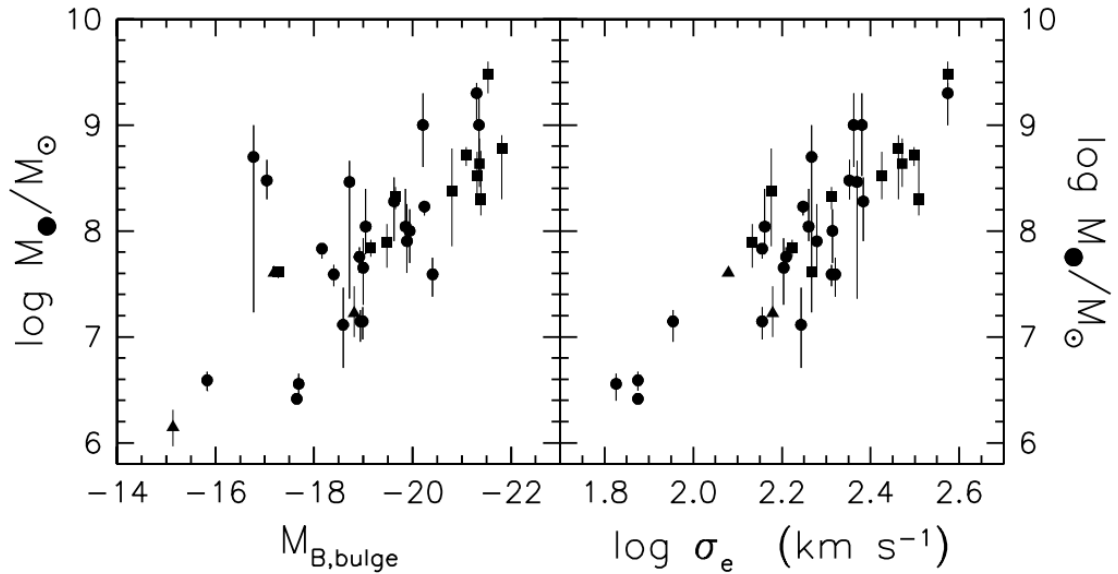


Figure 3.5: Correlation of the SMBH mass with the absolute magnitude of the bulge component of the host galaxy (left panel) and the luminosity-weighted mean velocity dispersion inside the effective radius of the bulge (right panel) for 37 galaxies observed with the Hubble Space Telescope. In both panels, filled circles indicate M_{\bullet} measurements based on stellar dynamics, squares are based on ionized gas dynamics, and triangles are based on maser disk dynamics. All three techniques are consistent with the same correlations. Credit: Kormendy & Gebhardt (2001).

The $M_{\bullet} - \sigma$ relation yields a tighter correlation, and can be parametrized as

³In general relativity, a black hole has only three properties: mass, angular momentum and electric charge. In astrophysical contexts, electric charge is generally unimportant, since it oscillates around zero, and measuring angular momentum of real black holes is extremely difficult – with only a few results obtained so far. Hence, the observed correlations involve only the mass of the SMBH.

$$\frac{M_{\bullet}}{10^8 M_{\odot}} \propto \left(\frac{\sigma}{200 \text{ km s}^{-1}} \right)^{\alpha}. \quad (3.9)$$

In fact, this relation does not depend on the morphology of the host galaxy (whether it is elliptical or spiral) – as shown in Fig. 3.6. It depends on the bulge type, though.

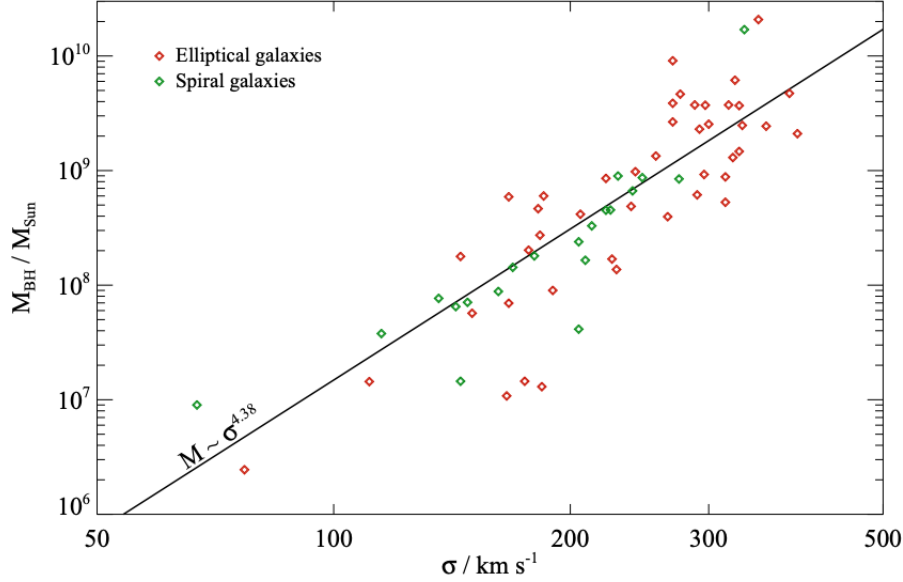


Figure 3.6: $M_{\bullet} - \sigma$ correlation for a sample of elliptical (red points) and spiral (green points) galaxies. The solid line indicates a best-fit of $M_{\bullet} \propto \sigma^{4.38}$. Data taken from [Kormendy & Ho \(2013\)](#). Credit: [Zubovas & King \(2019\)](#).

Of course, the best-fit measurements from Fig. 3.5 were obtained for a specific sample of galaxies; nevertheless, many subsequent observations have shown that these correlations are fundamental, suggesting that the histories of BH growth and star formation in the Universe are similar. In fact, the redshift evolution of the black hole accretion rate tracks the cosmic evolution of star formation rate (SFR), and both distributions seem to peak in the redshift range $1 \lesssim z \lesssim 2$, possibly due to a maximum merging rate between rich gas galaxies at this epoch. This is shown in Fig. 3.7.

These evidences indicate that the BH mass is determined in part by the amount of available fuel (i.e., cold gas supply), which, in turn, is connected with the total mass of the bulge (so that galaxies with large bulges have more massive black holes). Also, the AGN and the host galaxy share mutual mechanisms of self-regulation via feedback processes. The schematic diagram in Fig. 3.8 illustrates the self-regulatory feedback processes from black hole accretion and galaxy growth. Both processes are known to inject energy and

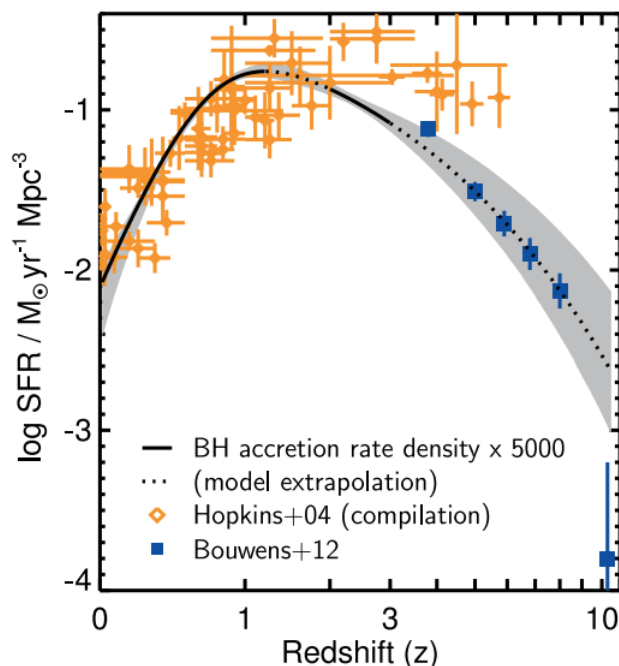


Figure 3.7: Redshift evolution of the black hole accretion rate (black line with gray band) scaled up by a factor of 5000 compared with the star formation rate distribution (orange dots). Credit: [Kormendy & Ho \(2013\)](#).

momentum (via radiation, winds and jets) that can reduce the availability of usable fuel through ionising, heating, shocking or expelling material.

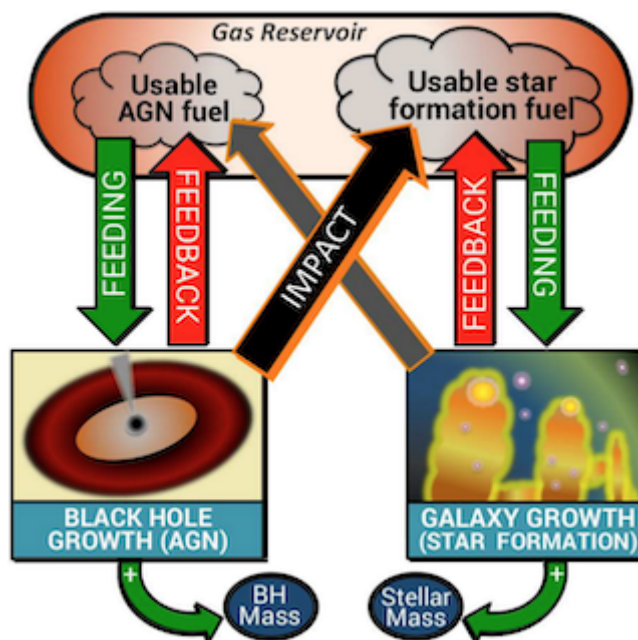


Figure 3.8: Schematic diagram of the relationship between fuel supply, galaxy growth and black hole growth. Both AGN and star formation are fueled by cold gas originating from a shared gas reservoir inside the galaxy halo. Not only the amount of gas, but also its ability to cool, are determinant in the availability of usable fuel for feeding black hole growth and star formation. In the case of providing the fuel for black hole accretion the material has the additional challenge of losing sufficient angular momentum to reach the inner sub-parsec region of the galaxy. Credit: [Harrison \(2017\)](#).

Detecting high-redshift quasars will give us insights not only about their clustering properties and black hole accretion rates, but also about the formation of their hosts across cosmic time. Therefore, studying the properties of the host galaxy, such as stellar mass, star formation rate and dust content, as well as measuring the number-density and rate of occurrence of AGNs, is very important for understanding the coevolution between galaxy growth and black hole growth.

As a final remark, observational results have shown that regarding the morphology AGN hosts are early-type or early-spirals, and are located in the transition region between the blue-cloud of star-forming galaxies and the red sequence of passive galaxies, forming the so-called “green valley” (Kauffmann et al. 2003; Salim 2014). Hence, the AGN activity is believed to be a short-lived, recurrent process in galaxies. In fact, Mallmann et al. (2018) suggest that the most luminous AGNs have been triggered by a recent supply of gas that has also triggered recent star formation in the central region, thus rejuvenating the stellar content of the nuclear region of the AGN hosts. Also, the negative feedback produced by the AGN has been proposed as a mechanism for fostering the transition from the blue cloud to the red sequence.

3.1.7 Black hole seeds

So far, we have presented an overview about why the basic energy production in AGNs must be related to a central supermassive black hole, and have shown some evidences to the existence of SMBHs in the center of massive galaxies and to the existing coevolution between black hole accretion and the host galaxy growth. One last fact worth mentioning regards the formation of these SMBHs.

Quasars are observed up to $z \sim 7^4$, i.e., at epochs when the Universe was only ~ 700 Myr old, and seem to contain SMBHs with masses of the order of $10^9 M_{\odot}$ (see Fig. 3.9 and Table A.2). This constitutes a significant challenge to the standard model of black hole growth: how is it possible to assemble these high masses at such early epochs? And what were the masses of the initial black hole seeds that allowed them to grow so rapidly to become the observed high- z SMBHs?

⁴To date, the most distant quasar was detected at $z = 7.54$ (Bañados et al. 2018).

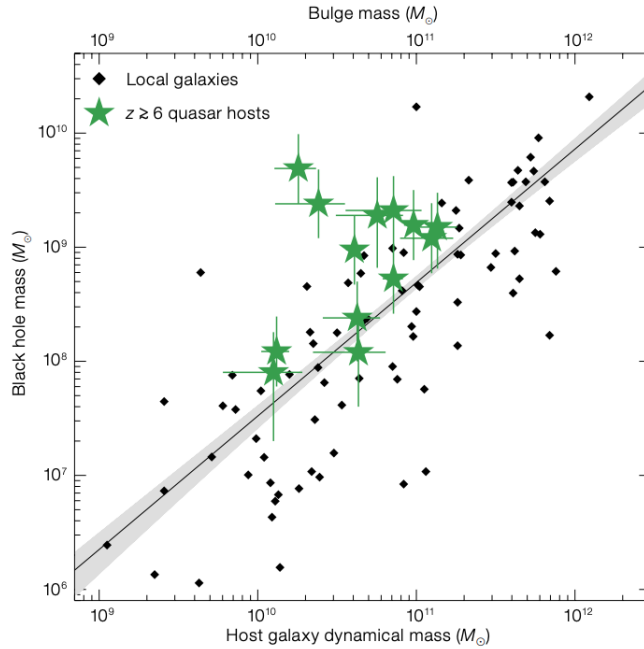


Figure 3.9: Comparison of black hole mass and dynamical mass estimates of $z \geq 6$ quasar host galaxies (green filled stars) and the bulge masses of local galaxies (black diamonds). The solid line and grey area show the local black hole to bulge mass relation as derived by Kormendy & Ho (2013). Credit: Venemans (2017).

Assembling such large black hole masses, this early in the history of the Universe, requires the convergence of several optimal conditions for the environment of the initial black hole seeds, including a vast gas reservoir and feedback mechanisms that do not interrupt the gas supply to the active nucleus. Currently, two distinct models are favoured: light and heavy seeds.

Light seeds are hypothesized as the remnants of massive Population III (Pop III) stars and, thus, to initialize with low masses (of order $100 M_{\odot}$). These seeds could, then, grow up to $10^5 M_{\odot}$ via accretion and mergers. Although easier to be formed, they may struggle in growing up to supermassive scales ($\gtrsim 10^6 M_{\odot}$) at $z \sim 6$, because they are expected to accrete with a lower Eddington limit and fragment more easily, moving their initial masses downward (e.g. Stacy et al. 2016; Pacucci et al. 2017).

Heavy seeds, on the other hand, would originate from the direct collapse of primordial gas clouds (Begelman et al. 2006). In this scenario, there is no growing phase from a low-mass black hole prior to the collapse: a supermassive primordial star (SMS) forms during catastrophic baryon collapse in atomically-cooling halos at $z \sim 15 - 20$, originating seeds with $\sim 10^4 M_{\odot}$. These SMSs must grow either (i) in the presence of Lyman-Werner UV sources that dissociate molecular hydrogen (hence, inhibiting star for-

mation); or (ii) in highly supersonic baryon streaming motions that delay the collapse of the halo – even in the presence of molecular hydrogen. Furthermore, the halo must also avoid significant metal pollution, in order to avoid fragmentation, and photoevaporation from ionizing radiation. Another possibility suggested by [Valiante et al. \(2017\)](#) is to have a Pop III star ionizing (or heating) the gas in the surroundings of the SMS, which allows a direct collapse into a $10^5 M_{\odot}$ seed.

Although rarer, these direct collapsing black holes (DCBHs) may be more successful than their lighter counterparts in explaining the origin of the masses of high redshift SMBHs. Moreover, they could emit gravitational waves during their tidal disrupt events, which could be detected today in the NIR. This means that they could be probed by the upcoming Laser Interferometer Space Antenna (LISA; [Amaro-Seoane et al. 2017](#)) and James Webb Space Telescope (JWST; [Gardner et al. 2006](#)), as well as by the next generation of extremely large telescopes (ELTs).

3.2 The unified model of AGNs

The first discoveries on the field of AGNs date back to 1943, when Carl Seyfert published a pioneering paper describing observations of nearby galaxies having bright nuclei that were sources of unusually broad emission lines. However, the unusual features of these spectra remained neglected for a long time. The development of radio astronomy acted as a catalyst to the understanding of AGNs, with the 3C radio survey playing an important role in the discovery of new radio sources and their optical counterpart positions. In photographic images, some of these objects were nearly point-like or quasi-stellar in appearance, being, thus, classified as quasi-stellar radio sources (which were, later, abbreviated as “quasars”). Nevertheless, a major breakthrough would only come in 1963 with the measurement of a redshift of 0.158 for the quasar 3C273 by [Maarten Schmidt](#). From then on, these objects were studied in greater details following the growing interest in their spectra and redshifts.

The realization that the extreme non-thermal (i.e., non-stellar) luminosity emanating from these sources was coming from a compact region at the center of the galaxy originated the nomenclature active galactic nuclei. As we have seen in §3.1, these AGNs

are powered by a supermassive black hole with $M_{\bullet} \sim 10^6 - 10^{10} M_{\odot}$. However, the study of the spectra of Seyfert galaxies (i.e., nearby elliptical galaxies with bright nuclei), identified two distinct groups: Seyfert 1 galaxies, which show strong, very broad ($2 \times 10^3 - 10^4 \text{ km s}^{-1}$) permitted and semi-forbidden emission lines; and Seyfert 2 galaxies, whose broadest lines do not exceed $\sim 1200 \text{ km s}^{-1}$. Such differences are now interpreted as arising from different viewing angles to the centers of these sources, and from a large amount of obscuration along the line-of-sight (Netzer 2013).

As they were discovered, the various AGNs received different nomenclatures, according to their appearances and emitted radiation. In this regard, the old unification scheme (Antonucci 1993; Urry & Padovani 1995) was a courageous attempt to construct a general picture connecting the various subgroups. The unified model describes the classification and main properties of this family of objects, which can be determined by (i) the black hole's mass, (ii) the rate of accretion onto the black hole, (iii) whether or not it has a powerful jet, and (iv) the angle at which we view the center of the galaxy. In Fig. 3.10 we show the main components of an AGN and the different nomenclatures depending on the inclination of the central source.

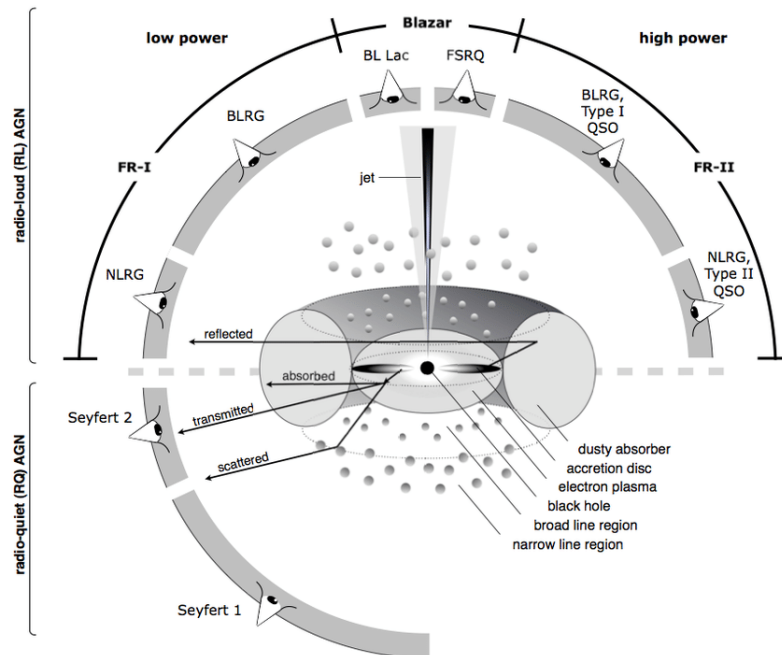


Figure 3.10: Unified model of AGNs. According to this scheme, an AGN is composed by a central accreting supermassive black hole, surrounded by an obscuring structure in a toroidal shape and possibly a jet. Depending on the orientation with respect to the line-of-sight, the object receives a different nomenclature. Credit: Beckmann & Shrader (2012).

One of the defining characteristics of AGNs is a broad spectral energy distribution (SED) that can be distinguished in several wavelength bands. It is customary to describe the SED in terms of the monochromatic luminosity per unit frequency ($[L_\nu] = \text{erg s}^{-1} \text{Hz}^{-1}$) or per unit wavelength ($[L_\lambda] = \text{erg s}^{-1} \text{\AA}^{-1}$), or yet in terms of their equivalent monochromatic fluxes (F_ν or F_λ) which receive an additional unit of cm^{-2} . In Fig. 3.11 we show a comparison of the SED of a normal galaxy (left panel, bottom row) to the SEDs of 7 different types of AGNs in the optical.

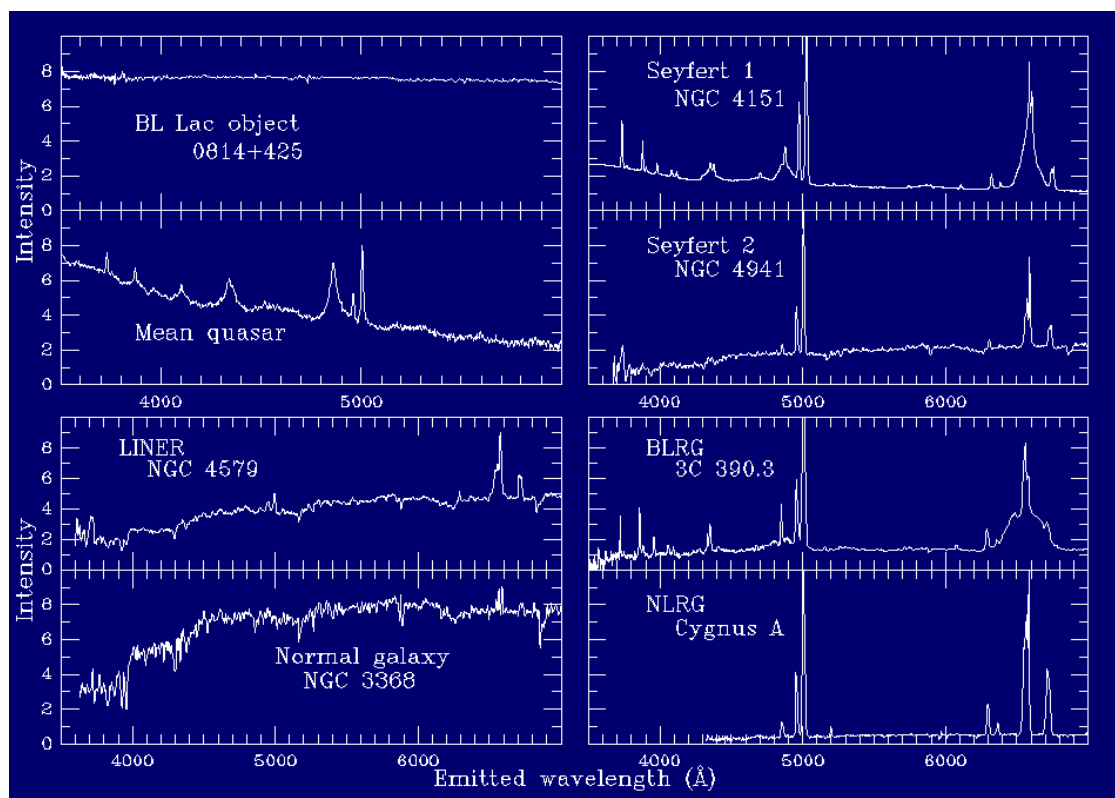


Figure 3.11: Comparison of the optical spectra of 7 AGNs and a normal galaxy (early-type spiral, NGC3368). BL Lacertae objects (left panel, top row) have featureless spectra, making their redshifts difficult to measure unless the surrounding galaxy can be detected. Seyfert 2 galaxies show much narrower lines than Seyfert 1s. Quasars, represented here by a composite produced from many individual objects (left panel, second row), have a family resemblance to Seyfert 1 nuclei, and in most cases, the bumps of Fe II emission are even more prominent in quasars. LINERs (left panel, third row) represent a lower-luminosity version of the processes seen in more traditional active nuclei and their galactic component is more evident. The BLRG and NLRG are broad and narrow-line radio galaxies, respectively. Although very similar to Seyfert galaxies, their spectra contain more profile structures. Credit: Bill Keel (2002).

In the following we describe some characteristics of AGNs according to their spectral signatures.

The nucleus of the galaxy is usually deeply embedded in dust. When heated by the primary AGN radiation source, these dust grains emit in near infrared (NIR) and

mid-infrared (MIR). Although AGN obscuration is more commonly associated with early stages of evolution, it can happen at any stage during the lifetime of the AGN (especially in systems that are actively interacting or merging). The standard unified model of AGNs distinguishes two phases in the material accumulated in the central region: (i) one portion is constituted by ionized gas at high speeds (ejected in the outflows); and (ii) the other by molecules at lower speeds, that can feed up the nucleus.

The deep gravitational potential of the SMBH leads to an accretion disk that radiates most strongly across the UV/optical spectrum, peaking in the near UV. This radiation originates a characteristic broad UV/optical feature known as the “big blue bump”. Furthermore, the intense magnetic fields expected to be present in the disk could provide a mechanism for jet collimation.

The central engine and accretion disk are surrounded by an extended, dusty, molecular toroidal region, involving the broad line region (BLR). The torus can be found at $\sim 1 - 20$ pc and dominates the spectral energy distribution out to $20 \mu\text{m}$. It may be composed by a smooth dusty distribution, high-density clumps (Nenkova et al. 2008a, 2008b), or even a two-phase medium combining both components. Such clumpiness suggests that the separation between type-I and type-II objects could be explained as the transition of an obscuring cloud along the line-of-sight (e.g., Audibert et al. 2017).

At viewing angles closer to the accretion disk, the obscuring torus hinders emission lines. In this case, there are strong radio emissions (sometimes, with prominent radio lobes that extend up to hundreds of kpc or even Mpc from the nucleus): these are the radio galaxies.

Broad emission lines from atoms in excited states originate close to the torus in clouds orbiting above the disk at high velocity, and are due to the Doppler effect. The emission lines can be of two types: permitted lines (i.e., those having high probability of occurring) and forbidden lines (or “improbable” lines that appear only in low density environments). These hot, high velocity, dense gas clouds near the black hole form the BLR. When the BLR is visible, a regular (type-I) quasar or Seyfert 1 galaxy is observed. On the left panel of Fig. 3.12 we show an example of SED fitting of luminous, type-I quasars, considering a standard torus model and two additional components: one contain-

ing pure graphite grains (hot dust) and the other a dusty NLR (warm dust) – Schweitzer et al. 2008 (for more details, see Netzer 2015). The right panel illustrates the large spectral differences between a type-I quasar and a radio galaxy.

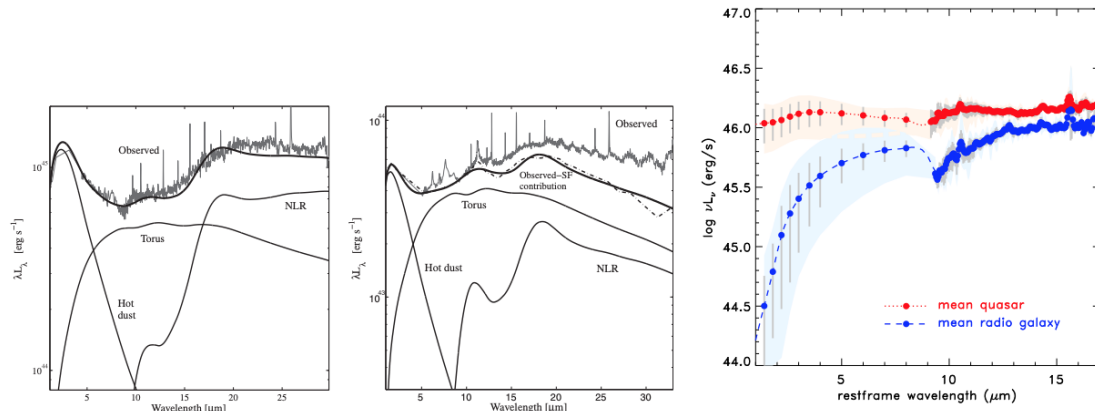


Figure 3.12: Left panel: Observed (gray) and star-forming contribution from the host subtracted (dot-dashed line) SEDs of two type-I AGNs. The diagrams show the three components used in the fit: hot pure graphite dust, a clumpy torus model with ISM-type dust, and dust emission from the NLR. The thick black curve is the sum of the three. Right panel: Composite type-I quasar (red) and radio galaxy (blue) SEDs at $z \sim 1.2$. Credit: Netzer (2015).

Another important component is the narrow-line region (NLR), which is further from the torus and is composed by cool, low-velocity, low-density gas clouds. When the broad-line region is hidden by the torus, but the NLR is still visible, the corresponding class is a type-II quasar or a Seyfert 2 galaxy.

The observation of broad lines in the polarized flux of the Seyfert 2 galaxies (e.g., NGC1068) gives evidence to support the unified model. It indicates that this type of galaxy must have a BLR that can only be observed with polarized light, corroborating to the presence of some obscuring structure along the line-of-sight that hides the inner central region.

In some cases, there are two-sided jets of relativistic particles⁵ emanating perpendicular to the plane of accretion. The X-ray emission is attributed to synchrotron radiation or inverse Compton scattering of high-energy density photons off relativistic electrons from the jet. When the jet points directly to the observer, we have a blazar, which can be of two types: flat-spectrum radio quasar (FSRQ), with prominent emission lines in the optical spectrum; or BL Lac Object, with weak emission lines or featureless continuum in the optical spectrum. Blazars are the only objects emitting in TeV.

⁵Due to the effect of relativistic beaming, sometimes only a single jet is visible.

Cold quasars (CQs; [Kirkpatrick et al. 2020](#)) constitute a rare subpopulation observed to host unobscured, X-ray luminous AGNs at the same time that it retains a cold gas supply fueling high star formation rates. This indicates the co-existence of an AGN with a starburst phase.

Another type of AGN are low-ionization nuclear emission line region galaxies (LINERs; [Heckman 1980](#)), which are characterized by the relative strength of emission lines from species of low ionization state, and whose $H\alpha$ luminosity is more similar to that of a typical giant HII region than to Seyfert galaxies and quasars.

BPT emission line activity diagnostic diagrams (named after Baldwin, Phillips and Terlevich [1981](#)) are useful for separating LINERs from normal star-forming galaxies (i.e., HII regions) and type-II AGNs on the basis of their $[\text{OIII}] \lambda 5007/H\beta$, $[\text{NII}] \lambda 6583/H\alpha$, and $[\text{SII}] \lambda 6716, 6731/H\alpha$ flux ratios. These diagrams use strong optical lines of close proximity in the ratios (limiting, hence, reddening and spectrophotometric effect) to distinguish between different classes of ionization. Type-I classification is more problematic because of the broad $H\alpha$ and $H\beta$ lines. The advent of SDSS and 2dF enabled us to fill these diagrams with details not seen before, providing a better understanding of the distribution and extent of emission-line galaxies. [Figure 3.13](#) shows one such example that has been used to separate star-forming galaxies from AGNs, as well as to distinguish between low-ionization (LINERs) and high-ionization (type-II) AGNs.

AGNs have been found to be variable over the entire electromagnetic spectrum. Indeed, variability was one of the first recognized properties of quasars, where significant variations ($\gtrsim 0.1$ mag) in the optical brightness of quasars occur on time scales as short as days, and blazars are characterized by extreme variability at all wavelengths. At a fundamental level, the physical origin of variations is not known yet; although it is believed that accretion-disk instabilities are involved on variations in short timescales, whereas variations on longer timescales are dominated by the fueling of gas into the nuclear regions plus regulation through feedback processes.

Changing-look quasars (CLQs; [Green et al. 2017](#)) are a newly-discovered class of AGNs that undergo rapid (< 10 yr) transitions between type 1 and type 1.9/2, with change in their continuum emission. They represent a challenge to the historical classifications

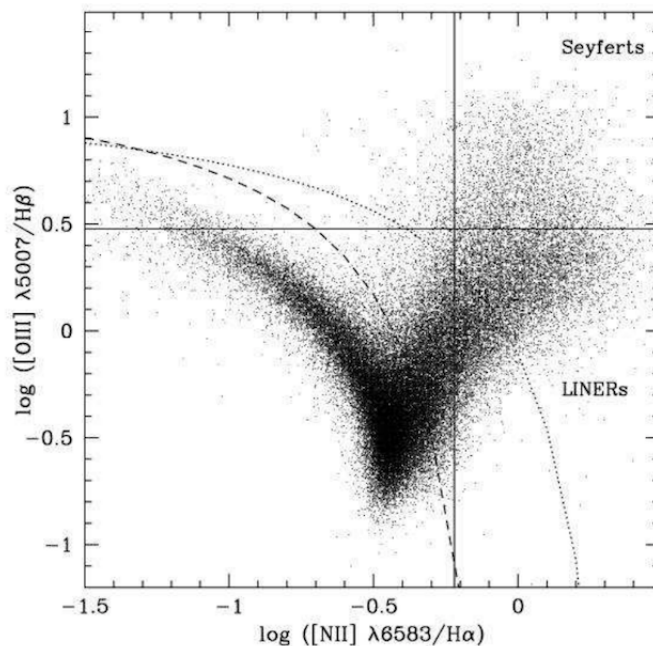


Figure 3.13: Example of a BPT diagram that uses four strong optical emission lines, $H\alpha$, $H\beta$, $[OIII]\lambda 5007$, and $[NII]\lambda 6583$, to distinguish SDSS galaxies from Seyfer 2 galaxies and LINERs. The dotted curve shows the demarcation between starburst galaxies and AGNs defined by Kewley et al. (2001). The dashed curve shows the demarcation proposed by Kauffmann et al. (2003). The AGNs and star-forming galaxies are well separated, whereas the division between the two AGN groups is less clear. Credit: Kauffmann et al. (2003).

in the unified model, because a transformation between classes in time scales of dozens of years cannot be explained by obscuration. In Fig. 3.14 (Gezari et al. 2017) we show an example of a dramatic change in the spectrum of transient iPTF 16bcc (a quasar at $z = 0.237$), that had its UV flux increased by a factor of 10, and transformed from a LINER galaxy (SDSS/2004) to a luminous type-I quasar (Keck2+DEIMOS/2016) in a dozen years.

The AGN spectral characteristics give us valuable information about the physical conditions in these extreme environments. Several well-established correlations in type-I AGNs appear under the name “eigenvector 1” because they were first found by a principal component analysis (PCA) of a large group of low-redshift AGNs (Netzer 2013). Such an analysis is powerful to isolate common properties – in this case, spectroscopic properties, that are related to each other and are not easy to find by other methods. Some of these correlations are (i) the anti-correlation between $[OIII]$ and optical FeII line strength, and (ii) the anti-correlation between FWHM of broad $H\beta$ line and the ratio of optical FeII line and broad $H\beta$ equivalent widths. In particular, the Baldwin effect (1977) is the anti-

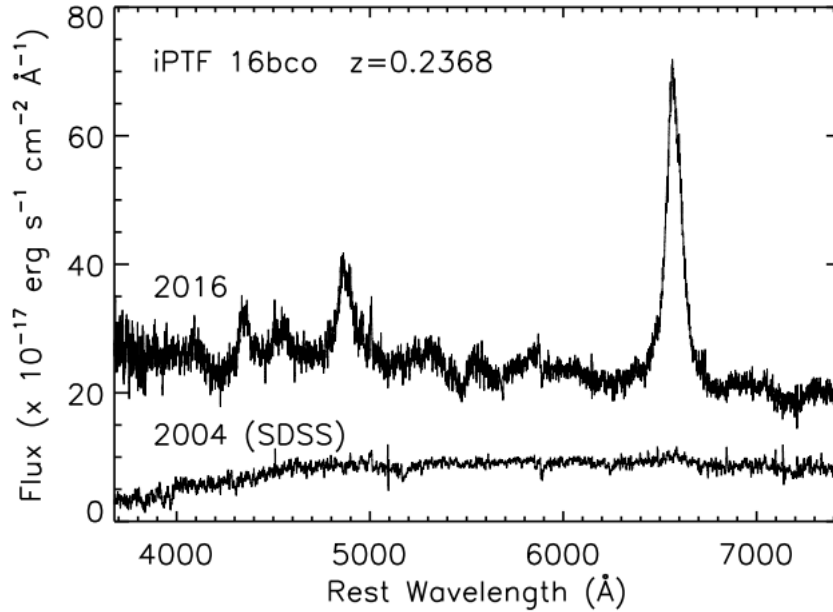


Figure 3.14: Spectral variation of quasar iPTF 16bco. The archival SDSS legacy spectrum was obtained on 2004 June 16, and the follow-up spectrum was obtained by Keck 2+DEIMOS on 2016 June 4. This quasar is located at $z = 0.237$ and was discovered “turning on” by the intermediate Palomar Transient Factory (iPTF). The transient iPTF 16bco shows two remarkable changes: its UV flux increased by a factor of 10, and it transformed from a LINER galaxy to a luminous type-I quasar. Credit: [Gezari et al. \(2017\)](#).

correlation between the equivalent width of some spectral lines (e.g. CIV, Ly α , etc.), and the continuum luminosity at 1350 Å. The observational consequence of this effect is a decrease in the ratio of brightness of the emission line to the brightness of the nearby continuum with increasing luminosity of the continuum.

Recently, two classification schemes were proposed as an alternative to the old unified model. The first, proposed by Padovani (2017), is based on the presence (“jetted”) or absence (“non-jetted”) of a relativistic jet – in opposition to the old nomenclatures “radio quiet” and “radio loud” sources. The second, proposed by Heckman & Best (see §3.1.5 for more details) separates the AGNs into two major groups: radiative mode and jet mode, depending on their nuclear activity.

3.2.1 Quasars

Since this thesis will focus on photometrically selected quasars, we dedicate the last paragraphs of this section to describe some additional properties of these sources.

Quasars are the most luminous sources in the Universe; their luminosities can exceed that of “normal” galaxies by a factor of $\sim 10^3$. In the optical range, the quasar

spectrum exhibits a characteristic non-thermal continuum often represented by a power-law shaped spectrum of the form $L_\nu \propto \nu^{-\alpha}$ accompanied by a set of strong emission lines, signatures of high ionization. The main quasar emission lines in the UV/optical are $\text{Ly}\alpha$, SiIV , CIV , CIII , MgII , $\text{H}\beta$, OIII and $\text{H}\alpha$ (a more complete list is shown in Table A.3).

In addition to the emission lines, their spectrum may also contain absorption features, 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. Neutral hydrogen intersected by the line-of-sight to the quasar will produce numerous narrow absorption lines at $\lambda_{obs} \lesssim (1 + z_{em}) \times 1215 \text{ \AA}$. 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.

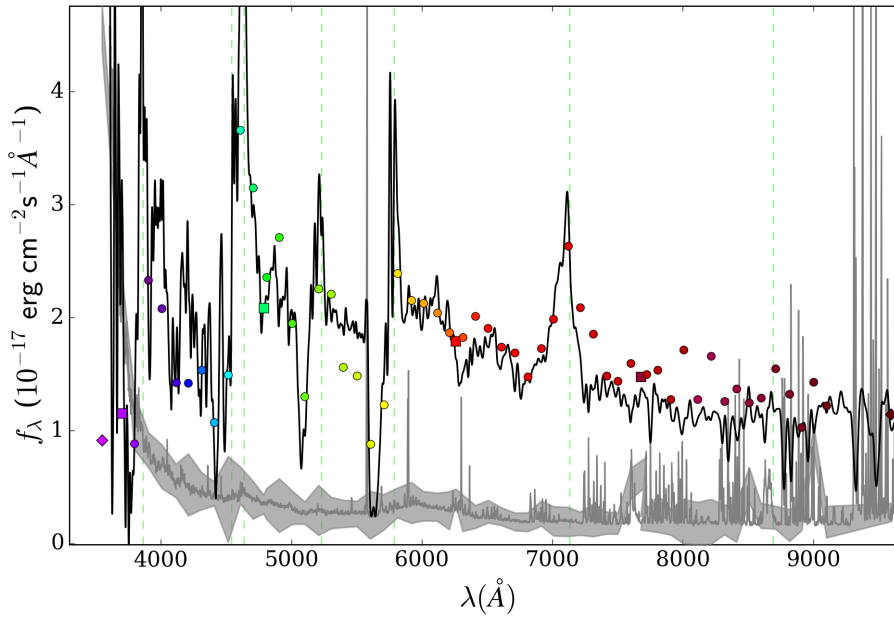


Figure 3.15: Example of a BAL quasar at $z_{\text{spec}} = 2.736$. The black solid line corresponds to the spectrum measured by SDSS DR16 and the gray solid line in the bottom to its associated uncertainty. The colored dots (squares) represent the photometric fluxes measured within an aperture of $3''$ by the narrow (broad)-band filters of the J-PAS system. The gray solid region at the bottom corresponds to the error bars of the photometric fluxes. The green dashed lines indicate the main emission lines. This quasar corresponds to the source 2406-11219 in the mini-JPAS catalog and has a magnitude of $r = 20.481$.

Another class of absorption are the broad absorption lines, known as BALs, which can be found in about 10 – 20% of type-I AGNs ([Shankar et al. 2008](#)) and originate from material in the AGN itself. In general, the redshift of the BAL 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. These troughs are associated in most cases with strong resonance lines of CIV λ 1549, SiIV λ 1397, NV λ 1240, OVI λ 1035 and Ly α , and sometimes to MgII λ 2798 or FeII lines. In Fig. 3.15 we show an example of a BAL quasar detected at $z_{\text{spec}} = 2.736$ by SDSS DR16. This quasar was also detected by the miniJPAS survey (for more details, see Chapter 6), and hence the figure depicts not only its spectrum (black solid line), but also the photometric fluxes (colored dots and squares) detected in the J-PAS filters. This quasar shows three broad absorption lines associated with Ly α , SiIV and CIV (from left to right).

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.

3.3 Searching for quasars

There are many techniques for detecting quasars⁶, each one probing different wavelength regimes of the electromagnetic spectrum. In this section we give an overview on the more traditional methods to discover quasars.

From their optical spectra featuring broad emission and absorption lines on top of an underlying non-thermal continuum, it is relatively easy to distinguish quasars from stars and galaxies. However, spectroscopy is not so efficient in discovering type-II quasars because of their relatively weak emission lines and narrow-lines that resemble those from galaxies with strong star formation activity. Besides, spectroscopic surveys usually rely on a photometric pipelines to select quasar candidates for follow-up spectroscopy, and the fibers end up allocated only to the brightest objects.

Next, still taking into account the differences in the SEDs, a simple and efficient way of discovering quasars uses broad-band multicolor photometry. Several color combinations, such as $u-g$ and $g-r$ from SDSS, and $W1-W2$ and $W2-W3$ from WISE, have been employed for separating quasars from stars in the sky (Richards et al. 2002; Nikutta

⁶From now on, we shall use indistinctly “quasars” to refer to type-I AGNs – unless specified.

et al. 2014; Peters et al. 2015). Color-color selections are useful for discovering low-redshift sources but fails to detect many high-redshift objects because the spectrum gets effectively red resembling the colors of nearby stars, although even low-redshift quasars can be confused with the local population of hot white dwarfs. In addition, some quasars are intrinsically red, or reddened by dust, which results in colors that are not very different from those of stars (see Fig. 2.7 and §2.2 for more details on the quasar target selection performed by SDSS). In fact, combining optical with infrared information is a powerful way of improving multi-object classifications and finding quasars. Moreover, selecting quasars as sources with zero proper motions in GAIA (Gaia Collaboration et al. 2018) can help us break some of these degeneracies (e.g., Heintz et al. 2018).

Mid-infrared techniques combined with X-ray data are also relevant to find highly obscured (Compton thick) quasars (Baloković et al. 2014; Del Moro et al. 2016), which have their SEDs dominated by warm dust emission as a result of the heating of the torus by the central engine.

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 $\lambda = 1215 \text{ \AA}$, due to the ionization of hydrogen, seen in the bluest filters. Moreover, most of the quasars in the range $5.5 < z < 6$ have been found by looking for sources with a large break between the optical i and z -bands (e.g. Fan et al. 2006), the so-called i -band dropouts. In Fig. 3.16 we show an example of the first high-redshift quasar detected by the Panoramic Survey Telescope and Rapid Response System 1 (Pan-STARRS; Kaiser et al. 2002) using the i -band dropout.

In order to find quasars beyond $z \sim 6.5$ wide-field surveys with coverage beyond $\sim 1 \mu\text{m}$ (i.e., NIR) are needed. The optical spectra of these $z > 6.5$ quasars is actually useful to map the intergalactic medium and put constraints, e.g., on the fraction of neutral hydrogen (Bolton et al. 2011) or on the metal enrichment (Simcoe et al. 2012) along the line-of-sight to the quasar.

Quasars have significant variability in the u -band, and the fraction of quasars with large two-epoch variability is much higher than that of stars. Targeting by variability can select both low-redshift quasars with UV excess and mid-redshift ($2 < z < 3.5$) quasars

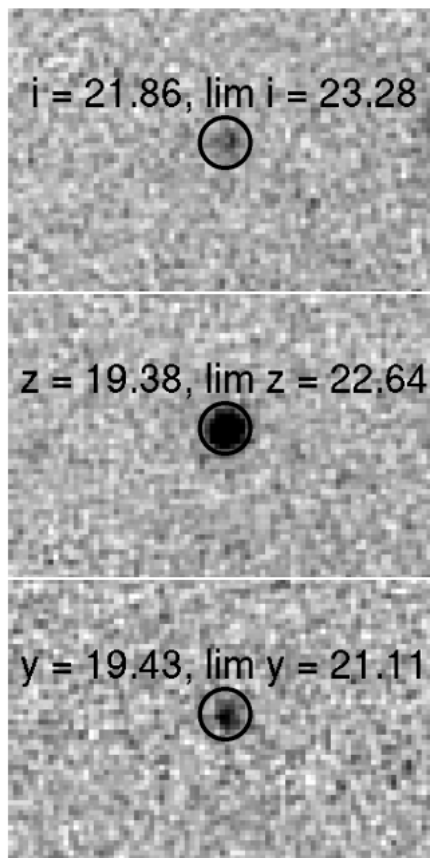


Figure 3.16: First quasar (PSO J215.1512-16.0417) discovered by Pan-STARRS 1 using the technique of *i*-dropout. Note that this quasar was detected in *z* and *y* bands (middle and bottom panels) but not in the bluer filter shown here (*i*, top panel). It has an estimated redshift of 5.73. Adapted from: [Morganson et al. \(2012\)](#).

where quasar selection by optical colors is inefficient ([Schmidt et al. 2010](#)).

Almost all AGNs are also strong X-ray emitters, providing an efficient route for locating optically faint quasars at high redshifts ([Barger et al. 2003](#); [Castander et al. 2003](#)) and discovering many type-IIs – since the amount of obscuration at higher energies (> 15 keV) is much smaller. However, this method is not so efficient in discovering high redshift quasars, because of the sharp drop of X-ray luminosity of such sources.

Finally, about 10% of all AGNs are core-dominated radio-loud sources; therefore, correlating the radio and optical positions provides an additional way to identify quasars ([White et al. 2000](#)). Since stars are extremely weak radio sources, this means that an optical point-source that is also a strong radio source is likely to be a radio-loud quasar.

All discovery methods have their own band-dependent biases and, in particular, flux-limited samples typically miss fainter sources in the band in question. Hence, understanding the selection biases, and combining surveys of different wavelength bands in a

proper statistical manner, are essential for revealing the underlying population properties.

Searching for quasars in photometric surveys is facilitated when images are available in a different number of filters covering the whole spectral energy distribution in the optical. In particular, the narrow-band surveys J-PLUS and S-PLUS have filters centered on key stellar spectral features, delivering low-resolution spectra for several types of sources. The J-PAS unique photometric system consisting of 54 contiguous narrow-band filters will have an even more outstanding performance for finding quasars, being able to resolve the broad emission-lines of type-I_s and detecting the narrow-lines of many type-II_s.

In this new era of massive data acquisition, we need effective statistical methods to classify the sources detected by these photometric surveys. This can be done with template-fitting methods (e.g. star/quasar separation in ALHAMBRA and J-PLUS: Queiroz, M.Sc. dissertation, University of São Paulo, 2015) or machine learning algorithms (Costa-Duarte et al. 2019; Baqui et al. 2020; Nakazono et al. 2020 – submitted; Villa-Nova, Abramo, Queiroz et al. in prep).

Recently, machine learning approaches have been more widely used in astronomy, being preferred when great statistical precision is needed. However, at the same time, it requires large and representative training sets, otherwise any uncertainties or incompleteness of the training set will be reflected on the resulting trained model.

In §3.3.2 we show some results using machine learning algorithms to identify quasars in S-PLUS and miniJPAS. The S-PLUS DR1 contains more than three million sources covering $\sim 336 \text{ deg}^2$ of the Stripe 82 area and have spectroscopic classifications for more than 100,000 objects. This means that the number of sources with trustworthy classifications is more than enough to be separated into training and test samples. However, this is not the case of the miniJPAS catalog, in which spectroscopic classifications are only available for a few hundreds of sources. To solve this problem, we constructed mock catalogs with fluxes simulated in the J-PAS filters, considering realistic error distributions, and employed these mocks for the training (see Chapter 5 for more details). In order to ensure that both the frequency and redshift-magnitude (or, equivalently, type-magnitude in the case of stars) distributions were properly represented in the mocks, we

derived the number densities expected by luminosity functions.

In the next section we present the luminosity function for quasars. We will discuss the luminosity functions employed for stars and galaxies later on (see §5.2).

3.3.1 Quasar luminosity function

The luminosity function (LF) is a key quantity in any study of galaxy evolution: it describes the distribution of a given class of sources (e.g., galaxies) with different intrinsic luminosities L (or magnitudes M). Since the main source of data coming from the distant past is provided by the light emitted (or scattered) by ordinary matter, the LF also supplies the primary information about the baryonic content of the Universe (Alcaniz & Lima 2004).

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, which means that a large volume must be surveyed. An additional effect comes from Malmquist bias, which leads to a preferential detection of intrinsic bright galaxies, visible at larger distances, in flux-limited surveys.

In general, the LF is defined by the differential expression $d\Phi(L) = \phi(L)dL$, where $d\Phi(L)$ is the number of galaxies per unit volume with luminosity in the interval $[L, L + dL]$. To a good approximation, the distribution of galaxies can be represented by the Schechter luminosity function (1976):

$$\Phi(L)dL = \phi_* \left(\frac{L}{L_*}\right)^\alpha \exp\left(-\frac{L}{L_*}\right) d\left(\frac{L}{L_*}\right) \quad (3.10)$$

where ϕ_* is the number density of galaxies (or “break” density), L_* is a characteristic luminosity, and α is the slope of the luminosity function for small L , typically assuming values in the range $-1.5 < \alpha < -1$. A schematic plot of the Schechter luminosity function is shown in Fig. 3.17.

This relation was inspired by the Press-Schechter theory, which predicts the number density of virialized halos above a certain mass. Similarly to the high-mass function behaviour, very luminous galaxies are rarer, implying that $\Phi(L)$ decreases exponentially for $L > L_*$. Note that Eq. 3.10 cannot be deduced from first principles, and as such, the

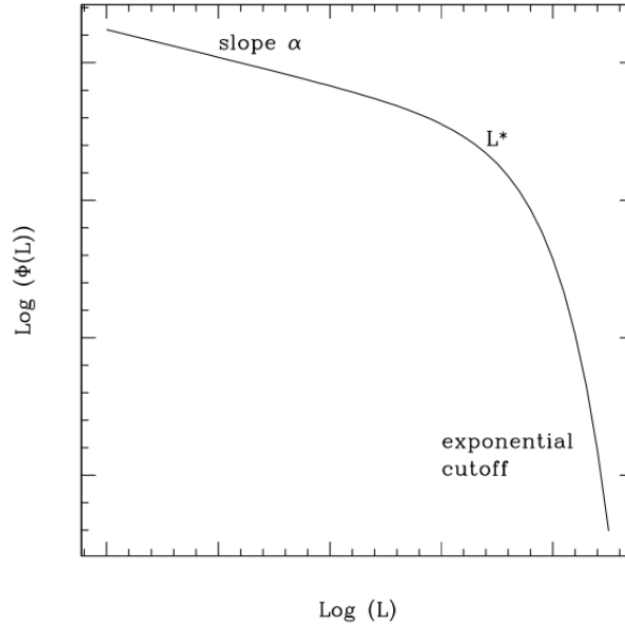


Figure 3.17: Schematic Schechter luminosity function, showing the slope α at low luminosities, the characteristic luminosity L_* , and the exponential cut-off for bright galaxies. Credit: [Schneider \(2006\)](#).

free parameters must be determined from astronomical observations, which means that these parameters are also sensitive to the kind of galaxy being sampled. Finally, feedback mechanisms, such as supernova explosions and AGN outflows, play an important role in modifying the luminosity function, leading to a constant density-core profile at low luminosities and suppressing the number of galaxies at high luminosities ([Silk & Mamon 2012](#)).

Actually, a magnitude⁷ representation of the luminosity function may be preferable because $\Phi(M)$ is more directly related to observations. Since M and L are related by $M - M_* = -2.5 \log(L/L_*)$, in terms of the magnitude the distribution of galaxies reads

$$\Phi(M)dM = 0.4 \ln 10 \phi_* 10^{-0.4(M-M_*)(\alpha+1)} \exp[-10^{-0.4(M-M_*)}] dM \quad (3.11)$$

where M_* is a cut-off magnitude.

The galaxy luminosity function may have different parameters for different populations and environments, i.e., it is not a universal function. Therefore, although the

⁷The absolute magnitude (M) can be measured as $M = m - 5 \log_{10}(D_L/1 \text{ Mpc}) - 25$, where m is the apparent magnitude, and D_L is the luminosity distance (in units of Mpc) – which depends on the cosmology.

Schechter function seems an adequate representation for the total distribution of galaxies, deviations from it are expected. In particular, the quasar luminosity function (QLF) is usually parametrized by a double power-law (Boyle et al. 2000; Croom et al. 2004; Richards et al. 2006a):

$$\Phi(L, z) = \frac{\phi_*(L)}{(L/L_*)^\alpha + (L/L_*)^\beta} \quad (3.12)$$

per unit luminosity, where α and β are respectively the power-law indices for bright and faint quasars, or

$$\Phi(M, z) = \frac{\phi_*(M)}{10^{0.4(\alpha+1)[M-M_*(z)]} + 10^{0.4(\beta+1)[M-M_*(z)]}} \quad (3.13)$$

per unit magnitude. Note that the QLF was found to evolve with redshift, and this evolution is encoded in the redshift dependence of the break luminosity, ϕ_* , and also potentially in the evolution of the power-law slopes. There are two possible scenarios for this evolution: (i) a Pure Luminosity Evolution (PLE), in which $\phi_*(M)$ remains constant with redshift but M_* declines from high redshifts ($z \sim 2$) towards low redshifts; or (ii) a Pure Density Evolution (PDE), in which the break density decreases with increasing redshift ($z \gtrsim 2$).

The physical origin of the characteristic double power-law shape of the QLF (at least, up to $z \sim 6$), in contrast to the Schechter-like form of the underlying dark-matter halo mass function, is to reproduce two distinct quasar populations: the brightest ones, hosted in the rarest most massive dark-matter halos; and the average-magnitude to faintest ones, hosted in less massive halos.

The quasar luminosity function is one of the most important observational signatures of quasar populations. It describes the spatial density (number per Mpc^3) of quasars per magnitude as a function of redshift. Hence, studying the change in the luminosity and number density of quasars across cosmic time is fundamental to understand, for example, the origin and nature of massive black holes, the formation and evolution of galaxies, the environments where quasars can be found (which can place important constraints on structure formation at early epochs), as well the contribution of quasars to reionization at $z \gtrsim 6$ (Fan et al. 2006). For instance, several optical and radio studies have suggested that

the periods over which the black hole is active can be short and repeat during the lifespan of the host galaxy, indicating that the SMBH is regularly re-ignited after a period of low activity (Woltjer 1959; Marconi et al. 2004; Best et al. 2005; Saikia & Jamroz 2009; Morganti 2017).

Large quasar surveys have allowed us to study the properties of the quasar population with unprecedented statistical precision. Several efforts have been performed to measure the QLF, such as the use of photometrically identified quasars (Richards et al. 2005; Hopkins et al. 2007; Croom et al. 2009; Ross et al. 2013), as well as quasars detected in the mid-infrared (Brown et al. 2006; Siana et al. 2008; Assef et al. 2011), and X-rays (Cowie et al. 2003; Hasinger et al. 2005; Aird et al. 2010; Fiore et al. 2012).

In this thesis, we adopt the pure luminosity evolution function from Palanque-Delabrouille et al. (2016), which assumes that the luminosity of all quasars scales up according to some function of redshift, and it allows the bright-end and faint-end slopes to be different on either side of a pivot redshift set as $z_{pivot} = 2.2$.

Figure 3.18 shows the differential quasar counts $\Phi_q(m, z)$ as a function of apparent magnitude (in the r -band) and redshift for the PLE luminosity function (QLF-PLE hereafter) over an area of 10,000 deg². Considering a perfect selection of objects, we find that a large-area ($\gtrsim 1/5$ of the sky) flux-limited ($r < 23.5$) survey could yield more than three million quasars up to $z = 6$.

Starting from $\Phi_q(m, z)$, we can define the total number of quasars with magnitude m_i over an area of solid angle $\Delta\Omega$ at any redshift as

$$\bar{N}_q = \int_0^\infty dz \frac{dV}{dz} \int_{m_i}^{m_i + \Delta m_i} dm W(m, z) \Phi_q(m, z) \quad (3.14)$$

where V is the comoving volume, $W(m, z)$ is an arbitrary selection function which, in principle, depends both on magnitude and redshift, and i denotes the index of the magnitude bin. In this representation, the probability of finding a quasar with magnitude m_i at redshift z_j is given by

$$P_{ij}^q = \frac{1}{\bar{N}_q} \int_{z_j}^{z_j + \Delta z} dz \frac{d\bar{N}_q(m_i, z_j)}{dz}. \quad (3.15)$$

At the high redshift range, the overall shape of the quasar luminosity function

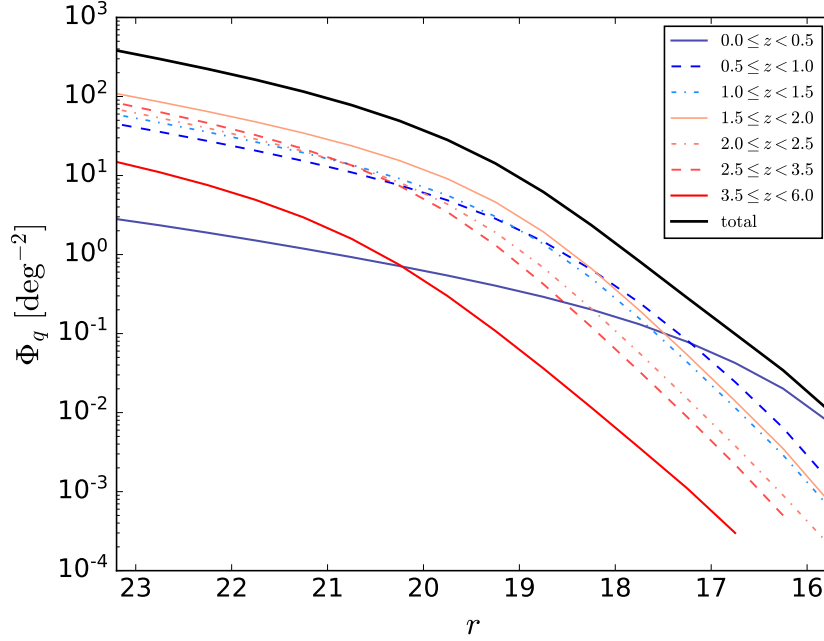


Figure 3.18: Cumulative distributions of the number density of quasars per deg^2 as a function of the r -band magnitude (reversed to show dim quasars on the left and bright quasars on the right). The different colored lines show the expected number of quasars at different redshift ranges, and the solid black line corresponds to the cumulative number of quasars up to $z = 6$. Here we adopted the PLE from [Palanque-Delabrouille et al. \(2016\)](#).

evolves only modestly up to $z \sim 6$ ([Matsuoka et al. 2018](#)), whereas the break density evolves strongly with redshift ([Jiang et al. 2016](#)).

When modeled as an exponential decay in redshift, [Fan et al. \(2019\)](#) have shown that the density of luminous quasars drops by a factor of ~ 3 per unit redshift in the range $z = 3 - 5$; this decline accelerates to a factor of ~ 6 per unit redshift in the range $z = 5 - 7$. Translating these results to an e-folding timescale, it implies in a quasar density evolution of about 400 million years at $z \sim 4$, and as short as 80 million years at $z \sim 7$. Assuming a radiative accretion efficiency of $\eta \sim 0.1$, the latter is comparable to the Eddington timescales of 45 million years for the SMBH growth, what suggests that the quasar density growth is mainly driven by the maximum rate of accretion onto the central BH. The quasar luminosity function at high redshifts is shown on the left panel of [Fig. 3.19](#).

Furthermore, by extrapolating the density evolution towards higher redshifts ($z > 6$), [Fan et al. \(2019\)](#) have predicted that there will be only ~ 1 luminous ($M_{\text{AB}} < -26$) quasar at $z \sim 9$ in the observable Universe. This result establishes $z \in [9, 10]$ as the probable epoch for the appearance of the very first luminous quasars. The right panel of

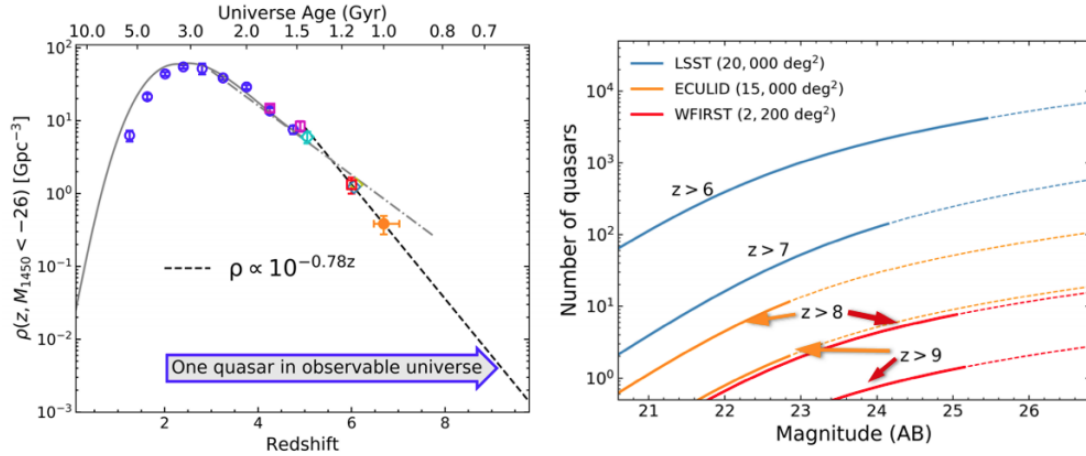


Figure 3.19: Left panel: Density evolution of luminous quasars. Extrapolating the measurements at $z = 5 - 7$, the expectation is of only ~ 1 luminous quasar powered by a billion-mass black hole in the entire observable Universe at $z \gtrsim 9$ (Wang et al. 2019). Right panel: Forecast of quasar discoveries in wide-field imaging surveys in the next decade. Dropout selection will discover quasars down to $m_{\text{AB}} \sim 23 - 25$ (solid lines), establishing a large sample of quasars at $z > 7$ (Vera C. Rubin Observatory/LSST: Ivezić et al. 2008), with a few dozen objects at $z > 8$ (Euclid: Laureijs et al. 2011; Nancy Grace Roman Space Telescope/WFIRST: Spergel et al. 2015). Credit: Fan et al. (2019).

Fig. 3.19 shows a forecast for the number of quasars that will be discovered in wide-field imaging surveys in the next decade, demonstrating that the high-redshift ($z > 7$) Universe is within reach of the capabilities of the next generation of surveys. Therefore, detecting large statistical samples of luminous quasars at $z = 7 - 9$ should be a priority for the high-redshift quasar community.

3.3.2 Classification using machine learning

Here we present some results on the classification of point-like sources detected in narrow-band surveys using machine learning.

Star/quasar separation in S-PLUS

In our initial approach, we considered a 2-class (i.e., star/quasar) separation. In the literature, there are different methods to perform the star-quasar separation, such as SDSS XDQSO (Bovy et al. 2011; see also §2.2.2), a probabilistic target selection technique that models the distributions of stars and quasars in flux space down to the flux limit by applying the extreme-deconvolution method to estimate the underlying density; and ELDAR (Chaves-Montero et al. 2017), a method to identify AGNs in medium- and

narrow-band filter surveys through the identification of emission lines. In our analysis, we tackled this problem by employing a series of machine learning methods to compute the posterior probability of each object belonging to a certain class.

Since this was done prior to the first data release from S-PLUS, we employed mock catalogs in our analysis. We will explain more thoroughly how we generated these mocks in Chapter 5, but for now it suffices to say that the fluxes were simulated by convolving the stellar and quasar spectra with the S-PLUS transmission curves, and the corresponding uncertainties were derived from the theoretical limiting magnitudes obtained for an aperture of 3" at a level of signal-to-noise ratio $S/N=5$. Then, we assumed Gaussian errors (which is not always true, but is a good approximation at first order) and added Gaussian fluctuations to those fluxes by setting the variance equal to the uncertainty in each filter. In order to have training sets as complete and representative as possible (and guarantee, thus, that the algorithm would be able to recognize most of the spectral patterns for each class), we simulated samples of stars and quasars and with both high and low signal-to-noise ratio fluxes (i.e., bright and faint sources up to $r \sim 21.5$).

For the quasars, fainter fluxes were obtained as follows. First, we binned the magnitude-redshift distribution from our sample in the same way as the quasar luminosity function (QLF-PLE). Next, we selected N_{ij}^Q objects to compose the bright subsample, where N_{ij}^Q is the number of quasars in the i -th magnitude bin and j -th redshift bin predicted by the QLF-PLE. Then, we completed the bins of fainter magnitude with the remaining objects by fixing their redshifts and randomly degrading their magnitudes within a given magnitude bin. By doing this, we were able to shift the original magnitude distribution from SDSS (which is only complete down to $r \sim 20.2$) to fainter values in order to resemble the limiting magnitudes expected for S-PLUS ($r \sim 21.5 - 22$). Note that we can only simulate fainter objects by adding noise to the original fluxes; the opposite (i.e. remove noise to simulate brighter objects) is not possible – at least, not in this simplified approach.

In the case of the stars, we simply degraded their magnitudes in a randomized way, so that they were shifted to the same magnitude range adopted for the quasars (i.e., $18 < r < 22$).

Our data set was composed of $\sim 96\,000$ quasar spectra with r -band magnitudes in the range $[18, 23.5]$ and redshifts $0.5 \leq z \leq 4.0$ selected from the SDSS DR16, and about 21 000 stellar spectra with $r > 16$, mainly A to M main sequence types plus white dwarfs selected from the SDSS Spectra of Everything. The Spectra of Everything is a subsample of stellar objects covering an area of approximately 240 deg^2 used as a substitute for a series of merged program plates (chunks `merger48` and `merger73`). This data set included a random sampling of all point sources with clean photometry, and the vast majority of the targets (later confirmed to be stars) were chosen from the densest core of the stellar locus in color-color space having reddening-corrected i -band PSF magnitudes brighter than 19.1.

The data set was separated into three different samples: training and validation sets, used to train and optimize the various machine learning methods; and an evaluation set, i.e., the sample to be tested. In the end, the quasars were separated as: $\sim 80\,000$ objects for the training, 10,500 for the validation and 1 725 for the evaluation. And, for the stars we had: $\sim 15\,000$ objects for the training and an equal number of about 2 600 objects for both the validation and evaluation sets. We ensured that these sets did not contain repeated objects.

From models of stellar population synthesis, e.g., Besançon Model (Robin et al. 2003) and TRILEGAL (Girardi et al. 2005), we expect to observe approximately 10-40 times more stars than quasars per square degree (depending on the area in the sky), which means that the count of stars in our sample is not realistic. Hence, while performing the classification, we had to take into account this problem of class dominance and balance the data set, otherwise the accuracy of the classification obtained would be biased by the dominant class – always being numerically high.

Furthermore, luminosity priors are important to avoid non-realistic solutions (Salvato et al. 2009), and they should be chosen depending on the type of objects that we want to target. Therefore, we applied the following criteria:

- for the bright end, we rejected all objects i for which all the filters α satisfied $\frac{\sigma_{f_{\alpha,i}}}{f_{\alpha,i}} \lesssim 0.04$, i.e. $S/N \gtrsim 25$;
- for the faint end, we rejected all objects i for which at least 5 filters α satisfied

$$\frac{\sigma_{f_{\alpha,i}}}{f_{\alpha,i}} \gtrsim 0.2, \text{ i.e. } S/N \lesssim 5.$$

The classification was performed using the publicly available code ANNz2 (Sadeh et al. 2016), which utilizes multiple machine learning methods, such as artificial neural networks and boosted decision/regression trees. In Fig. 3.20 we show an example of an artificial neural network (ANN) architecture, made up of several layers, each consisting of a number of interconnected nodes. In order to recognize and identify the main features in the evaluation set, the ANN usually employs an adaptive learning process, in which every connection carries a weight, until one obtains acceptably low (i.e., below some threshold) error rates.

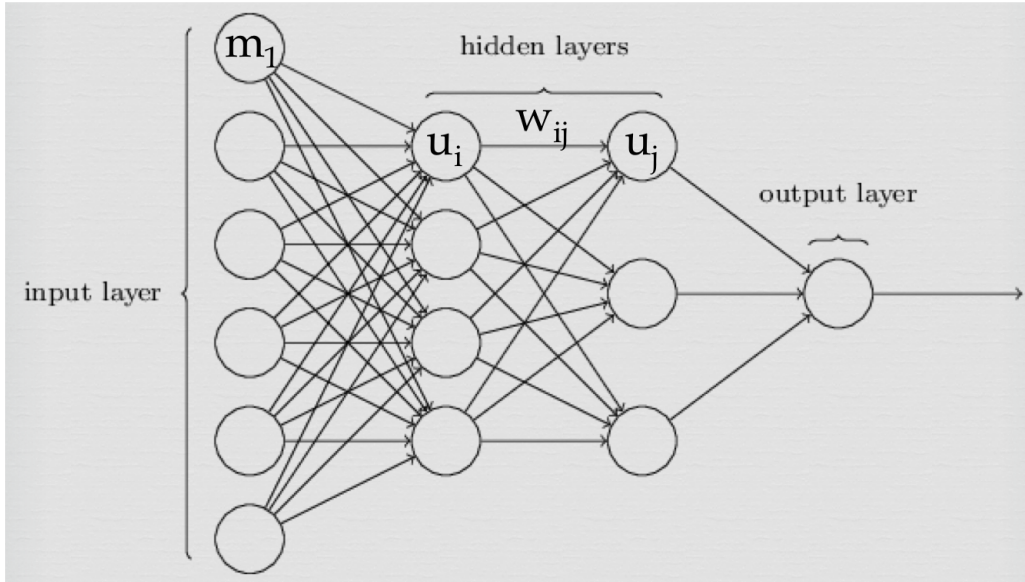


Figure 3.20: Example of an artificial neural network architecture. In the figure, m_1 is the value of the first parameter in the input layer (e.g. the magnitude in the first filter). The node u_i in a given layer is connected to all nodes in the adjacent layers, and every connection carries a weight w_{ij} . Adapted from: Collister & Lahav (2004a).

In particular, in our analysis we employed the randomized classification mode of ANNz2, which applies an ensemble of random machine learning methods during the training. We have trained many (typically, order of 100 or more) different machine learning methods (MLMs), based on the same datasets (i.e., training and validation sets). After the MLMs have finished training, they are able to assign values (scores S_i^n) between 0 and 1 to each object ($i = 1 \dots N_{obj}$) of the evaluation set, yielding $n = 1 \dots N_{MLM}$ predictions for the classification of each object.

Notice that even though the values 0 and 1 are associated with the confidence that an object is either a quasar or a star, the threshold of the score which separates the two

classes is not identical for all MLMs: some are closer to $S \sim 0.5$, but some may be closer to $S \sim 0.4$ or $S \sim 0.6$. Another important aspect to take into account is that not all trained MLMs are successful: based on the validation sets, we can perform tests (such as the Kolmogorov-Smirnov test, also known as K-S) to reject those MLMs that have failed.

After discarding unsuccessful MLMs ($N_{\text{MLM}} \rightarrow N'_{\text{MLM}}$) and rescaling the scores ($S_i^n \rightarrow S'_i{}^n$), we produced histograms of the scores of all MLMs for each individual object. The histograms are then identified (in the limit of $N'_{\text{MLM}} \rightarrow \infty$) with the likelihood that the object i is of a certain type. The mean of the PDF for the type of an object can be estimated as:

$$\bar{S}_i = \frac{1}{N'_{\text{MLM}}} \sum_{n=1}^{N'_{\text{MLM}}} S'_i{}^n. \quad (3.16)$$

The mean score, applied to a validation set consisting of 10 500 quasars and 2 600 stars, is shown in Fig. 3.23.

Besides the mean, we can also estimate the variance of an object's PDF as:

$$\sigma_i^2 = \frac{1}{N'_{\text{MLM}}} \sum_{n=1}^{N'_{\text{MLM}}} (S'_i{}^n - \bar{S}_i)^2. \quad (3.17)$$

The variance (or, if preferred, some other measure of the dispersion, such as the absolute mean deviation) is a useful measure of the uncertainty associated with the prediction, and can be used to separate good from bad classifications.

In Fig. 3.21 we show how the variance of the PDFs can be used to help in the classification of the quasars (blue dots) and stars (in red) of our evaluation sample. We have split the two types of objects in the left and right panels for visualization purposes – in reality, they would obviously appear mixed in the same plot. The vertical axes correspond to the mean renormalized scores \bar{S}_i of each object, while the horizontal axes (which is reversed in the case of stars) corresponds to the variance (σ_i^2) of the PDFs.

For instance, if we make an arbitrary quality cut (represented by the black dashed lines in the plot), it means that the stars inside the red dashed region ended up wrongly classified as quasars, corresponding to the contaminants of the population of quasars, while the quasars inside the blue dashed region are the ones that ended up wrongly classified as stars.

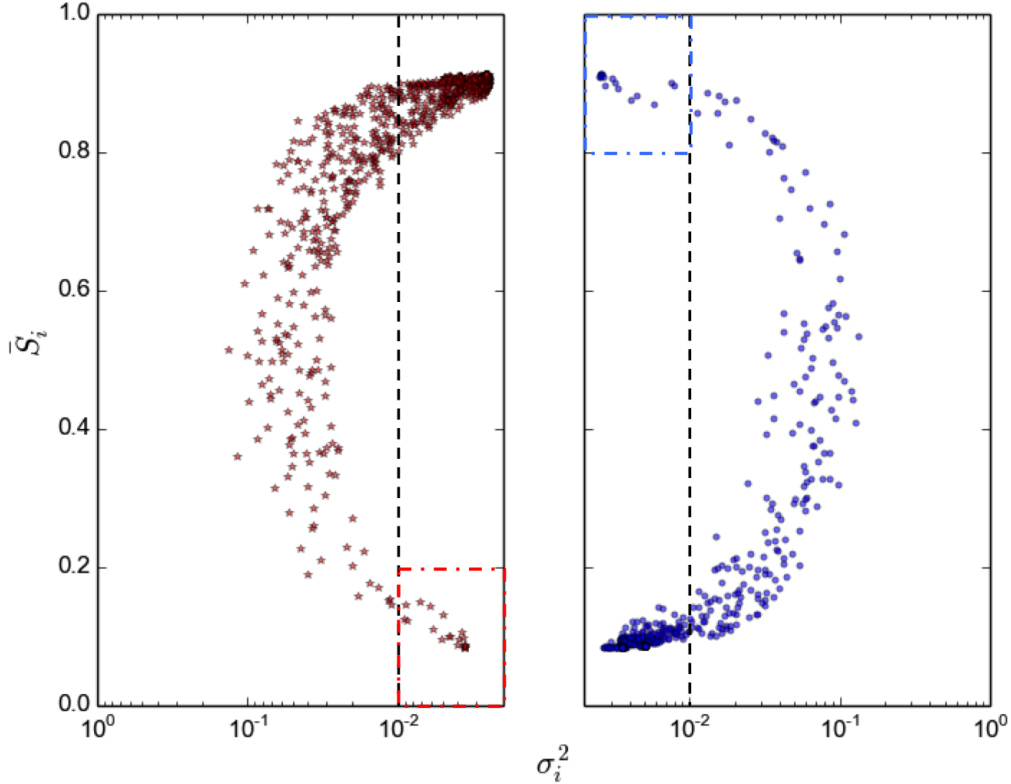


Figure 3.21: Star (left panel) and quasar (right panel) classification with ANNz2. The x -axis (reversed in the case of stars) corresponds to the variance of the PDF; the y -axis corresponds to the mean renormalized scores of each object, from 0 to 1. In this convention, a classification equal to “0” corresponds to a quasar, while a classification equal to “1” corresponds to a star. The black dashed lines represent an arbitrary quality cut: the stars inside the red dashed region correspond to the contaminants of the population of quasars; the same argument applies for the quasars inside the blue dashed region.

In Fig. 3.22 we show the histograms for the mean classification (over all of the randomized machine learning methods) for 9 randomly selected quasars and stars. In Fig. 3.23 we show the histograms for the mean classification for all the quasars and stars.

The method seems to have a high performance level, because only a few stars ended up classified as quasars (i.e., the contamination sample is small). By taking into account a more realistic relative frequency between the star and quasar samples (a factor of 40 in our case), we can make an assessment about the performance of the classification by computing the purity p_k and the completeness C_k for an object of kind k (quasar or star):

$$p_k = \frac{N_k}{Q_k + S_k} \quad (3.18)$$

where N_k is the number of real objects of kind k that were classified as kind k , while Q_k

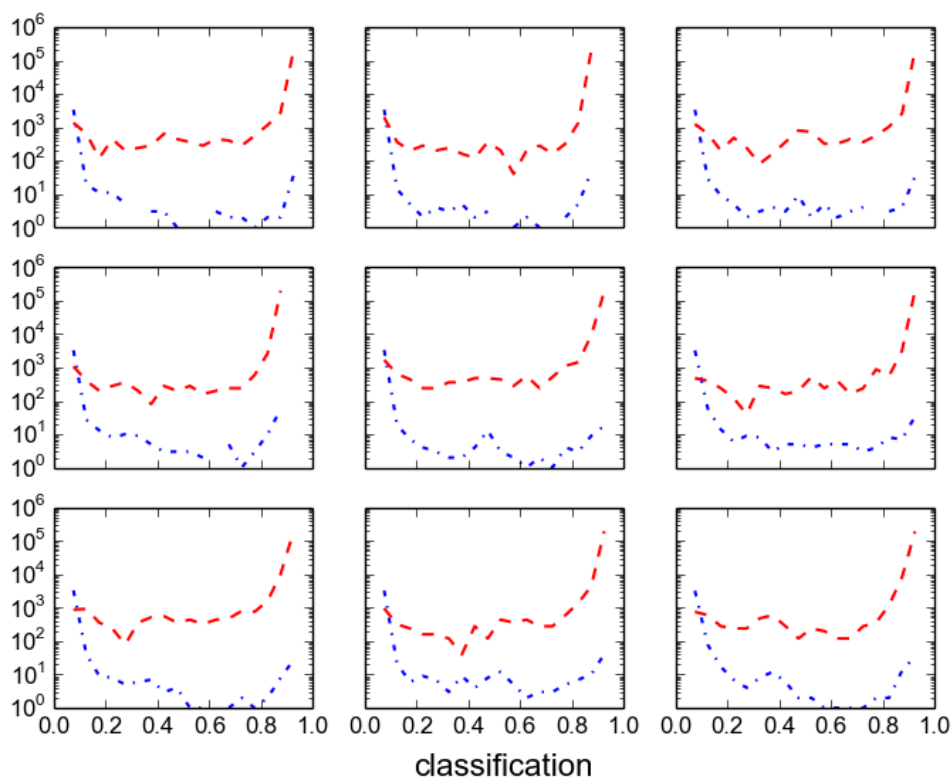


Figure 3.22: Mean classification over all the MLMs for 9 randomly selected stars (red dashed lines) and quasars (blue dot-dashed lines).

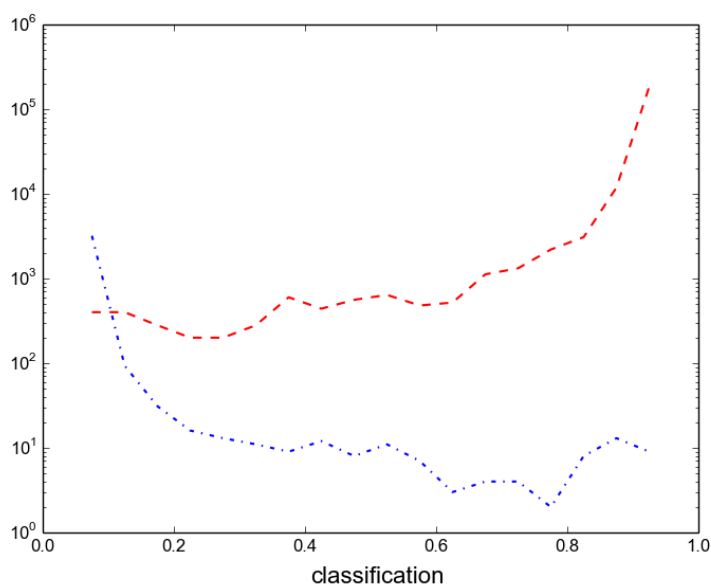


Figure 3.23: Mean classification over all the MLMs for all the stars (red dashed line) and quasars (blue dot-dashed line).

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

$$C_k = \frac{N_k}{T_k} \quad (3.19)$$

where T_k is the total number of objects of kind k in our evaluation sample (i.e., 1 725 quasars and $40 \times 2\,600$ stars).

We can also compute the purity and the completeness as a function of some quality factors. In this case, we have chosen three quality cuts as a function of the variance: high ($\sigma_i^2 < 0.01$), medium ($\sigma_i^2 < 0.05$) and low ($\sigma_i^2 < 0.1$).

In tables 3.1 and 3.2 we show the results for the classification of the quasars and stars, respectively. In the end, we want both the quasar completeness and the purity of stars to be as high as possible, to guarantee a complete (and with a low level of contamination) catalog of quasars. As we can see from the tables, we obtain a high level of purity for the stars (99.9%) – independently of the quality cut. As for the quasar completeness, we obtain 94.2% for smaller values of variance, and up to 98.1% for less strict quality cuts.

$$\bar{S}_i = \frac{1}{N'_{\text{MLM}}} \sum_{n=1}^{N'_{\text{MLM}}} S'_i{}^n. \quad (3.20)$$

Table 3.1: Purity and completeness for S-PLUS quasars as a function of the quality cut.

Quality cut	Purity	Completeness
high	80.3%	94.2%
medium	57.1%	97.1%
low	50.5%	98.1%

Table 3.2: Purity and completeness for S-PLUS stars as a function of the quality cut.

Quality cut	Purity	Completeness
high	99.99%	93.4%
medium	99.98%	97.6%
low	99.97%	98.3%

Regarding a star/quasar separation, even in the case of a high quality cut we would still have a 20% level of contamination from stars (expected to be uniformly distributed

when averaged over large distances) in the quasar sample, as shown in Table 3.1. This would affect any clustering measurements of quasars by inducing, e.g., correlations on large scales, even though we expect the angular distribution of quasars to be uncorrelated at different redshifts (Pullen & Hirata 2013; Ho et al. 2015). These stars would eventually end up as outliers in the photometric redshift estimation with LePhare and QPz (see Chapter 4).

In Fig. 3.24 we show some of the PDFs obtained for the classification. In general, the PDFs are well behaved. However, note that the quasar in the middle panel of the bottom row and the star in the right panel of the top row ended up wrongly classified.

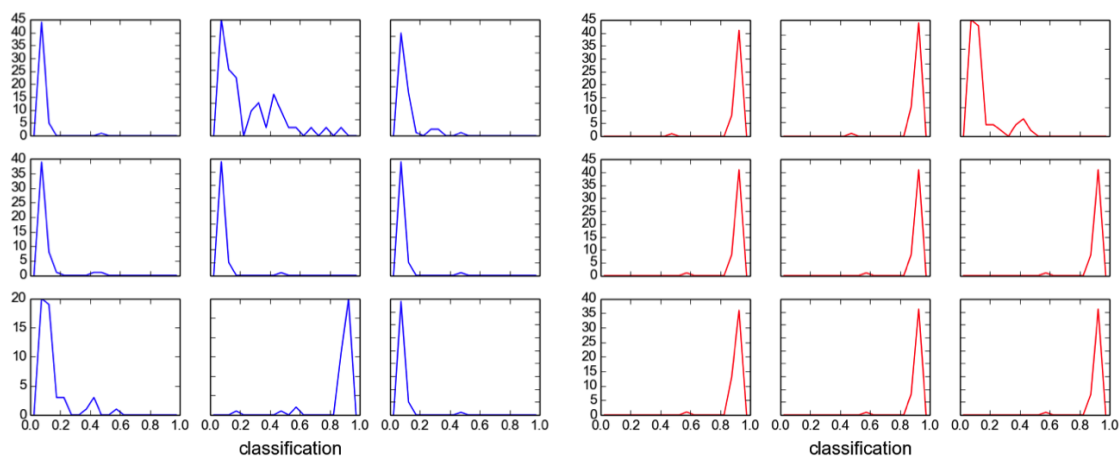


Figure 3.24: Probability density distribution for the classification of 9 quasars (left panel) and 9 stars (right panel) from our S-PLUS evaluation sample. A classification equal to “0” corresponds to a quasar and equal to “1” corresponds to a star.

Note that galaxies can also be misclassified as point-like sources. In particular, Costa-Duarte et al. 2019 have performed a star/galaxy classification for the S-PLUS DR1 using a Random Forest (Breiman 2001) algorithm. They have shown that the inclusion of morphological parameters – specially the full width at half maximum (FWHM) improves the overall accuracy of the classification from 88.1% to 95% for a limited-magnitude sample ($r < 21$).

A star/galaxy classification using machine learning was also performed for the miniJPAS catalog. Baqui et al. (2020) selected two data sets by cross-matching the miniJPAS dataset with SDSS and Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP, Aihara et al. 2019) data, whose classification is trustworthy within the intervals $15 \leq r \leq 20$ and $18.5 \leq r \leq 23.5$, respectively. Then, they trained and tested 6 differ-

ent algorithms on the two cross-matched catalogs: K-Nearest Neighbors, Decision Trees, Random Forest, Artificial Neural Networks, Extremely Randomized Trees (ERT) and Ensemble Classifier – a hybrid algorithm that combines ANN and RF. The best performances were obtained with RF and ERT. They have shown that for the full magnitude range of $15 \leq r \leq 23.5$ the area under the curve (AUC) of the receiver operating characteristic (ROC) curve – a parametric plot of the true positive rate (or completeness) and false positive rate, is 0.957 with RF using only photometric information, while $\text{AUC} = 0.986$ with ERT when using photometric and morphological information (being FWHM the most important feature).

Note, however, that a more thorough assessment of the contaminating objects in the quasar sample can only be obtained with a three-class separation (i.e., with the inclusion of galaxies).

In another exploratory analysis, we performed a 3-class separation of the spectroscopically confirmed sources from S-PLUS DR1 without any near-infrared data by employing a Random Forest algorithm. Again, the training sets contain synthetic fluxes with the same level of noise as in the S-PLUS observation, constructed from SDSS DR12 spectra for main sequence stars and white dwarfs, quasars in the redshift range $0.0 < z < 4.0$, as well as red and blue galaxies. The performance of this method was tested in a sample of about 40k point-like sources detected in the Stripe 82 region. By applying probability cuts on a magnitude-limited sample ($r < 20.5$), we reached a completeness of 76% and a purity of $\sim 94\%$ for the quasars⁸ (of which 1.6% are stars, and 4.2% are galaxies). The purity of the sample with different magnitude cuts is shown in Fig. 3.25 — in particular, note that for $r < 18$ no stars were classified as quasars. Extrapolating these results to the Main Survey (MS) area, we forecast a total number of approximately 703 000 quasars in S-PLUS brighter than $r = 20.5$, with $\sim 94\%$ purity.

Nakazono et al. 2020 (submitted) have provided a catalog of stars, galaxies and quasars based on the S-PLUS DR1. The classification was performed with two different supervised learning algorithms: Support Vector Machine (Cortes & Vapnik 1995) and

⁸Here we are showing two different analyses for both star/quasar and star/quasar/galaxy separation in S-PLUS just to illustrate some applications of machine learning for the classification problem. Note, however, that these results cannot be directly related to the completeness and purity values shown in Table 3.1, since the data sets are not exactly the same. A more thorough comparison between these results is beyond the scope of this thesis.

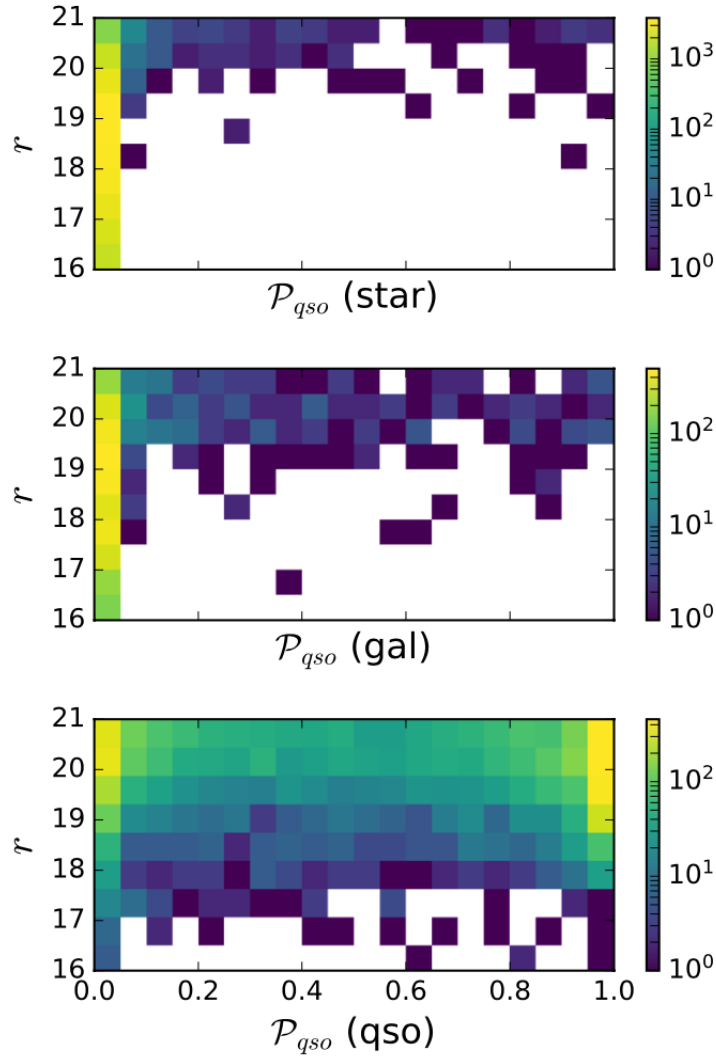


Figure 3.25: Number of S-PLUS quasars detected in the Stripe-82 region that are classified as stars (upper panel), galaxies (middle panel) and quasars (bottom) in different magnitude bins. Credit: C. Queiroz (Mendes de Oliveira et al. 2019).

Random Forest, in addition to the morphological information – encompassed in the parameters FWHM, major (A) and minor (B) semi-axes, and Kron radius (KrRadDet). In terms of the quasar classification, they achieved a completeness of 85.31% and a purity of 85.61% for the random forest. They have also shown that, when available, including the information from the $W1$ and $W2$ magnitudes from WISE improves considerably the completeness and purity of the quasar sample (95.96% and 96.33%, respectively).

Star, galaxy, quasar separation in miniJPAS

The results presented here were obtained in collaboration with a student from our group, Natália Villa Nova (Villa-Nova, M.Sc. dissertation, University of São Paulo,

2020).

In Villa Nova et al. 2021 (in prep.) we classify the miniJPAS sources using four supervised machine learning models: (i) CNN mag&err, (ii) CNN 1D Flux, (iii) RF and (iv) LGBM. These models differ from each other in the training strategy and/or input data. The first one is a convolutional neural network that receives as input the set of observed magnitudes modified to take into account the corresponding uncertainties (as illustrated in Fig. 3.26). The second one is also a convolutional neural network but that has as input the set of observed fluxes. The third one is a random forest using as input the magnitudes. Finally, the fourth is an implementation of gradient boosting decision trees, LightGBM⁹, for which the input data is the set of fluxes.

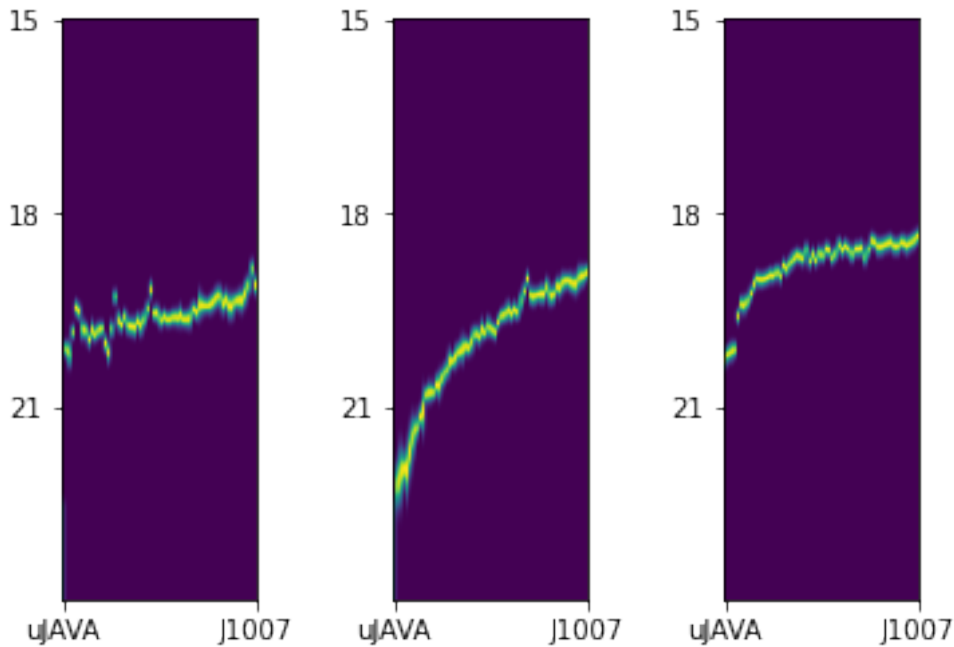


Figure 3.26: New representation of the miniJPAS photospectra for a quasar (left), galaxy (middle) and star (right) used as input for the CNN. The x -axis corresponds to the central wavelengths of the miniJPAS filters, and the y -axis corresponds to the magnitudes in the r -band.

The area imaged by the miniJPAS survey contains a wealth of multi-wavelength observations, especially spectroscopic data from SDSS and DEEP3 (Cooper et al. 2011a; Cooper et al. 2012a). However, the number of sources with trustworthy spectroscopic classifications is insufficient to be used for the training. The solution found by our group was to build mock catalogs for quasars, stars and galaxies following representative number densities in magnitude, redshift and/or stellar type (see Chapter 5 for more details).

⁹<https://lightgbm.readthedocs.io/en/latest/>

Furthermore, since we had real data observed in the J-PAS filter system, we could model more realistic errors directly from the miniJPAS catalog. We trained and tested the four machine learning methods with the mocks, and we also tested them with miniJPAS real observations (obtained by cross-matching the miniJPAS data set with SDSS DR16 Superset). For the classification, we are not considering the cross-matched sample with DEEP3 yet because, since this survey was designed to reach limiting magnitudes of $R \sim 24.1$, the miniJPAS cross-matched objects are typically noisier and their spectral DEEP3 classifications are less reliable.

The classification of the miniJPAS data is not only important for developing automated tools that will be later applied for the J-PAS data (once it is fully operational), but also to provide primary quasar targets to be observed by the WEAVE-QSO survey during its science verification program. Therefore, to attend the scientific goals of WEAVE-QSO, which focus on mapping the Ly α forest of high-redshift quasars, the objects were initially divided into four different classes: stars, galaxies, quasars 1 (at $z < 2.1$) and quasars 2 (at $z \geq 2.1$). We also considered four magnitude bins: $r < 18$, $18 \leq r < 20$, $20 \leq r < 22$ and $22 \leq r \leq 23.5$.

So far, the best performance was obtained for the CNN mag&err. In terms of the quasar classification with the mock test set, we obtained 96.7% (92.5%) and 99.0% (99.2%) for the completeness (purity) of the quasars 1 and 2, respectively. For the cross-matched sample, the performance decreases: we obtained 81.3% (82.4%) and 80.8% (72.4%) for the completeness (purity) of the quasars 1 and 2, respectively. This indicates that the error pattern observed in the real data was not properly modelled in the mocks. In particular, a correct assessment of correlations between contiguous filters and/or filters from the same tray for non-detections and non-observations, and the way that this is informed to the training sets could help us improve the classification with the real sources.

In Fig. 3.27 we show the confusion matrices for the four ‘‘classes’’ obtained with the test sample of the mocks (top panels) and miniJPAS observations cross-matched with Superset (bottom panels). On the left, we show the absolute numbers for each predicted label as a function of the true label; on the right, we show the fractions after normalizing by the total number of real objects of a given class (i.e., normalization by the sum of the

rows).

Currently, we are also combining these algorithms with other machine learning methods. So far, the results seem very promising because the combined algorithms always seem to outperform the individual performances in all magnitude ranges (see §5.4 for more details).

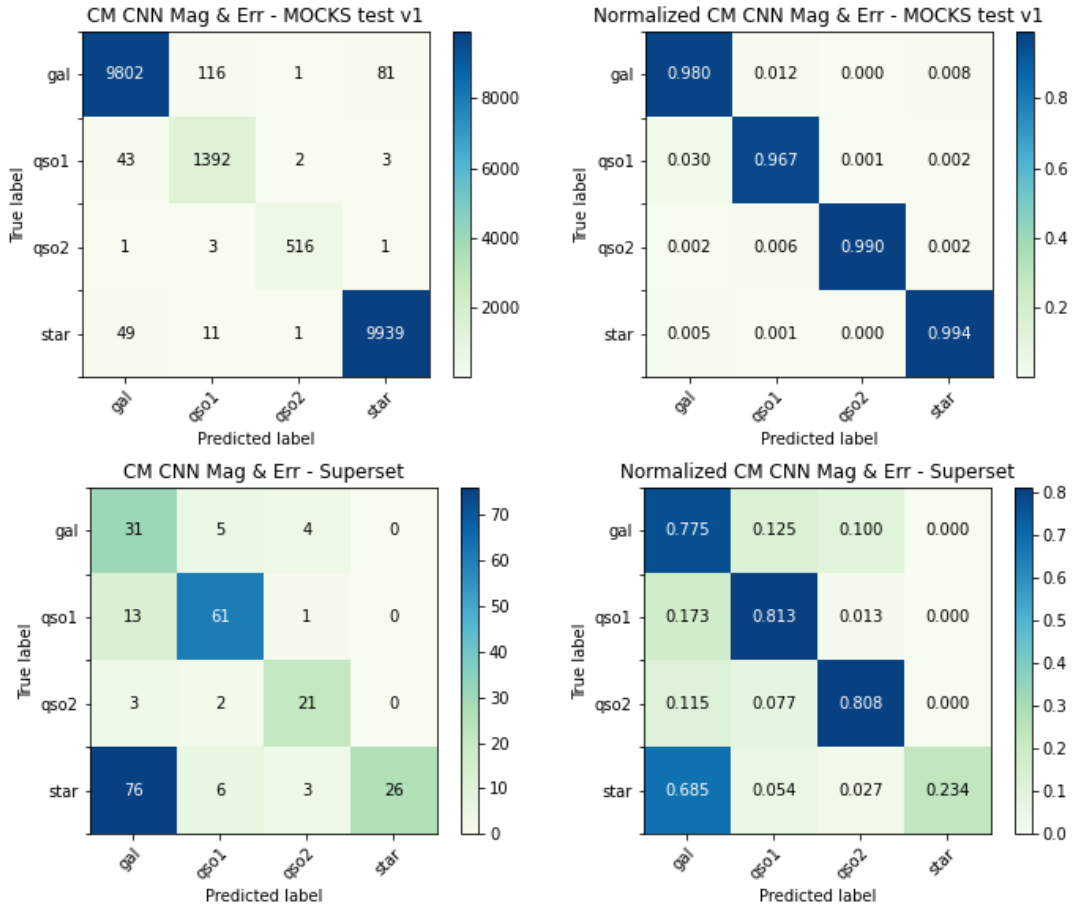


Figure 3.27: Confusion matrices obtained with the test set of the mocks (top panels) and the cross-matched miniJPAS data (bottom panels) and using CNN mag&err. Left panel: Absolute numbers of objects of a given predicted class as a function of the true label. Right: Normalized fractions for each class (Villa-Nova et al. in prep).

3.4 Final remarks on quasar classification

So far, we have talked about SMBHs and some evidences for their presence at the center of galaxies, and have shown how the galaxy “feels” their presence through the scaling relations. We presented the unified model of AGNs and described the main properties of the members of this family, focusing on the search for quasars and how we expect their number densities to evolve with redshift – as expected by the luminosity

function.

Nevertheless, the focus of this thesis is not the study of the physics of black holes: we are rather interested on their observations in the optical when they are in their active phase. AGNs are one the most powerful astrophysical events, and yet not so well-understood. How can quasars help us better constrain cosmological parameters and give us one more piece on the puzzle about the expansion of the Universe? How can the light of quasars indirectly tells us about the composition of the IGM? In order to answer these questions we must first succeed in correctly identifying them in multi-band photometric surveys (not an easy task, one must say).

Now, suppose that after applying a classification scheme we have identified a sample of quasar candidates with high confidence levels in a photometric survey: how can we properly estimate their distances, or in other words, their photometric redshifts? That is what we try to answer in the next chapter.

CHAPTER 4

PHOTOMETRIC REDSHIFT ESTIMATION

Measuring distances is extremely useful in astrophysics and cosmology: either to study the expansion history and growth of structures in the Universe, or to understand the physical mechanisms responsible for the formation and evolution of a given astronomical object.

For example, the distances to the nearest stars can be measured by parallax, which is the apparent shift in the stellar position due to Earth's orbit around the Sun. Parallaxes are useful to measure distances up to perhaps a few thousand light years. Beyond that, they are so small that they cannot be measured with present-day instruments. Another method is based on variable stars, such as Cepheids and RR Lyrae, which pulsate at a rate that depends solely on their absolute magnitudes. Since the period-luminosity relationship is well-known for these stars, once the pulsation period is obtained, the absolute magnitude is automatically determined, and hence by measuring the apparent magnitude one obtains its luminosity distance (which, by its turn, depends on the cosmology). Cepheids are considered standard candles and they can be used to measure distances to nearby galaxies.

Nevertheless, one of the easiest ways of measuring distances for galaxies (specially for those located at large distances) is based on the determination of their redshifts, specified by their spectra. Due to their quantum mechanical properties, atoms and molecules have characteristic spectral signatures at rest-frame frequency, which are assumed to be the same whether they are located at Earth or in outer space. This means that any emission and/or absorption features in the spectrum will inform us about the composition and motion of the source. Therefore, by looking at the shift of known spectral lines due to the

cosmic expansion, one can compute the so-called spectroscopic redshifts.

The Sloan Digital Sky Survey (SDSS) is doubtlessly the largest sky survey to-date. In twenty years of operation, it has obtained more than four million spectra (as of DR14), comprising galactic and extragalactic sources. Among other successful spectroscopic surveys, there are: 2dF (Colless et al. 2001), zCOSMOS (Lilly et al. 2007), 6dF (Jones et al. 2009), DEEP2/DEEP3 (Cooper et al. 2011b, Cooper et al. 2012b, Newman et al. 2013), WiggleZ (Parkinson et al. 2012), VIPERS (Garilli et al. 2014), MUSE (Urrutia et al. 2019) – just to cite a few. Upcoming and future multi-fiber surveys, such as Euclid (Lau-reijs et al. 2011, Amendola et al. 2013), PFS (Takada et al. 2014), Nancy Grace Roman Space Telescope (Spergel et al. 2015), DESI (DESI Collaboration et al. 2016, Levi et al. 2019), WEAVE (Dalton 2016, Pieri et al. 2016), and 4MOST (Richard et al. 2019), will expand the cosmic volumes probed with galaxy redshifts by several orders of magnitude. We summarize some of the upcoming ground-based multi-fiber facilities in Table 4.1.

Table 4.1: Summary of four spectroscopic survey facilities under construction. Mid-resolution is typically a few thousand; high-resolution is typically around 20K, but for a more limited bandpass. Adapted from: Levi et al. (2019).

Name	Telescope	Fibers	FOV (deg ²)	Bandpass (nm)	Resolution
PFS	Subaru 8-m	2400	1.5	380–1260	mid
DESI	Mayall 4-m	5000	8	360–980	mid
WEAVE	WHT 4-m	960	3	370–960	mid&high
4MOST	VISTA 4-m	2436	5	370–950	mid&high

In particular, the WEAVE survey is a multi-object spectrograph designed for the 4.2-m William Herschel Telescope (WHT) in the Canary Islands that will start observing by mid-2021. Part of the WEAVE strategy is to follow-up high-redshift ($z > 2.1$) quasars in order to conduct a Ly- α forest survey (WEAVE-QSO, Pieri et al. 2016). Dense spectroscopic samples of large-scale structure tracers are required to measure, e.g., the BAO feature accurately, but they can be observationally expensive. Mapping the Ly α forest is one way in which we can overcome this difficulty, using the fact that a single spectrum contains the signature of several intervening structures – but for that, we must be able to identify $z > 2.1$ quasars that can serve as background light sources to illuminate those

structures. The targets for the WEAVE-QSO survey will be provided by J-PAS, which is the only instrument in the world capable of identifying quasars in large enough numbers and down to the depths needed by WEAVE to do its science. This means that the task of identifying quasar targets from J-PAS photometry, as well as estimating their redshifts, is more pressing than ever.

One of the main drawbacks of spectroscopic surveys is the long integration times needed to resolve galaxy spectra sufficiently well to yield a good redshift estimate. Another limitation is the number of fibers that can be deployed in any single exposure and area, which is presently reaching $\sim 10^3$ fibers per deg^2 . Because of these factors, the number of targets selected for spectroscopic follow-up is usually much smaller than the total imaged sample, which significantly limits the survey depth and number densities that can be achieved.

In view of these difficulties, an alternative way of computing distances arises via photometric redshifts, which are estimated solely from the magnitudes and/or colors detected in a small number of wide frequency bands. Using photometry allows us to measure redshifts for larger samples of sources, in a much more efficient and faster way. In addition, imaging detectors can cover greater areas of the sky than multi-object spectrographs. For the same exposure times, photo-zs can be recovered for much fainter galaxies than spectroscopic redshifts, and so photometric surveys have also the ability to cover significantly larger volumes, making them potentially more suitable for constraining cosmological observables on ultra-large scales. Nevertheless, the accuracy of photometric redshifts is strongly dependent on (*i*) the number and the width of filters (e.g. broad, intermediate or narrow-band), as well as (*ii*) the redshift range and type of galaxies of interest (passive, star-forming or active). Since they provide low-resolution spectral information, photometric redshifts usually have much larger uncertainties than their spectroscopic counterparts.

For fields where extensive photometric data sets are available, the photometric redshift technique has been employed with reliable results. For instance, COMBO-17 (Wolf et al. 2004), CFHTLS (Ilbert et al. 2006), COSMOS (Scoville et al. 2007b), UKIDSS Ultra-Deep Survey (Williams et al. 2009), ALHAMBRA (Molino et al. 2014), Dark Energy Survey (DES; Hoyle et al. 2018), Javalambre Photometric Local Universe Survey

(J-PLUS; [Molino et al. 2019b](#)), Southern Photometric Local Universe Survey (S-PLUS; [Molino et al. 2019a](#), Lima et al. 2021 submitted), Physics of the Accelerating Universe Survey (PAUS; [Eriksen et al. 2020](#)). Yet, not all deep and wide fields have photometric redshifts of comparable accuracy, and above all, very few have tuned photometric redshifts for quasars.

Most of the current imaging surveys operate with photometric systems composed of a few (4-5) optical/NIR broad-band filters (i.e. widths $\sim 1000 \text{ \AA}$), which makes them more prone to color-redshift degeneracies (affecting, e.g., the ability to unambiguously determine the redshift for a galaxy, even if observed at relatively high S/N). Alternatively, in order to increase the spectral information, one could use more, narrower filters (with widths down to 100-200 \AA) but obviously at the expense of compromising significantly the photometric depth (at fixed exposure time). Nevertheless, “photometric depth is not equivalent to photometric redshift depth”, as shown by [Benítez et al. \(2009\)](#). In particular, they have demonstrated that for a larger number of non-overlapping filters (shown in Fig. 4.1) the effective photometric redshift depth actually decreases much more slowly with filter width than expected from the reduction in signal-to-noise. In addition, by decreasing the width of the passbands, it is possible to increase the redshift precision. These results are shown in Fig. 4.2. In fact, the photometric systems of surveys such as ALHAMBRA, J-PLUS, S-PLUS, and J-PAS were born from this recipe, and are a proof of this concept, allowing us to actually identify individual emission lines from galaxies and quasars (and, thus, not only recreate their continuum – as done by broad passbands).

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 must 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. Besides, 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.

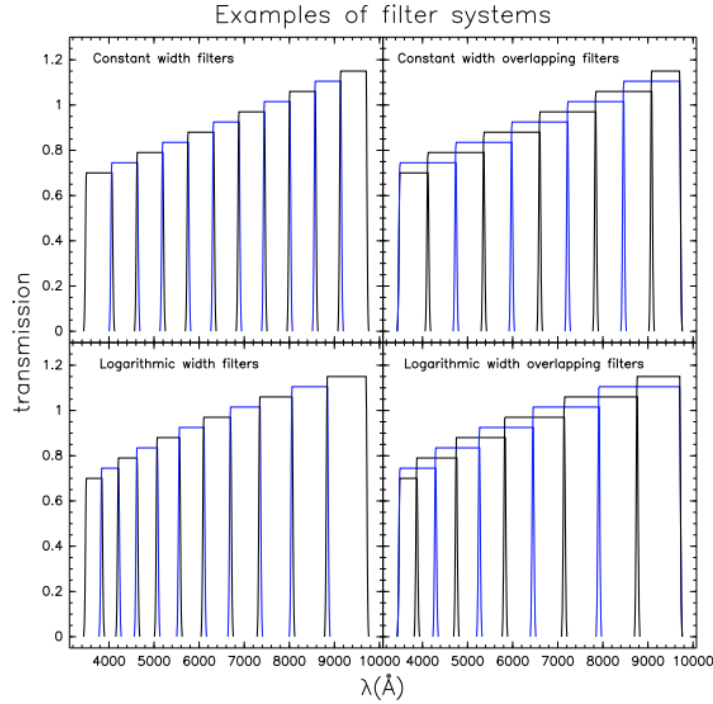


Figure 4.1: Example of four types of optimal filter sets, each with 11 filters. The filter transmissions are represented without taking into account the CCD or the telescope+optics transmission. The height of the filters were slightly increased with wavelength and alternate colors in successive filters were used to help visualization. Credit: [Benítez et al. \(2009\)](#).

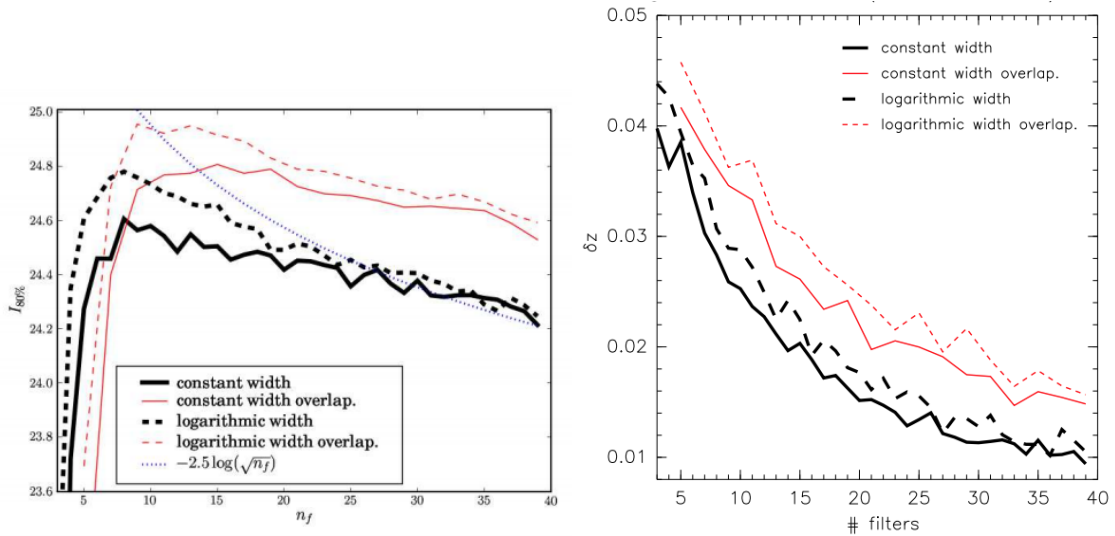


Figure 4.2: Left panel: Effective magnitude depths for different types of filter sets as a function of filter number for optical/NIR observations. They correspond to the magnitude at which the accumulated number of objects with Bayesian odds ≥ 0.99 is 80% of the total number of objects. The blue dotted line shows how the completeness magnitude would change with number of filters if it mimicked the behavior of the photometric limiting magnitude. Right panel: Redshift precision $\delta z = (z - z_{\text{best}})/(1 + z)$ for galaxies with odds ≥ 0.99 as a function of the number of filters for four types of filter systems. Credit: [Benítez et al. \(2009\)](#).

In this chapter we propose a method (dubbed QPz) to accurately estimate photometric redshifts for quasars. This method is inspired on a standard chi-square minimization: given a photospectra, it fits a model defined by a linear combination of quasar eigen-

spectra derived from a principal component analysis (PCA) and finds the full probability distribution function (PDF) for the redshifts. In order to validate the results obtained with our method, we compare its performance with a standard template-fitting code publicly available – LePhare.

4.1 Redshifts

Hubble’s law states that nearby galaxies in the local Universe move away from each other at a rate $cz = H_0 r$. As the physical separation of two fundamental observers is $a(t)dr$, Hubble’s law can be rewritten as

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (4.1)$$

At small cosmological distances one can define the redshift z in terms of the recession velocity as:

$$\frac{\lambda_{\text{obs}}}{\lambda_{\text{intrinsic}}} = 1 + z \approx 1 + \frac{v}{c}. \quad (4.2)$$

A more general expression for redshift comes when considering the photon’s null geodesic. For a radial geodesic, the metric yields:

$$r = \int \frac{dt}{a(t)}. \quad (4.3)$$

Since the comoving distance is constant for this case, the integral above leads to:

$$\frac{dt_{\text{intrinsic}}}{dt_{\text{obs}}} = \frac{a(t_{\text{intrinsic}})}{a(t_{\text{obs}})} \quad (4.4)$$

which means that there is a time-dilation for photons emitted from distant galaxies, which is proportional to the expansion of the Universe. This effect also appears on the observed wavelength, resulting in a more general expression for the redshift:

$$1 + z \equiv \frac{\lambda_{\text{obs}}}{\lambda_{\text{intrinsic}}} = \frac{1}{a(t)} \quad (4.5)$$

where $a(t = t_0) = 1$.

4.2 Photometric redshift methods

There are two classes of methods to compute photo-zs: template fitting (Loh & Spillar 1986) and empirical (originally proposed by Connolly et al. 1995). The first one uses a set of standard SED templates, either based on theoretical models (synthetic templates) or real data (empirical templates), which are supposed to represent all the possible spectral types observed (accounting for, e.g., different star-formation histories, chemical abundances, mixtures of dust and stars, luminosities) but should not be too large to avoid degeneracies. This class is, thus, very sensitive to template choice. The template-fitting methods map the templates in redshift space, using small dz steps. Then, through a chi-square minimization, they compare the theoretical fluxes from the redshifted templates with the observed fluxes until finding the best correspondence – both in type and redshift. Another issue in this class of methods is that they are strongly affected by systematic errors in the observables, such as differences in zero-point calibrations, especially if they are wavelength dependent. Some examples of algorithms that implement this method are: LePhare (Arnouts et al. 1999, Ilbert et al. 2006), BPZ (Benítez 2000), HyperZ (Bolzonella et al. 2000), ZEBRA (Feldmann et al. 2006), and EAZY (Brammer et al. 2008).

The other class relies on a “true” sample of sources for which it is known not only their magnitudes or colors, but also their true spectroscopic redshifts. This sample, dubbed training set, is employed to feed the training procedure and determine a functional relationship between colors and redshifts, which will then be applied to the photometric sources from the sample of interest (the so-called test set) to estimate their photometric redshifts. Therefore, to ensure that the empirical relation will be able to reproduce most of the spectral variations found in the test set, the training set should ideally contain a large variety of spectral properties, different ranges of luminosity, and realistic frequencies for each spectral type in each redshift bin. In addition, a new empirical relation must be derived for each photometric system configuration. In this class are included polynomial fits, nearest neighbors, Gaussian processes, neural networks, random forest, boosted decision trees, self-organizing maps, etc. Some examples of algorithms found in the literature are: ANNz (Collister & Lahav 2004b), ANNz2 (Sadeh et al. 2016), Weights (Lima

et al. 2008), ArborZ (Gerdes et al. 2010), Multilayer Perceptron Artificial Neural Network (Vanzella et al. 2004), and SOMz (Carrasco Kind & Brunner 2014).

Empirical methods usually outperform template methods when a representative training set is available (e.g. Abdalla et al. 2011). However, template fitting methods can extrapolate better at redshift bins or luminosity ranges where there is a lack of training objects. Hence, in the future, hybrid methods that combine the main characteristics of each technique may be preferable to ensure more precise photo-zs.

There are also some methods tuned specifically to identify and estimate redshifts for quasars. We discuss two of them in the following.

The first aims to identify emission lines for quasars in medium- and narrow-band filter surveys. The Emission Line Detector of Astrophysical Radiators (ELDAR; Chaves-Montero et al. 2017) adopts a two-step procedure to search for quasars in the ALHAMBRA fields. First, it employs a template fitting method (LePhare) to pre-select quasar candidates and obtain a redshift probability distribution (PDZ) for them. The second step is a spectro-photometric confirmation step, in which it detects the main quasar emission lines and refines the photo-z estimation. This allows the rejection of galaxies and stars that were wrongly assigned to quasar templates and also to discriminate between multiple solutions in the PDZ. For a magnitude limited sample of $F814W = 23$, they estimate a completeness of 67%, a redshift precision of 0.86% and an outlier fraction of 5.8%.

The second implements a random forest algorithm to classify quasar optical spectra and estimate their redshifts. The Spectroscopic QUasar Extractor and redshift (z) Estimator (SQUEzE; Pérez-Ràfols et al. 2019) mimics the process of human inspection to identify all significant emission lines. To each peak, a different redshift is assigned (z_{try}) and then passed to a random forest classifier, which will assess the quality of this trial. This allows one to recover redshifts from the SDSS DR12 with a purity and completeness of 99.59% and 98.81%, respectively, for a high redshift subsample ($z_{\text{try}} > 2.1$). In a second paper, Pérez-Ràfols & Pieri assess the performance of SQUEzE when low-resolution spectra (pixels of 100 Å of width) are considered – hence, a forecast for JPAS-like data. In this case, they show that the purity remains unchanged for the high-z sample, while the completeness decreases only $\sim 1\%$ when compared to the original sample. This means

that SQUEzE performs satisfactory not only on spectra from, e.g., DESI and WEAVE, but also on J-PAS photospectra.

In the following sections, we describe the configuration of the photo-z methods employed in this thesis.

4.3 LePhare++

LePhare performs SED template fitting through a simple chi-square minimization. Additional extinction laws can also be applied to the templates. The publicly available version of this code is written in FORTRAN language which only allows for a maximal size to be allocated to the arrays. In practice, this means that we cannot estimate photo-zs for a photometric system containing more than 45 passbands and considering redshift steps as small as $\Delta z \sim 0.003$ (i.e., of order of the precision that we expect to obtain with J-PAS). For this reason, in our analyses we employed a new version, LePhare++ (O. Ilbert and S. Arnouts), which is written in C++ and is not publicly available yet.

To build the quasar library, we started from two lists previously compiled for the COSMOS field (Salvato et al. 2009) and the ALHAMBRA fields (Chaves-Montero et al. 2017). These libraries served only as an initial guess and were successively refined to properly fit our data sets (composed of spectroscopically confirmed quasars detected with miniJPAS and S-PLUS DR1), in such a way that we guaranteed an increase in the redshift precision (σ_{nmad}) and a decrease in the outlier fraction (η) – see §4.5 for more details on how these quantities were computed. First, all normal galaxy templates and SEDs that did not fit any quasar from our sample were rejected. Then, we adopted a redshift precision of $\Delta z = 0.003$ (since values smaller than this generated a segmentation fault). We also applied a luminosity prior of $M_B < -20$ to avoid unrealistic solutions at low z – and this modification was invaluable to improve the performance of the photo-z estimation. A correct modelling of dust attenuation (A_V) is also required; in the case of quasars, we expect it to be dependent on both redshift and orientation with respect to the observer. We tested different extinction laws, and the starburst model from Calzetti et al. (2000) provided the best performance (we also tested combinations of more than one extinction law, but these also generated a segmentation fault). The intrinsic galactic absorption was

computed with values $E(B - V) = 0, 0.025, 0.05$.

Finally, we used the diagnostic broad-band color-redshift plots from Richards et al. (2001) to obtain our final library. In Fig. 4.3 we show a comparison between the broad-band colors ($g - r$ and $r - i$) of the templates shifted up to redshift $z = 5$ (colored lines), and the distribution of quasars that were spectroscopically confirmed by SDSS DR16 Superset (gray dots). In this figure, we also show the locus occupied by the miniJPAS quasars (red dots).

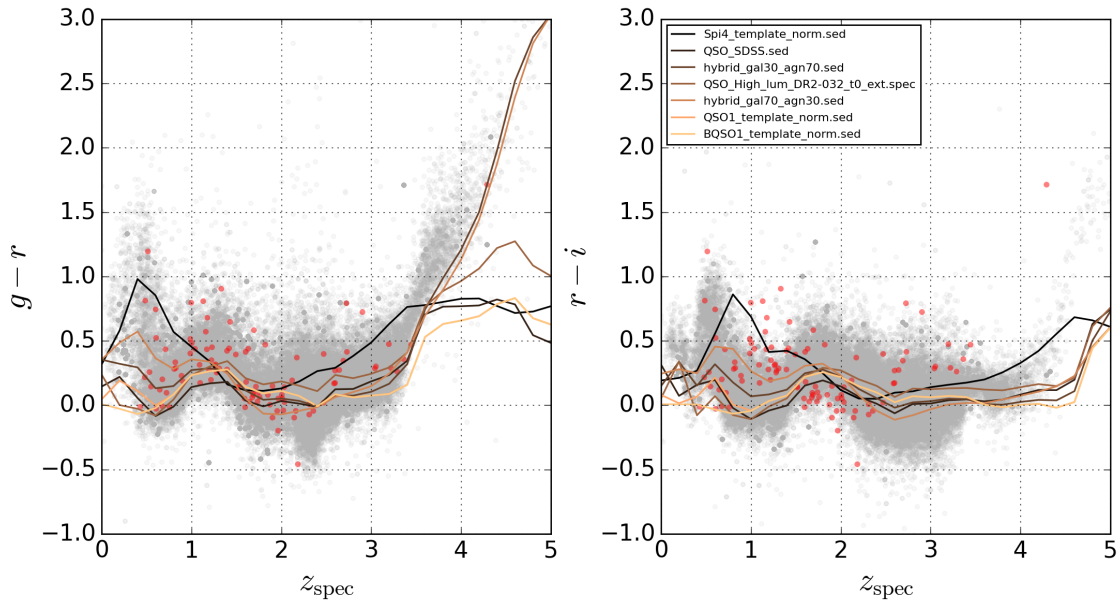


Figure 4.3: Color-redshift distribution for the final sample of 7 templates used in LePhare++. We compare these distributions with that of spectroscopically confirmed quasars from SDSS DR16 Superset (gray dots) and miniJPAS quasars (red dots).

Throughout the procedure of building the quasar library, we were aware that SED templates can often be combined in an arbitrary manner that may not be physically realistic in terms of the properties that would be inferred for the source. For this reason, we tested different template orders and tried to include twice the same template in the list, but without any improvements. The final configuration that we obtained (shown in Fig. 4.3) seems to be the one that best describes our data sets¹.

There are still some internal issues that should be resolved with LePhare++ (especially for S-PLUS), but the overall performance is already very satisfactory for miniJPAS – as we show in Chapter 6.

¹The same library was applied for both miniJPAS and S-PLUS data.

4.4 Quasar Photometric Redshifts (QPz)

In the UV-optical wavelengths, most quasar spectra are characterized by a non-thermal continuum, a blue bump and a series of (mostly) broad emission lines. We know, however, that their spectra are far from being uniform: the spectral slope, as well as the emission lines, can differ a lot among quasars, and even within a single spectrum, the widths of the emission lines can actually be remarkably different. If, on one hand, these differences provide invaluable insights about the physical environments in the vicinity of quasars, on the other hand, they present significant challenges to the modelling of broad and narrow-line regions.

A useful tool to quantitatively understand the variation in quasar spectra and determine the minimum number of dimensions required to describe the multidimensional distribution of variables (which, in this context, are represented by the observed flux densities in the wavelength bins of the spectrum) is the principal component analysis (PCA; [Chatfield & Collins 1980](#)). The PCA employs an orthogonal transformation to model the fluxes of real astrophysical objects to find a set of linearly uncorrelated vectors – the so-called principal components. These principal components (hereafter eigenspectra) are hierarchically organized in such a way that the first component accounts for the largest possible variance, and they form the most meaningful basis that represents the essential physical properties of the data sample.

For example, [Boroson & Green \(1992\)](#) applied this technique to study the properties of various emission lines in the rest-frame optical spectra of quasars. [Francis et al. \(1992\)](#) were the first to apply PCA to the spectral pixels themselves, where they found that the mean spectrum plus the first two principal components are able to describe the majority of the variations seen in the UV-optical spectra of quasars. [Yip et al. \(2004\)](#) applied a variant of the PCA technique, the Karhunen-Loève (KL) transform, to find the spectral classification for a sample of 16 707 quasar spectra from the SDSS early DR1 ([Abazajian et al. 2003](#), [Schneider et al. 2003](#)). They found that the spectral classification of quasars depends both on redshift and luminosity, since it was not possible to derive a single compact set (with $\lesssim 10$ modes) of eigenspectra capable of describing most variations of the

entire catalog. In 2005, [Suzuki et al.](#) introduced the idea of using PCA to study the continuum of low-redshift ($z \sim 0.14 - 1.04$) quasars and construct a model to predict the intrinsic continuum of Ly α absorbed regions in the spectra of high-redshift quasars. This approach was extensively used in other works to predict the blue-side quasar continuum from the red-side spectrum (e.g. [Pâris et al. 2011](#), [Lee et al. 2012](#), [Lee et al. 2013](#), [Davies et al. 2018](#)).

In this work, we applied the quasar eigenspectra derived by [Yip et al. \(2004\)](#) (Yip04 henceforth) to develop a method of photometric redshift estimation for quasars detected in multi-band surveys. This idea of fitting spectra with a linear combination of redshifted template “eigenspectra” in combination with a low-order polynomial (to account for Galactic extinction) is not new. [Bolton et al. \(2012\)](#) applied this approach to construct an automated pipeline to classify and estimate redshifts for stars, galaxies and quasars detected in the SDSS DR9 ([Ahn et al. 2012](#)). However, as far as we know, this is the first time that it is applied to fit *quasar photospectra*.

One stringent assumption of the PCA technique is the linearity. This means that the data set is re-expressed as a linear combination of its basis vectors, and this is the same underlying assumption in the photo- z estimation. Obviously we are applying a linear approach to a non-linear problem; still, our results are very robust and reveal a significant lower fraction of catastrophic errors (i.e., outliers) when compared to those obtained with a standard template-fitting method (LePhare).

In the following we describe how this method was implemented.

4.4.1 Global eigenspectra

The redshifts of the quasar sample used to derive the eigenspectra range from 0.08 to 5.41, the i -band absolute magnitude from -30 to -22, and the rest-frame wavelengths from 900 Å to 8000 Å. All magnitudes were corrected for Galactic extinction using dust maps from [Schlegel et al. \(1998\)](#). The signal-to-noise ratios of the spectra generally meet the requirement of $(S/N)^2 = 15$ per spectroscopic pixel ([Stoughton et al. 2002](#)). Finally, at least one prominent line in each spectrum from the DR1 quasar catalog has full width at half maximum FWHM $\geq 1000 \text{ km s}^{-1}$, such that type II quasars and BL Lacs are not

included in the quasar catalog.

Different rest-frame wavelength coverage, removal of skylines, and bad pixels on the CCD chips can lead to incomplete spectra. In order to correct for the missing pixels, Yip04 initially repaired the missing data with a polynomial interpolation, constructed a set of eigenspectra from the gap-repaired spectra, and then corrected the gaps in the original quasar spectra with a combination of these eigenspectra. This procedure was iterated until convergence (~ 10 steps).

Yip04 found that in order to account for approximately 99% of the total sample variance, about 50-60 modes were required. In their work, they derived 100 modes. In Table 4.2 we show the partial sum of weights (i.e., cumulative weights) in different orders of the global² eigenspectra. Note that here these weights are equivalent to the eigenvalues of the correlation matrix.

Table 4.2: Partial sum of weights of the global quasar eigenspectra. The information of w_{100} is not available. Adapted from: Yip et al. (2004).

Number of modes	Weights (w_n)	Partial sum
1	0.5609	0.5609
2	0.1193	0.6802
3	0.0758	0.7560
4	0.0614	0.8174
5	0.0469	0.8643
10	0.0139	0.9194
20	0.0032	0.9689
50	0.0009	0.9957
100	–	0.9992

This means that the first eigenspectrum accounts for about 56% of the total sample variance, while the first 10 modes account for about 92%. In Fig. 4.4 we show the first four modes at rest-frame and the main quasar emission lines on the top panel.

One of the most interesting characteristics of employing eigenspectra instead of

²We will denote by global eigenspectra those modes constructed from the whole sample as a way of distinguishing from the sets of eigenspectra constructed in specific redshift and luminosity bins.

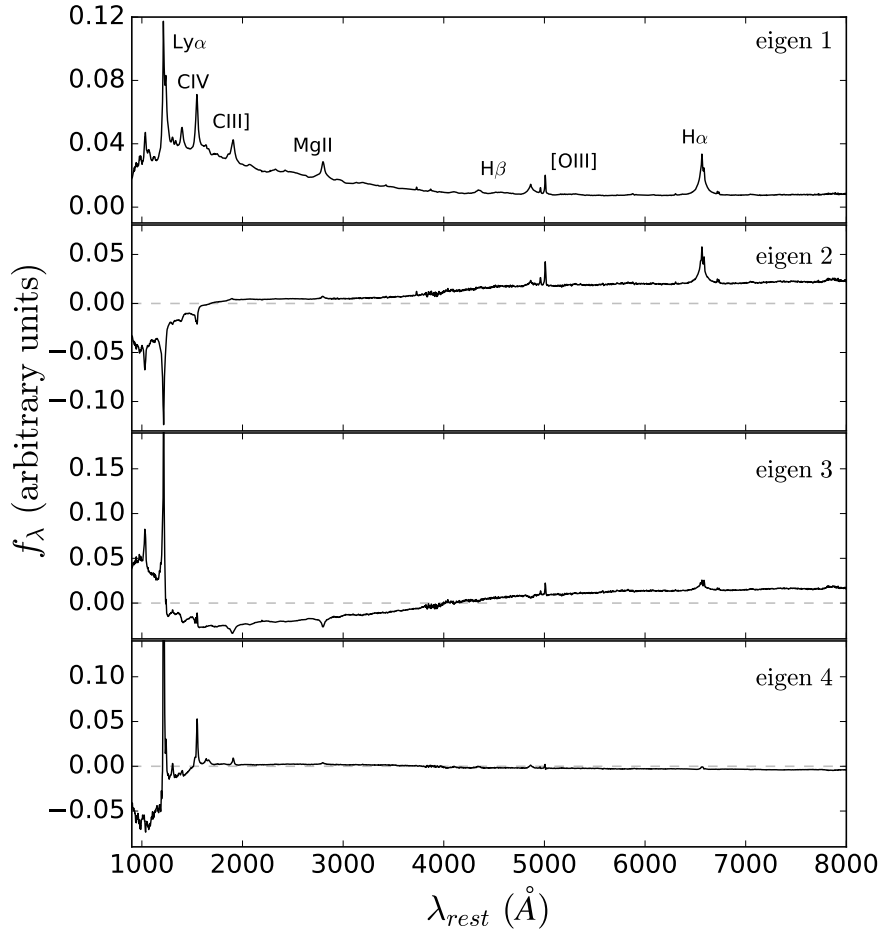


Figure 4.4: First four global quasar eigenspectra at rest-frame. The lowest-order component (eigen 1) is the most important one: it represents the mean quasar spectrum and, therefore, it carries all the main type-I emission lines. The fluxes were normalized by the first mode.

templates is that each mode can be associated to a different physical property of the quasars. For example, the first four quasar eigenspectra reveal the following features:

- 1st mode: it represents a mean quasar spectrum. It has approximately a power-law shape with prominent broad emission lines;
- 2nd mode: it represents the change in the spectral slope, to account for the host-galaxy contribution;
- 3rd mode: it carries the information from the UV-optical continuum slope. It is very similar to the host contribution;
- 4th mode: it shows the correlations between broad emission lines, $\text{Ly}\alpha$, CIV, SiIV+OIV], CIII], MgII, [OIII] λ 5008, and also the Balmer emission lines, $\text{H}\alpha$, $\text{H}\beta$, $\text{H}\gamma$, $\text{H}\delta$ and $\text{H}\epsilon$.

By construction, higher-order components account for small modulations on the continuum slope and can also show broad absorption lines. This means that the signatures of BAL quasars (which are not dominant in the sample) will show up preferentially at higher-orders in the global eigenspectra.

Quasars with prominent narrow-lines usually have low luminosities, in which case the contamination from the host galaxy can obscure the true physical nature of these lines. Note, however, that since we can detect significant features from the host, this could be used as an application to remove its contribution.

Given a photospectrum of a source q , characterized by a set of observed fluxes and corresponding uncertainties $\{f_{q,\mu}; \sigma_{q,\mu}\}$, our quasar photometric redshift estimator, QPz, utilizes a chi-square minimization

$$\chi_q^2(z) = \sum_{\mu} \frac{[f_{q,\mu} - \mathcal{F}_{\mu}(z)]^2}{(\sigma_{q,\mu}^2 + \sigma_{\text{Ly}\alpha}^2)} \quad (4.6)$$

to find the best-fit model $\mathcal{F}(z)$ at each redshift from a linear combination of redshifted eigenspectra, $\xi_{\mu}^n(z)$, convolved with the corresponding photometric band, μ , and combined with an extinction power law:

$$\mathcal{F}_{\mu}(z) = \sum_n c_n \xi_{\mu}^n(z) \left(\frac{\lambda_{\mu}}{\lambda_{\text{eff}}} \right)^{-\alpha} \quad (4.7)$$

where n is the mode number, c_n are the eigencoefficients allowed to vary in the range $\pm 3w_n$, λ_{μ} is the filter effective wavelength, $\lambda_{\text{eff}} = 5000 \text{ \AA}$ is a scale wavelength that marks the modulation of the slope between the blue and red sides, and α is the power-law index allowed to vary in the range $[-1.5, 1.5]$. In Eq. 4.6, $\sigma_{\text{Ly}\alpha}$ is an additional variance of 2.5% of the flux for the filters located on the blue side of the Ly α line.

In order to shift the eigenspectra, we first extend the blue and red sides by assuming a constant flux (of course, this is just a first-order approximation). Then the redshift is allowed to vary up to 5.13 in steps of $\Delta z = 0.001$. This choice of z_{max} was made to match the limiting redshift in which the redshift-luminosity eigenspectra are defined; it is arbitrary, though, and can be adapted accordingly to the properties of the sample that we want to test. Note, however, that at $z \gtrsim 5.2$ the Ly α line is in the transition between the

red and NIR part of the spectrum, which means that they are not any significant emission lines left to be identified with this method.

For each source q , the method determines the set of parameters $\{c_n, \alpha\}$ that minimize the chi-square function in Eq. 4.6 at each redshift. In the end, it provides the full redshift probability distribution:

$$\mathcal{P}^q(z) \sim \exp \left[-\frac{\chi_q^2(z) - \min(\chi_q^2)}{2} \right]. \quad (4.8)$$

The value of redshift that globally maximizes \mathcal{P}^q corresponds to the photometric redshift and will be denoted by z_{phot} . Although most metrics utilize these single point-estimates to assess the quality of the photometric estimation, it is important to bear in mind that the full PDF actually incorporates the complete information about the uncertainty in the photo-z, being preferable in any cosmological analysis.

Also, although in the beginning we assume that the photospectrum corresponds to a quasar, by construction this method will find a best-fit for the photospectrum of any source – even if it lacks any physical meaning.

For the implementation of QPz, we use a suitable tool for minimization problems: the Python module `scipy.optimize.least_squares`³ with the trust region reflective (trf) algorithm, a robust approach that quickly moves to another region in the parameter space whenever a negative curvature is encountered (i.e., whenever the local minimum is far from the solution).

4.4.2 Number of eigenspectra

As can be expected, increasing the number of eigenspectra used for modeling a photospectrum increases not only the completeness (understood as the ability to correctly fit more spectral features) but mainly the impurity of the resulting fit because it increases the possibility of finding spurious results. This is equivalent to what happens when we increase too much the number of templates in a template-fitting code.

In order to determine the most suitable number of eigenspectra, we tested four

³https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html

different sets: with 4, 5, 10 and 50 modes. We have not considered less than 4 modes to ensure that the effect of correlation between Balmer emission lines is included in the resulting model; on the other hand, 50 was the maximum number of eigenspectra that we had access to. We tested these different sets with 124 spectroscopically confirmed quasars detected by the miniJPAS and compared their performances with regard to the redshift error (in terms of σ_{nmad}) and the fraction of outliers (see §4.5 for more details on how we computed these metrics). The performances are illustrated in Fig. 4.5, and the quality of the photo-z estimation is shown in Table 4.3.

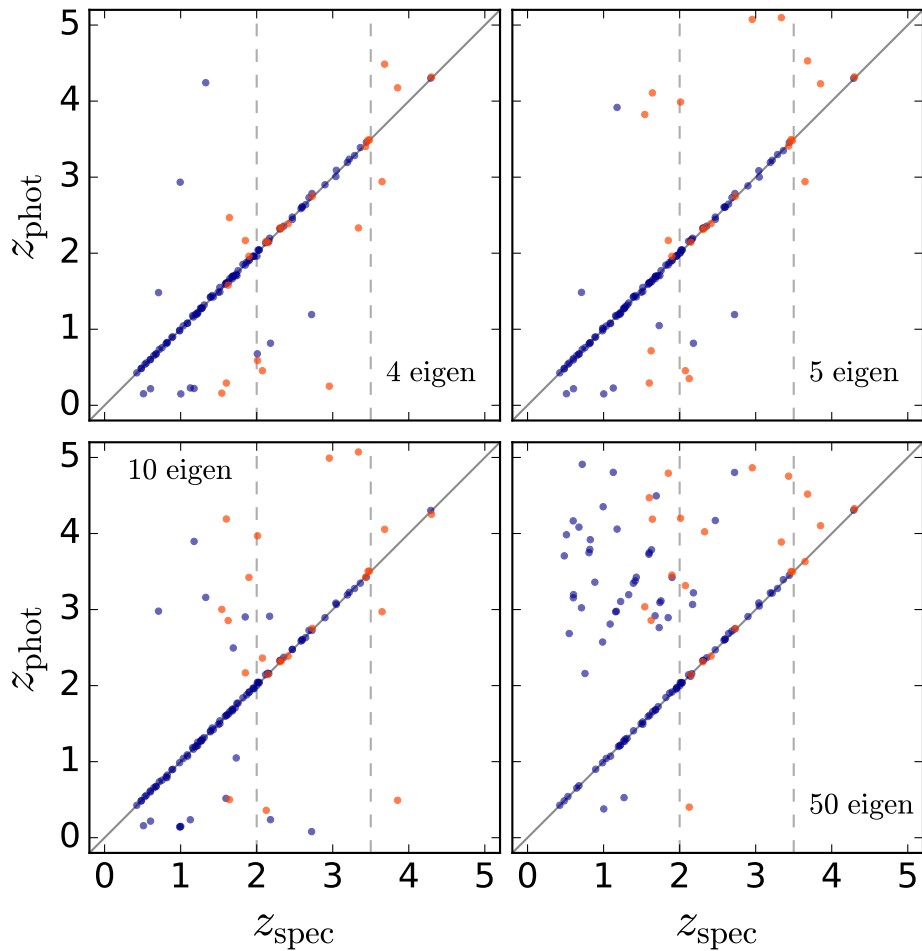


Figure 4.5: Comparison of the performance for different numbers of eigenspectra: estimated photometric redshift as a function of the spectroscopic redshift for spectroscopically confirmed quasars from miniJPAS. The blue dots correspond to quasars from Superset (101) and the orange dots to quasars from DEEP3 (23). The diagonal line indicates $z_{\text{phot}} = z_{\text{spec}}$. The vertical gray dashed lines at $z_{\text{spec}} = 2$ and $z_{\text{spec}} = 3.5$ illustrate the redshift range where we expect to obtain the most precise photo-zs since there are several quasar emission lines entering the filters.

From Table 4.3 we can see that running QPz with a set of 4 eigenspectra provides the best performance ($\sigma_{\text{nmad}} = 0.0057$, $\eta = 17.7\%$). Also, as already expected, we found a tendency towards an increase in the outlier fraction in all redshift ranges as the

number of modes increased. Although we are showing here the results for a small sample of quasars, which is more subjected to statistical fluctuations, we verified that this same tendency remained even with larger samples of quasars and, hence, all our results are shown in terms of a linear combination of the first 4 eigenspectra.

Table 4.3: σ_{nmad} , bias and outlier fraction obtained with QPz using different numbers of modes and for a sample of 124 quasars from miniJPAS.

Number of modes	σ_{nmad}	δz	Outlier fraction (η)
4	0.0057	-0.021	17.74%
5	0.0070	-0.01	17.74%
10	0.0067	-0.023	24.19%
50	0.0240	-0.409	45.97%

4.4.3 Luminosity-redshift dependency

Yip04 also investigated a possible redshift evolution and luminosity effects in the quasar spectra. For this, they divided their sample into five redshift bins: $z_1 = [0.08, 0.53]$, $z_2 = [0.53, 1.16]$, $z_3 = [1.16, 2.06]$, $z_4 = [2.06, 3.33]$, and $z_5 = [3.33, 5.13]$, and four luminosity (absolute magnitude in the i -band, M_i) bins: A = $(-30, -28)$, B = $(-28, -26)$, C = $(-26, -24)$, and D = $(-24, -22)$.

They noted that the first modes constructed in all these subsamples were always very similar to each other (with more than 99% of commonality), which demonstrated that a single mean spectrum can be constructed across the whole redshift coverage. This is not a trivial finding because there is always the possibility of having an evolving quasar population at different cosmic epochs. However, as higher orders were considered, the commonality leveled off and the eigenspectra subspaces became more disjoint. For 20 modes or higher, the resulting eigenspectra contained mainly noise. This means that eigenspectra of same order but derived from quasars of different redshift and luminosity bins describe different spectral features. Or, in other words, eigenspectra derived from quasars of a particular redshift and luminosity bin in general are not able to predict quasar spectra of other redshifts and luminosities.

Taking into account this redshift-luminosity dependency, we also implemented a second version of QPz in which, instead of using the global eigenspectra, we used the first four eigenspectra derived in (M_i, z) -bins.

The binned eigenspectra were identified to represent the same physical interpretations as in the set of global modes, except maybe for the third one, which is now specifically related to the anti-correlation between FeII (UV) and optical continuum around $H\beta$. In Fig. 4.6 we show the first four eigenspectra in the 12 different redshift-luminosity bins.

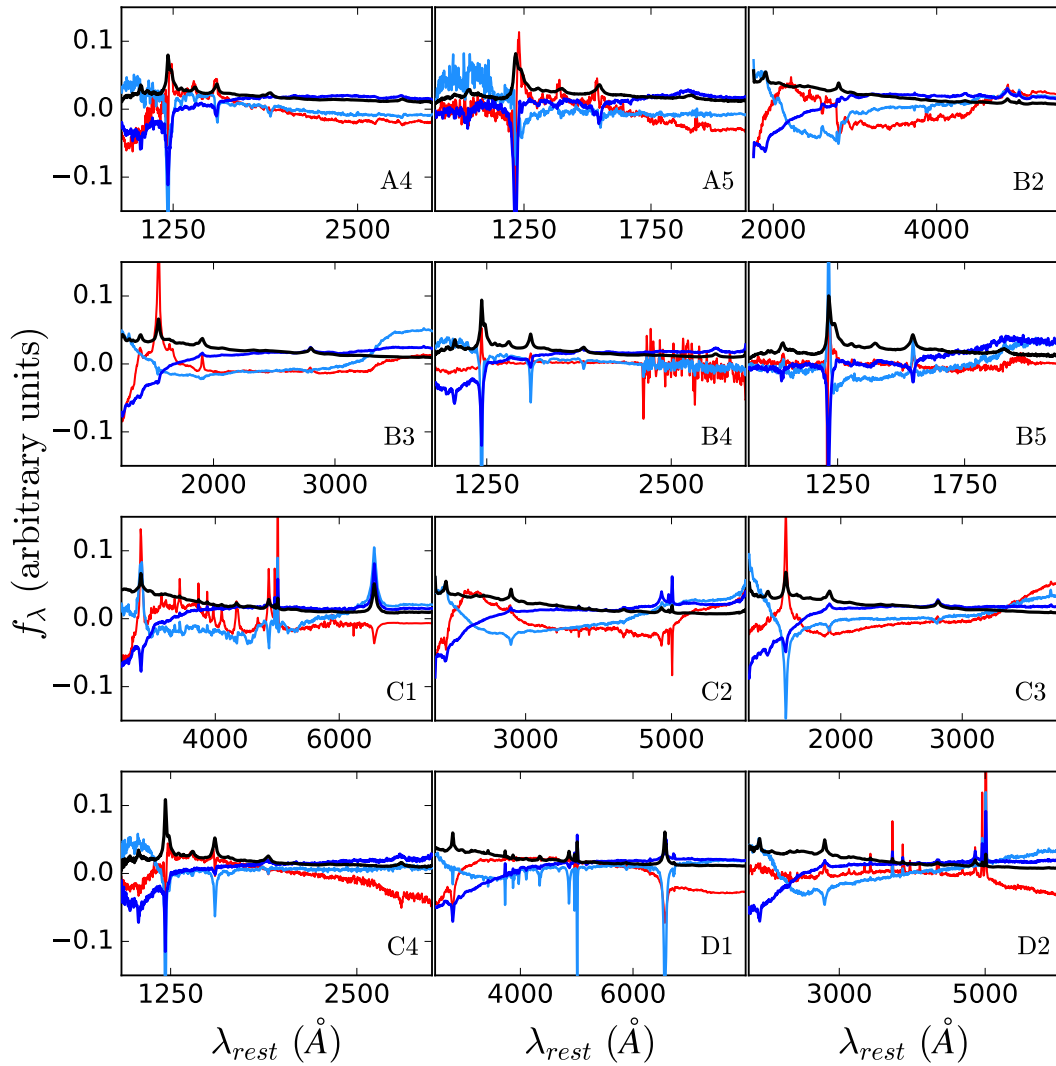


Figure 4.6: Quasar eigenspectra in the 12 redshift-luminosity bins (different panels). The 1st, 2nd, 3rd and 4th modes correspond, respectively, to the black, dark blue, light blue and red lines. In each bin, the eigenspectra span different wavelength coverages and have different weights. The fluxes were normalized by the first mode.

4.4.4 Not only photospectra

Although designed to model the photospectra for quasars detected in multi-band surveys and estimate their redshifts, QPz is flexible enough to model high-resolution *spectra* as well. To test this, we considered the spectra of the same SDSS quasars detected by miniJPAS (there are 101) and mimicked their spectral pixels as very narrow bands (of a few Å-width). Now, convolving the global eigenspectra with these new “filters”, we obtained a new set of eigenspectra, and then employed these to model the SDSS spectra. Not surprisingly, we were able to recover their redshifts with an extremely high precision ($\sigma_{\text{nmad}} = 0.00067$), without any outliers. These results are shown in Fig. 4.7.

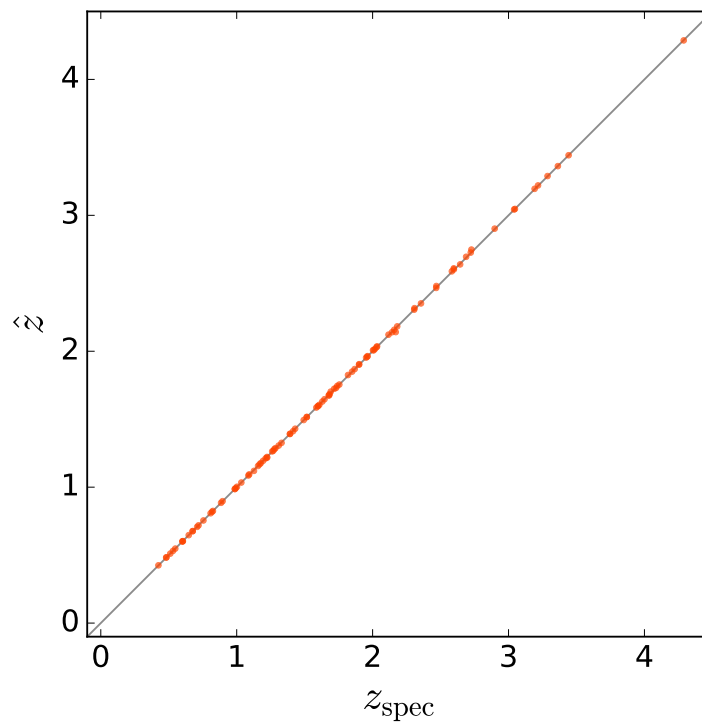


Figure 4.7: QPz redshift estimation for 101 quasar spectra from SDSS DR16. The diagonal line indicates where the estimated redshift \hat{z} is equal to the spectroscopic redshift.

4.4.5 Type-II quasars

Since, in principle, Yip04 did not include quasars with exclusively narrow emission lines in the sample utilized to derive the eigenspectra, we also tested if QPz would be able to recover the photo-zs of type-II quasars. We did not find any spectroscopically confirmed type-II quasar in the miniJPAS area, so for this exercise we employed synthetic fluxes generated from SDSS spectra. In Fig. 4.8 we show two examples of simulated

type-II quasars at redshifts $z_{\text{spec}} = 2.400$ (left panel) and $z_{\text{spec}} = 3.324$ (right panel). QPz is able to recover the redshifts for both objects with high accuracy: $z_{\text{spec}} = 2.402$ and $z_{\text{spec}} = 3.336$, respectively.

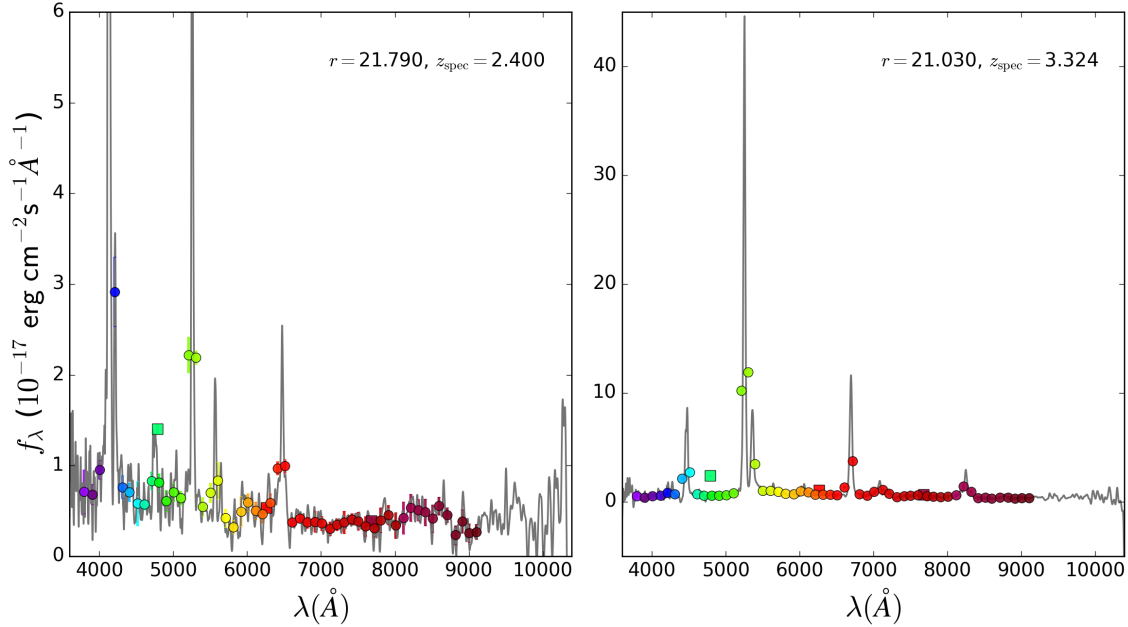


Figure 4.8: QPz redshift estimation for type-II quasars at different redshifts. The gray line corresponds to the SDSS spectrum and the colored dots correspond to the synthetic fluxes in the miniJPAS passbands. The redshift values estimated with QPz are $z = 2.402$ (left panel) and $z = 3.336$ (right panel).

This demonstrates the flexibility of QPz, and also the power of the J-PAS photometric system of detecting narrow emission lines.

4.5 Photo-z quality

In this section, we present the metrics adopted to assess the quality of the photo-z estimates.

4.5.1 Normalized Median Absolute Deviation (NMAD)

In the standard deviation, the distances from the mean are squared; this means that large deviations are more heavily weighted, being more affected by catastrophic results (i.e., outliers). The normalized median absolute deviation (NMAD), on the other hand, is a more robust estimator and more resilient to the influence of outliers. It is defined as (e.g., [Matute et al. 2012](#)):

$$\sigma_{\text{nmad}} = 1.48 \times \text{median} \left| \frac{\Delta z}{1 + z_{\text{spec}}} \right| \quad (4.9)$$

where $\Delta z = z_{\text{phot}} - z_{\text{spec}}$.

Note that the standard deviation alone, given by $\delta_z = \Delta z / (1 + z_{\text{spec}})$, already constitutes a way of estimating the accuracy of the measure. However, σ_{nmad} takes into account the fluctuations around the diagonal line where $z_{\text{phot}} = z_{\text{spec}}$ and, therefore, is less sensitive to catastrophic errors. In the limiting case of a null fraction of outliers, one would expect $\sigma_{\text{nmad}} = \delta_z$.

We also considered the scatter σ_N , which corresponds to the half of the width of the Δz distribution where $N\%$ ($N = 50, 75$) of the data is enclosed.

4.5.2 Bias

The bias b_z gives information about systematic errors in the photo-z estimates, and it is defined as:

$$b_z = \langle z_{\text{phot}} - z_{\text{spec}} \rangle. \quad (4.10)$$

Note that b_z is expected to be positive when the method overestimates the photo-z, and negative when the photo-z is underestimated. Ideally, a method with a high performance should provide $b_z \sim 0$ for the most part of the redshift coverage of the data sample. Therefore, the bias value is more enlightening when evaluated as a function of the redshift bin.

4.5.3 Outlier fraction

The outlier fraction η is related to catastrophic errors (led by spurious solutions) and, in general, is defined in terms of some factor of σ_{nmad} . In our analysis, we define it as the fraction of objects that satisfies the condition $|\Delta z / (1 + z_{\text{spec}})| > 0.05$.

4.5.4 Odds

One further quality parameter is the odds, which in general can be defined as the redshift probability integrated within an interval Δz around the maximum likelihood value:

$$odds = \frac{\int_{z_{peak}-\Delta z}^{z_{peak}+\Delta z} \mathcal{P}(z) dz}{\int_{z_{min}}^{z_{max}} \mathcal{P}(z) dz} \quad (4.11)$$

where z_{peak} is the redshift corresponding to the maximum likelihood peak, z_{min} and z_{max} define the redshift range considered in the photo-z estimation, and Δz depends on the survey. Narrow distributions result in values close to 1 since most of their integrated probabilities will be contained within Δz . On the other hand, broad or multi-peaked distributions will result in values close to 0 since, in this case, the integrated fraction will be very small. Hence, the higher the odds, the more precise the photo-z estimation and the smaller the redshift dispersion δ_z .

4.5.5 Redshift-magnitude priors

Using the magnitude-redshift distribution of the spectroscopic sample as a prior information is believed to be extremely powerful to remove spurious peaks from the PDF(z) and, thus, increase the odds.

In our analysis, we defined our priors in terms of the quasar luminosity function, QLF-PLE (see Eq. 3.15). Therefore, for a quasar q in the magnitude-redshift bin (m_i, z_j) , the final redshift probability is denoted by

$$P^q(z) = P_{ij}^q \mathcal{P}^{q, ij}(z). \quad (4.12)$$

4.5.6 Kolmogorov-Smirnov (K-S) statistics

The K-S statistics is a non-parametric test that provides the maximum value of the absolute difference between two cumulative redshift distribution functions. This test is also fairly robust to outliers.

The null hypothesis is that the testing groups were sampled from populations with

identical distributions (in our case, we will be referring to the spectroscopic and photometric redshift distributions). Therefore, the test searches for any violation of the null hypothesis, being powerful to detect changes in the shape of the redshift distributions. Note, however, that it does not provide any information about confidence intervals.

4.6 Final remarks on photo-z estimation

We employ QPz on the photo-z estimation of quasars detected with miniJPAS (Chapter 6) and S-PLUS (Chapter 7), and compare its performance with LePhare++ (or simply LePhare henceforth) in terms of σ_{nmad} , bias and outlier fraction.

In the next chapter we describe how we generated the mock catalogs.

CHAPTER 5

MOCK CATALOGS

Constructing mock catalogs is an optimally approach to exploit data from upcoming surveys or surveys that are not fully operational yet. With synthetic galaxy catalogs one can estimate the uncertainties in deriving galaxy properties, study selection effects, quantify the impact of different sources of errors, and make forecasts regarding the most appropriate scientific strategies for a particular survey. In particular, they have been extremely useful for training machine learning algorithms to classify point-like sources from miniJPAS, since the number of sources with spectroscopic identification is not large enough.

In this chapter, we will focus on the discuss of generating mock catalogs for miniJPAS. Note, however, that the procedure described here can be easily employed for any other multi-band survey, such as S-PLUS.

The synthetic fluxes for quasars, stars, and galaxies are constructed from SDSS spectra, and in order to have realistic mocks there are several requirements.

First, we select spectra from the SDSS DR16 Superset (§2.2.4), a catalog containing a set of sources targeted as quasars for eBOSS with great completeness and purity due to visual inspection procedures to correct the automated classifications and redshifts from the pipeline. This ensures a secure identification of quasar spectra, as well as the inclusion of the main contaminants to the quasar sample.

Second, we consider only good-quality spectra and useful quality information in the median signal-to-noise ratio per resolution element by setting the flags `zWARNING=0` and `snMedian>0`.

Third, we draw realistic redshift-luminosity frequencies for each type (sub-type) of source and angular distributions on the sky (in the case of stars). For the quasars, we draw

those frequencies from the QLF-PLE. For the galaxies, instead of using a phenomenological luminosity function, we mapped the redshift-magnitude distributions of DEEP3 galaxies detected in the AEGIS fields. For the stars, we considered the Besançon Model of stellar population synthesis of the Galaxy (Robin et al. 2003), which provides the stellar counts in the same angular position of the AEGIS fields as a function of some reference broad-band magnitude (in our case, r -band). We did not make any further assumptions on the sub-types of extragalactic sources, and in the case of stars we considered the main sequence, white dwarfs and carbon stars. Drawing samples with realistic frequencies is an important step because due to selection effects the SDSS catalog is shallower than the miniJPAS observations (which reach $i \sim 23.5$); this means that we need a sample that includes less luminous objects than SDSS does. A fair sample of faint sources can be constructed from our initial sample of bright SDSS spectra by making random shifts in the original magnitudes up to $r = 23.5$ (at fixed redshift, for quasars and galaxies, or fixed sub-type, in the case of stars). Therefore, from a single source we generate several fainter sources, and this procedure also help us increase the size of our sample of spectra.

Finally, we make an assessment of the error distribution in a subsample of point-like sources (i.e., $\text{CLASS_STAR} > 0.5$) detected within an aperture of $3''$ with miniJPAS and try to reproduce the same signal-to-noise ratios in each band.

Of course, this is just a first approximation to a much more complex problem since, e.g., (i) we are not taking into account the sub-types of quasars and galaxies; (ii) by simply shifting bright quasar spectra to the faint end we are not properly modeling physical effects such as the Baldwin effect; (iii) we do not simulate fluxes considering a specific aperture and, still, the uncertainties are selected from the observations within an aperture of $3''$; and (iv) we are not directly modeling systematic effects that might affect real observations. If the uncertainties of the observations were not initially properly estimated (e.g., if they were underestimated), this will also propagate to the mocks. Nevertheless, with the current version of the mocks we are already obtaining satisfactory results (as shown in §3.3.2).

In the following we describe in more details how we generated these mock catalogs.

5.1 Photometry

In photometry, the observations are performed with a grid of detectors called Charge-Coupled Devices (CCDs). A CCD is a light sensitive silicon chip electrically divided into a large number of independent picture elements (the so-called pixels). These detectors work through a physical process similar to the photoelectric effect: instead of measuring the energy deposited (such as calorimeters do), they measure how much light falls on each pixel by counting the number of photons accumulated on their surface. Then, the incident photons are converted into a charge that can be measured and recorded in a digital format (image), consisting of a matrix of numbers, one per pixel, each number being related to the amount of light that falls on that pixel, but that also includes other effects such as read out noise (i.e. noise per pixel due to the read out process), bias frame (electrical offset or background) and dark signal (associated with the thermal excitation of electrons when there is no light falling in the CCD). Since the random variations in the number of photons that strike the CCD follow a Poisson distribution, the statistical noise of this measure is proportional to the root square of the number counts. In addition, note that the CCD is an integrating device, which means that the number counts also depend on the amount of time during which the detector was exposed.

Although a CCD detects individual photons, not even the best CCD is able to detect every single photon that falls on it. The fraction of photons falling on a CCD that are actually detected is called the quantum efficiency (QE), which depends on the wavelength. In general, one expects low QE in the blue side because the electronic layers tend to absorb more blue than red light.

A measurement made with this type of detector basically returns the number C of photons collected by the primary mirror of the telescope (of effective area A), captured by the detector during an exposure time $\Delta\tau$:

$$C = A\Delta\tau \int_0^{\infty} T(\lambda) f_{\text{obs}}(\lambda, \Delta t_{\text{obs}}) \frac{\lambda}{hc} d\lambda \quad (5.1)$$

where $T(\lambda)$ is the total transmission function of the photometric system (i.e., fraction of photons that are effectively detected, taking into account the atmospheric and filter

transmission functions, as well as the detector and telescope efficiencies).

5.2 Synthetic fluxes

It is common to classify the light coming from astrophysical sources in terms of the magnitude, a logarithmic scale of the flux relative to a known source (normally, a standard star). One of the most used systems is the AB magnitude system, calibrated in absolute units (or spectral flux densities). For a flux per unit frequency, the magnitude is given by

$$m_{\text{AB}} = -2.5 \log_{10} f_{\nu} - 48.6. \quad (5.2)$$

To convert the spectral flux density from f_{ν} to f_{λ} , it is useful to use

$$f_{\nu} = \frac{\lambda^2}{c} f_{\lambda}. \quad (5.3)$$

When dealing with photometric bands, we assume that the value of λ above corresponds to the effective wavelength of the bandpass, defined as

$$\lambda_{\text{eff}} = \frac{\int T(\lambda) \lambda d\lambda}{c \int T(\lambda) / \lambda d\lambda}. \quad (5.4)$$

The synthetic fluxes in the mocks are obtained by the convolution of the SDSS spectra, $S_q(\lambda)$, with the filter transmission functions:

$$f_{q,\mu}^s = \frac{\int T_{\mu}(\lambda) S_q(\lambda) d\lambda}{\int T_{\mu}(\lambda) d\lambda} \quad (5.5)$$

where $f_{q,\mu}^s$ is a flux density per unit wavelength (and the index s indicates that this is a synthetic flux). In general, the spectrum is not calibrated, which means that the observed magnitude (i.e., photometric magnitude) in a reference broad-band – let’s say r -band – does not correspond to the the magnitude value obtained by convolving the spectrum with that same band (i.e., synthetic magnitude). So, $S_q(\lambda)$ is actually the spectrum “scaled” by the reference band.

Since the eBOSS wavelength coverage (360-1000 nm) is not sufficient to fully

cover the bluest and reddest J-PAS filters, prior to the convolution we perform a template fitting to the spectra, extending their coverage. For quasars, we adopt a Vanden Berk composite spectrum (Vanden Berk et al. 2001), whereas for galaxies and stars we employ standard templates from LePhare.

Equivalently, the noise in each filter is obtained by adding the uncertainty from the SDSS spectrum in each spectral bin in quadrature ($\sigma_{\text{SDSS},\mu}$). But in order to mimic the signal-to-noise of the miniJPAS observations, we also adopt the following procedure: first, we compile a catalog of point-like sources (of size ~ 4000) and sort the observed magnitudes by filter. Then we search in this catalog for a magnitude value that resembles the synthetic magnitude $m_{q,\mu}^s$, and save the corresponding observed uncertainty ($\sigma_{\text{obs},\mu}$). Finally, we compare these two values of noise: since the noise coming from the observations are typically greater than the one coming from the spectrum, the final uncertainty in the synthetic flux is given by $\sigma_{\mu}^s = \sigma_{\text{obs},\mu}$. Nevertheless, since we cannot remove noise from the spectrum, whenever $\sigma_{\text{SDSS},\mu} > \sigma_{\text{obs},\mu}$, we set $\sigma_{\mu}^s = \sigma_{\text{SDSS},\mu}$.

Lastly, we introduce Gaussian fluctuations to the synthetic fluxes by selecting a random number from a Gaussian distribution with variance equal to σ_{μ}^s .

In Fig. 5.1 we show four examples of synthetic photospectra generated for quasars at different redshifts.

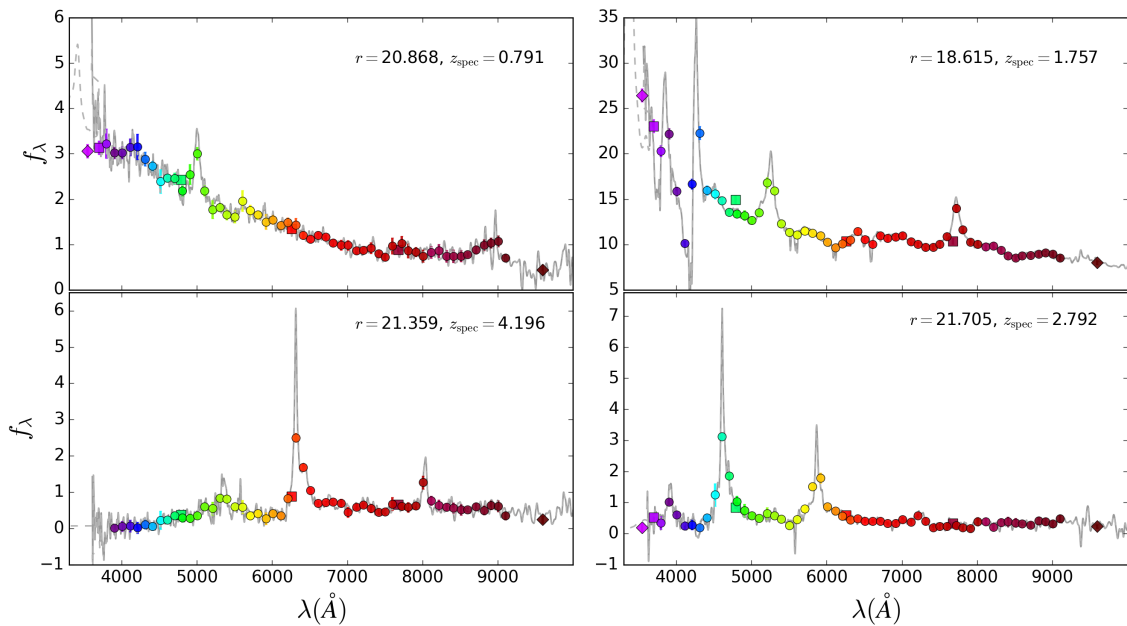


Figure 5.1: Synthetic Jspectra for four quasars at different redshifts. The gray solid line corresponds to the SDSS spectrum, the colored dots are the synthetic fluxes and the gray dashed line illustrates the best-fit using the Vanden Berk composite spectrum.

In Fig. 5.2 we show the signal-to-noise ratio distributions in the r -band for the quasars, galaxies and stars in the mocks, and how it is comparable to the original S/N in the miniJPAS catalog of point-like sources.

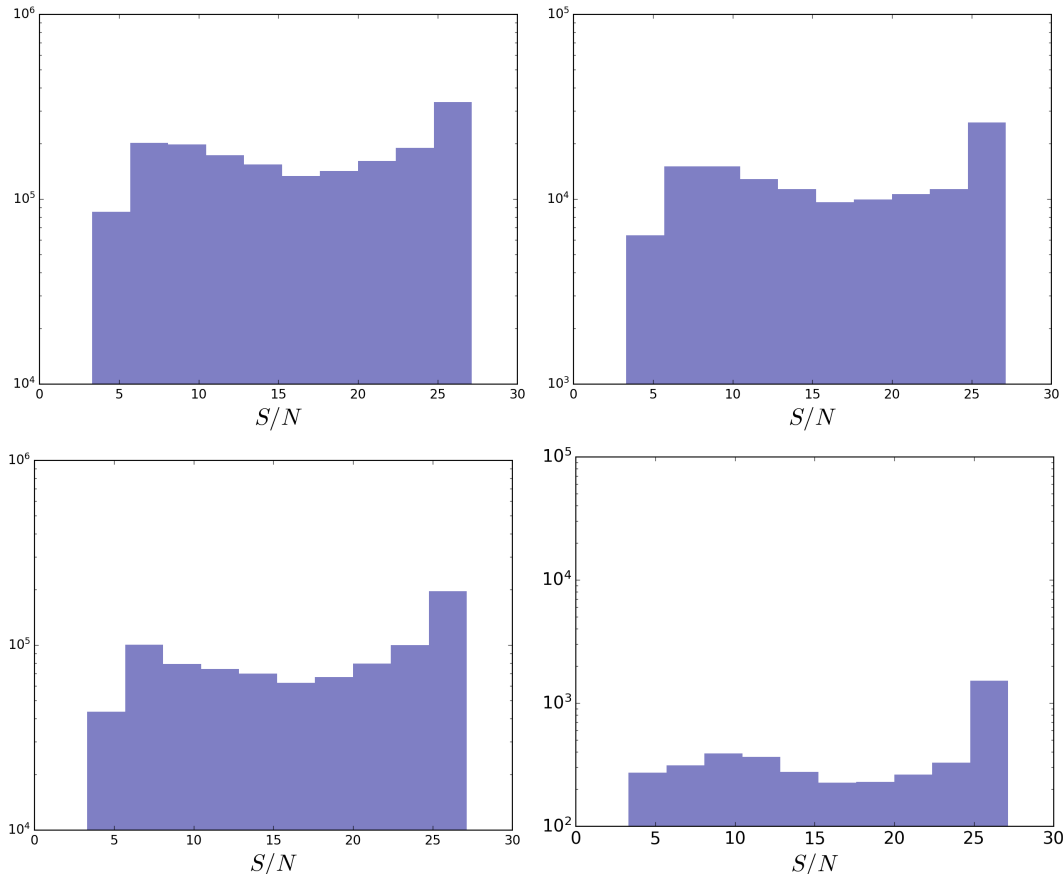


Figure 5.2: S/N distributions in the r -band for quasars (left, upper panel), galaxies (right, upper panel) and stars (left, bottom panel) in the mocks. We also compare them to the S/N distribution in the miniJPAS catalog of point-like sources (right, bottom panel).

5.3 Galactic extinction correction

Since the 10th data release, SDSS has adopted a policy of not correcting the spectra for Galactic extinction, because Milky Way stars have extinctions that differ from that given in Galactic dust maps. Nevertheless, they state that the extinction has a relative small effect over most of the survey area, since the median $E(B - V)$ over the survey is around 0.04 mag¹.

The extinction can be understood as the sum of two physical processes: (i) absorption, efficient for particles (i.e. ISM dust grains) with physical sizes $a > \lambda$, where

¹<https://www.sdss.org/dr16/spectro/caveats/>

λ is the wavelength of the incident radiation; and (ii) scattering, efficient when $a \sim \lambda$. The density of ISM dust grains decreases with the size; this means that the extinction is, in general, more significant for small wavelengths (i.e., the blue part of the spectrum). This effect of higher absorption by smaller wavelengths results in the “reddening” of the spectrum.

Extinction can have two different origins: atmospheric or from the interstellar medium. As the local atmospheric transmission effects were already taken into account in the J-PAS filter transmission functions, here we only correct for the interstellar reddening.

The most common (and simple) measure of reddening is in terms of an excess of color $E(B-V)$. So, the mean Galactic attenuation (or reddening curve) can be represented as

$$A_V = R_V E(B - V) \quad (5.6)$$

where R_V is the ratio of extinction to reddening in the optical, which varies for different regions of the Universe. For instance, it can assume the following values: $R_V = 3.1$ for the Milky Way, $R_V = 4.05$ for the Calzetti law, and $R_V = 2.72$ for the Prévot law.

We use the NASA/IPAC Infrared Science Archive² (IRSA, [Schlegel et al. 1998](#)) service to estimate the Galactic dust reddening $E(B - V)$ at the position (RA, dec) of each spectrum. Then, assuming that most of the reddening in the spectrum is due to the Milky Way, we set $R_V = 3.1$ and the resulting dereddened spectrum is given by

$$S_q^{\text{dered}}(\lambda) = 10^{0.4R_V E(B-V)} S_q(\lambda). \quad (5.7)$$

5.4 Training sets for classification

The classification of sources from miniJPAS has been performed with different machine learning algorithms (e.g., neural network, random forest, convolutional neural network, light gradient boost machine and SQUEZE) in a joint-effort between members from J-PAS and the WEAVE-QSO collaborations. Each method is trained individually

²<https://irsa.ipac.caltech.edu/applications/DUST/>

using balanced samples containing 100 000 objects of each type and frequencies proportional to what we expect in 1 deg². Then, the classifiers are combined using a linear support vector classification (LSVC) and a random forest algorithm, and trained once again on a validation sample containing 10 000 objects of each type. Finally, the combined model is tested on a sample containing 10 000 objects of each type, as well as on the miniJPAS catalog (i.e., real observations).

A perfect classifier would find all the quasars, and only the quasars. However, in practice, this is not possible and some quasars are missed (due to, e.g., the lack of enough emission lines or low signal-to-noise ratio) and some contaminants are also included in the final quasar catalogs. Of course, confidence parameters can be tuned to recover more quasars (at the expense of including more contaminants) or to reject more contaminants (but rejecting true quasars as well). This means that, in the end, the performance of the classifiers depends on a delicate balance between having high purity (p) and high completeness (\mathcal{C}). This balanced performance can be described in terms of the F-score

$$F_1 = 2 \frac{p \mathcal{C}}{p + \mathcal{C}} \quad (5.8)$$

where in an ideal world a perfect classification would imply in $F_1 = 1.0$, i.e. 100% purity and 100% completeness.

Our preliminary results on the test samples of the mocks (Pérez-Ràfols et al. in prep) indicate that the combined algorithms outperform any individual classifier in all magnitude ranges and achieve F-score values very close to 1 (see Table 5.1). Also, all classifiers perform better for high-redshift ($z \geq 2.1$) quasars than low- z quasars – which is not surprising given that low- z quasars lack strong spectral features and have a more significant spectral contribution from the host galaxy, what makes it more difficult to correctly detect them.

As the algorithms were trained on mock data, when we apply the trained models on real miniJPAS observations the performance is not as accurate³. We are still investigating

³Clearly, simulated fluxes are always simpler than real observations, but so far these mock catalogs are the only information available to us in order to perform the classifications of miniJPAS sources. With the WEAVE-QSO spectroscopic follow-ups, we will be able to improve the simulated fluxes, in particular at the faint end (which, currently, has been simulated by degrading the magnitudes of brighter objects). In the future, our idea is to train the algorithms with actual J-PAS data once we have those observations.

Table 5.1: F-score values of the combined performances (linear support vector classifier and random forest) of different machine learning methods in the classification of (miniJPAS) sources from the test sample of the mocks.

Combined algorithm	$r \leq 18$	$18 < r \leq 20$	$20 < r \leq 22.5$	$22.5 < r \leq 23.6$
LSVC	1.0	0.947	0.957	0.951
RF	1.0	0.983	0.997	0.996

the sources of this issue, but one possibility is that our mocks have failed to properly model some of the features present in the real data. One of these features, the frequency of non-observations, is briefly discussed on the next section.

5.5 Non-observations and non-detections

In astronomy, we refer to a non-observation (NO) whenever a certain filter or group of filters could not be observed in a specific night (due to, e.g., the sky conditions), and to a non-detection (ND) whenever the source is fainter than the limiting sensibility of the detector. In practical terms, the magnitudes and corresponding errors are treated separately following these criteria: $m = -99.0$ and $\sigma_m = 0.0$ for non-observations, and $m = 99.0$ and $\sigma_m = m_{\text{lim}}^{S/N}$ for non-detections (where $m_{\text{lim}}^{S/N}$ is a threshold magnitude at some S/N level that depends on the filter and the observation conditions). Note, however, that in the miniJPAS catalog non-detections and non-observations were treated indistinctly, and the magnitude values set to 99.0 in both cases.

Because of the nature of the filter wheel available for the JPAS-Pathfinder camera, the miniJPAS observations were performed in groups of six filters and carried out with different sky conditions. Moreover, the reddest filters were observed last, when the AEGIS fields reached the lowest elevations. This implies in a resulting catalog containing several sources with observations (i.e., $m_\mu \neq 99.0$) in less than 60 filters. This is shown in Fig. 5.3 for the four pointings in the AEGIS fields.

In our approach of selecting separately the uncertainties in each filter directly from the catalog of point-like sources, when we sort the magnitudes in each filter we are actually ignoring possible correlations between passbands with non-detections or non-

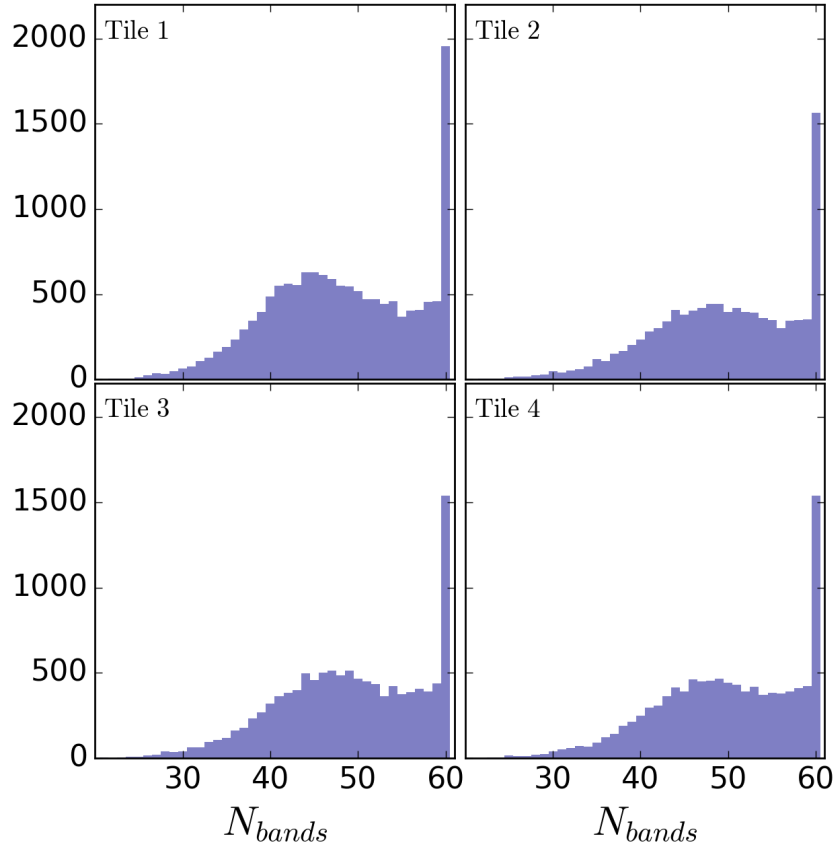


Figure 5.3: Histogram with the number of observed filters in miniJPAS as a function of the tile.

observations. Therefore, we need to have a better assessment of the NOs and NDs in the miniJPAS catalog so that the mocks can mimic in a more realistic way these effects. Moreover, since the traditional machine learning algorithms do not take into account the noise of the observations, this distinction should also be properly addressed in each code, in order to improve the classifications with real data.

Our first efforts to mimic the NO and ND pattern from the miniJPAS catalog in the training set of mocks are shown in Fig. 5.4.

5.6 Final remarks on mock catalogs

Until J-PAS is fully operational, our catalogs of synthetic fluxes are our best (and only) available option to train machine learning classifiers. These mock catalogs are under constant fine tuning, so that they can incorporate more and more features that we expect to find in the real observations.

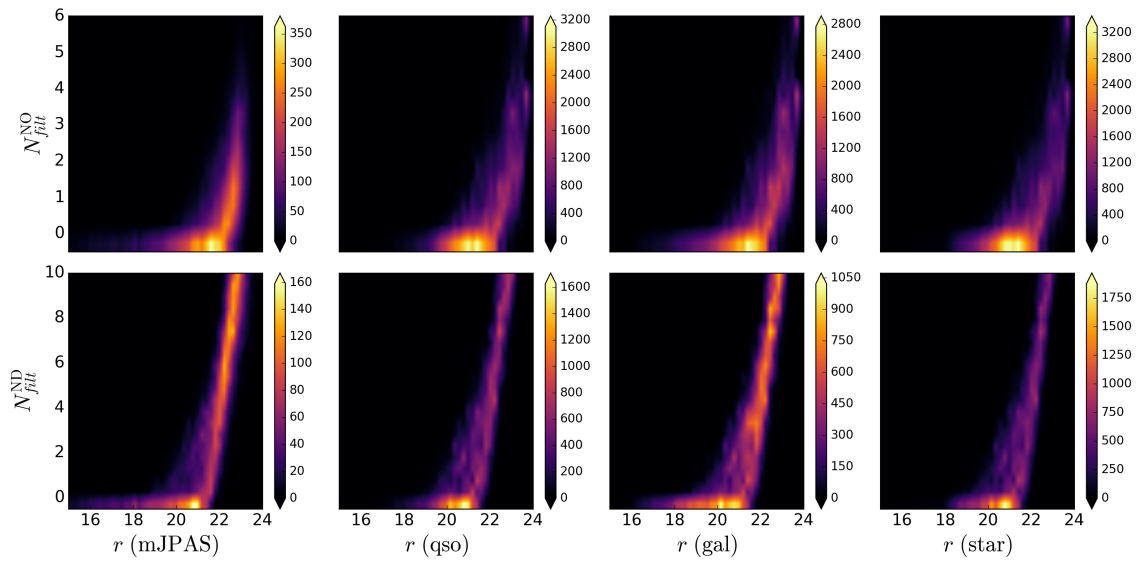


Figure 5.4: Histograms with the number (sum) of filters with non-observations (upper row) and non-detections (bottom row) as a function of the r -band magnitude. Here we compare the NO/ND patterns found in the miniJPAS catalog of point-like sources (left, upper and bottom panels) with the patterns mimicked in the training sets of quasars, galaxies and stars. For fainter sources ($r > 20.5$) we expect an increasing number of filters without observations or detections. The colors indicate the number of objects.

CHAPTER 6

THE MINIJPAS SURVEY

The Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS; [Benitez et al. 2014](#)) will start to scan the northern sky by July 2021 with a unique set of 56 filters and the JPCam instrument ([Taylor et al. 2014](#), [Marín-Franch et al. 2017](#)) onboard the 2.5-m telescope (JST/T250) at the Javalambre Astrophysical Observatory ([Cenarro et al. 2014](#)) in Teruel, Spain.

The J-PAS filter system contains 54 narrow band (NB) filters ranging from 3780 Å to 9100 Å plus two medium-band filters at 3497 Å (named uJAVA) and 9316 Å (named J1007) in the blue and red parts of the optical wavelength range. The 54 NB filters present a full width at half maximum (FWHM) of 145 Å and are equally spaced every ~ 100 Å, whereas the FWHM of the uJAVA band is 495 Å and J1007 is a high-pass filter. These are complemented with 3 SDSS-like broad-bands (u_{JPAS} , g and r). This photometric system is illustrated in Fig. 6.1.

This set of filters is a powerful tool that combines the power of both spectroscopy (with a reliable estimation of physical properties of galaxies) and photometry (with the measure of light in a spatially resolved manner, avoiding the need of pre-selection of targets).

Besides, this filter system was optimized to (i) obtain accurate photo- z measurements and carry out cosmological experiments (such as constrain BAOs, dark energy and modified gravity cosmologies, and probe the large-scale structure) using different tracers at different epochs. For instance, we expect to detect 1.3×10^7 LRGs up to $z < 1$ with $\sigma_z = 0.003(1 + z)$, and 10^8 ELGs up to $z < 1.3$ with $\sigma_z = 0.0025(1 + z)$ ([Benitez et al. 2014](#)); (ii) to deliver a low-resolution ($R \sim 60$) photo-spectrum or J- spectrum that allows us to identify and characterize the galaxy populations up to $z \sim 1$, as well as detect

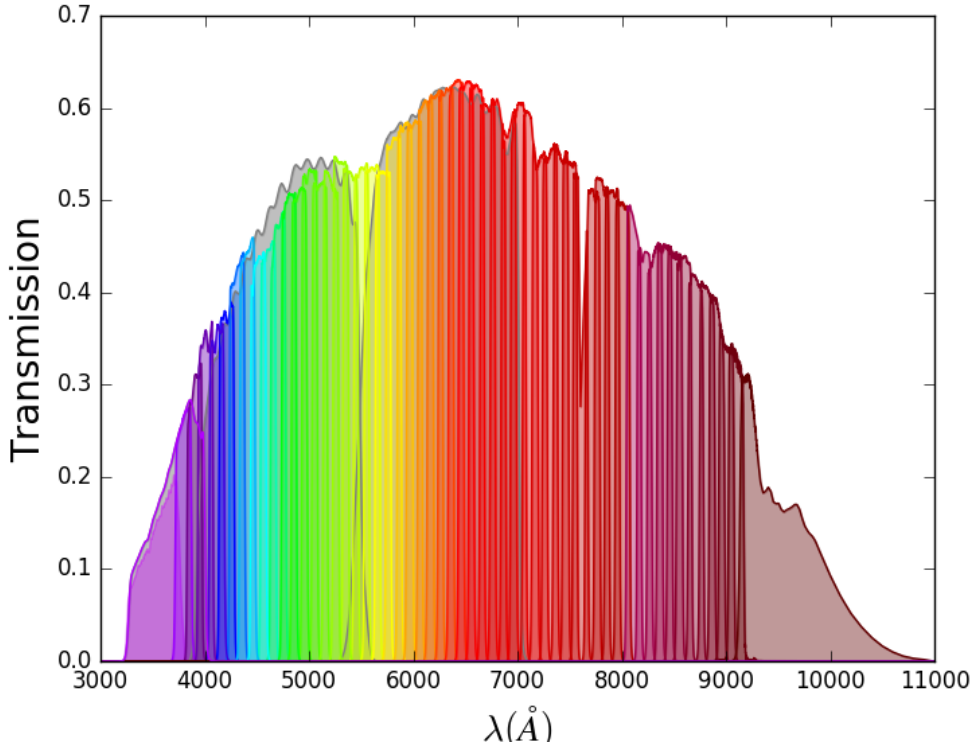


Figure 6.1: J-PAS unique photometric system consisting of 54 narrow-band filters, 2 intermediate-band filters, and 3 broad-band filters. Effects from the CCD quantum efficiency, the entire optical system of the telescope and sky absorption are included.

a large number of clusters and groups ($\sim 10^5$; [Zandivarez et al. 2014](#)); and *(iii)* measure emission lines in galaxies and broad emission line features of quasars and supernovae (e.g., we expect to find $\sim 10^4$ SNe with no need of spectroscopic follow-ups; [Xavier et al. 2014](#)). In particular, the J-PAS filters are sufficiently narrow to marginally resolve the characteristic emission lines that will not only allow the identification of $\sim 2 \times 10^6$ quasars but also a precise redshift determination [$\sigma_z = 0.0015(1+z)$; [Abramo et al. 2012](#)].

When compared to broad-band filters, the flux of photons collected by the narrow-bands is smaller, which implies in a smaller signal-to-noise ratio and, consequently, higher exposure times are needed. The basic J-PAS strategy requires a total minimum exposure time per filter of 480 s, being 4x120 s for the NB plus u (and 960 s for the reddest filters), and 4 x 4 x 30 s for the BBs.

The JPCam is a 1.2 Gpixel camera composed of 14 CCDs and has an effective field-of-view of 4.2 deg². This main instrument has completed the assembly, integration and verification phases and is currently being installed and commissioned. Before the arrival of JPCam, the JST/T250 telescope was equipped with a first light instrument, single

CCD camera, called JPAS-Pathfinder, specifically built to test the optical performance of the system and carry out the first scientific operations.

The Pathfinder Camera conducted the imaging in 60 filters (59 J-PAS filters plus *i*-band) with 4 pointings covering an effective area of 0.895 deg² on the Extended Growth Strip region (also known as AEGIS fields) in the North Galactic hemisphere. This survey is dubbed miniJPAS and the data is publicly available¹. In this catalog, there are calibrated fluxes for 64 293 sources. When we apply some quality cuts, `FLAGS= 0` and `MASK_FLAGS= 0`, we end up with 46 440 sources.

6.1 miniJPAS

In Table 6.1 we show the sky positions of the 4 pointings of the miniJPAS survey. This area was targeted due to (*i*) its location, which can be observed at latitudes $> 30^\circ$ from the OAJ from February to July; and (*ii*) the availability of a wealth of multi-wavelength data, including SDSS and HSC-SSP (Aihara et al. 2019).

Table 6.1: Central coordinates of the four miniJPAS pointings. Credit: Bonoli et al. (2020).

Tile	RA J2000 (deg)	dec J2000 (deg)
miniJPAS-AEGIS1	214.2825	52.5143
miniJPAS-AEGIS2	214.8285	52.8487
miniJPAS-AEGIS3	215.3879	53.1832
miniJPAS-AEGIS4	213.7417	52.1770

The availability of spectroscopic redshifts for galaxies and quasars in the miniJPAS footprint makes this dataset an ideal test case to characterize the performance of photometric redshift estimates for different magnitudes and redshift ranges. In particular, we report 101 quasars from SDSS Data Release 16 Superset and 23 quasars from the DEEP3 Galaxy Redshift Survey in the AEGIS fields. The cross-matches were performed within 1'' and some rigid conditions were applied: (*i*) we only considered sources from miniJPAS with `FLAGS= 0` and `MASK_FLAGS= 0`; (*ii*) we only considered cross-matches with spectroscopically confirmed sources; (*iii*) we imposed cuts of `zWARNING= 0` for

¹<http://archive.cefca.es/catalogues/minijpas-pdr201912>

SDSS quasars, and $z_{\text{QUALITY}}=3$ for secure redshifts with DEEP3; and (iv) many of the sources classified as AGNs by DEEP3 at low redshifts seem to be misclassified, and since the quality of the spectra does not allow reliable visual inspection, we make an additional cut of $z > 1.5$ for DEEP3 quasars.

Fig. 6.2 shows the AEGIS fields with the quasars identified from SDSS DR16 (black circles) and DEEP3 (blue circles). The dots are colored in spectroscopic redshift. In Fig. 6.3 we show the redshift and r -band magnitude distributions of the quasar sample: in blue, quasars from SDSS, and in red, from DEEP3. Note how the DEEP3 quasars span fainter magnitudes, complementing the SDSS cross-matches.

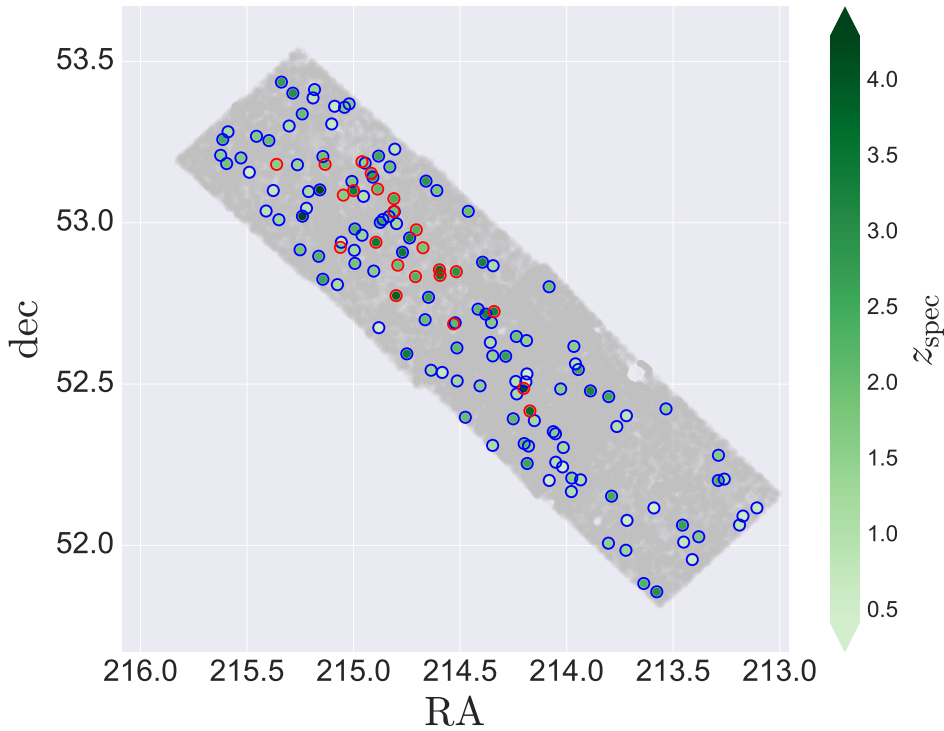


Figure 6.2: Footprint of the miniJPAS survey (gray area). The dots indicate the positions of spectroscopically confirmed quasars (blue circles for SDSS quasars and red circles for DEEP3 quasars), colored by spectroscopic redshift.

The resulting depths in each filter are shown in Fig. 6.4. The minimum target depths are reached in all the filters, with most actually reaching fainter magnitudes. The differences in depth from band to band depend both on the net effect of sky brightness when the observations were acquired and on the final number of combined images (Bonoli et al. 2020).

We provide a foretaste of J-PAS science with quasars in Fig. 6.5, where we present three examples of J spectra for miniJPAS quasars at different redshifts. In particular, note

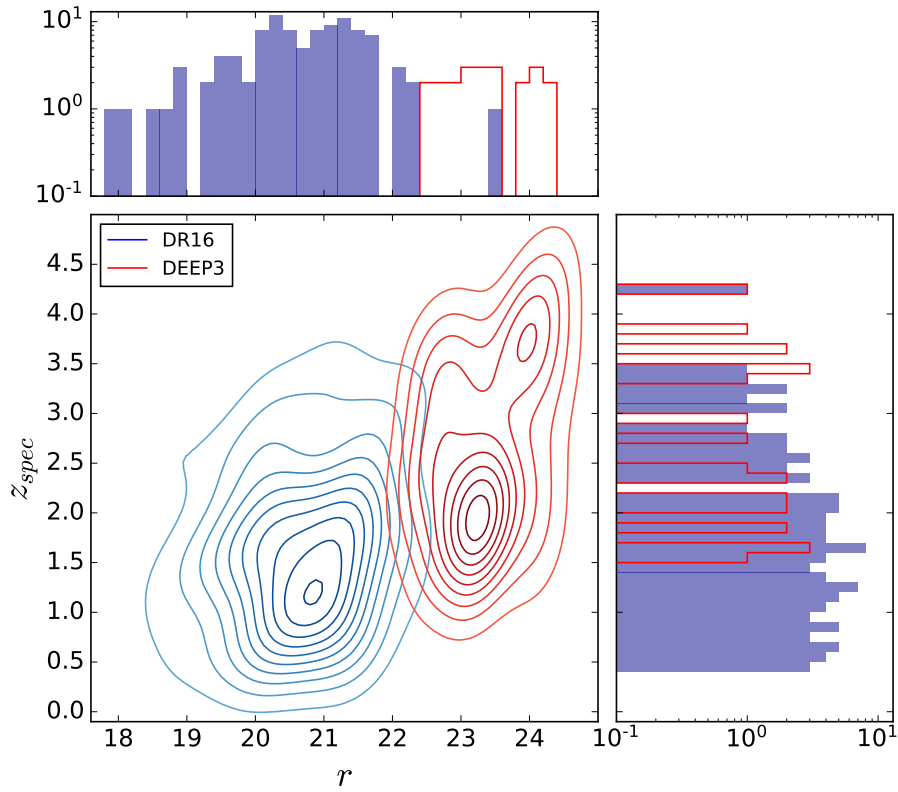


Figure 6.3: Magnitude-redshift distribution of quasars in miniJPAS. Quasars from SDSS DR16 Superset are shown in blue, and quasars from DEEP3 in red. The r -band magnitude corresponds to an aperture of $3''$.

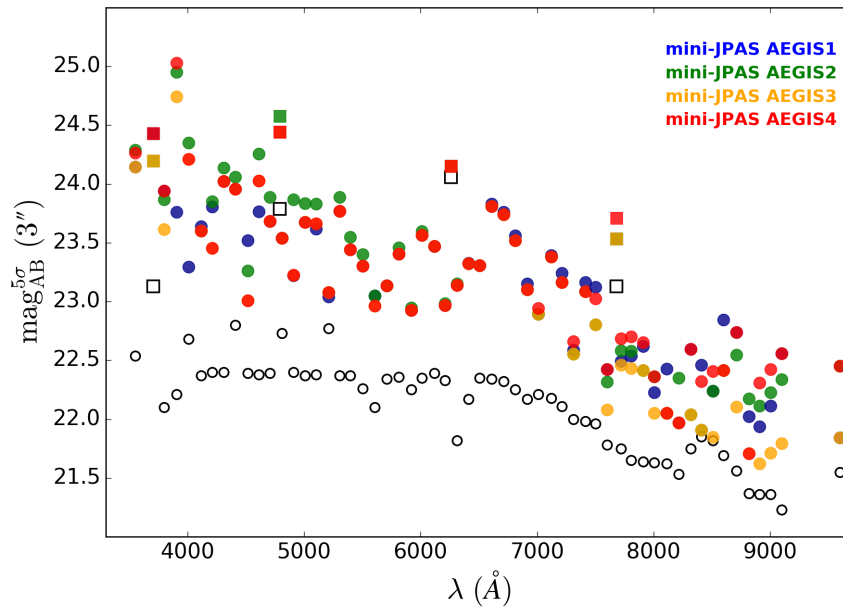


Figure 6.4: Estimated depths (5σ at $3''$) computed from the noise in each tile, for the narrow-bands (dots) and broad-bands (squares). Each color represents a different tile. Black unfilled symbols indicate the approximate targeted minimum depth defined in Benitez et al. (2014). Here λ indicates the filter effective wavelength.

how well the narrow-bands are able to detect not only broad emission lines, but also the broad-absorption lines of the BAL quasar in the middle panel.

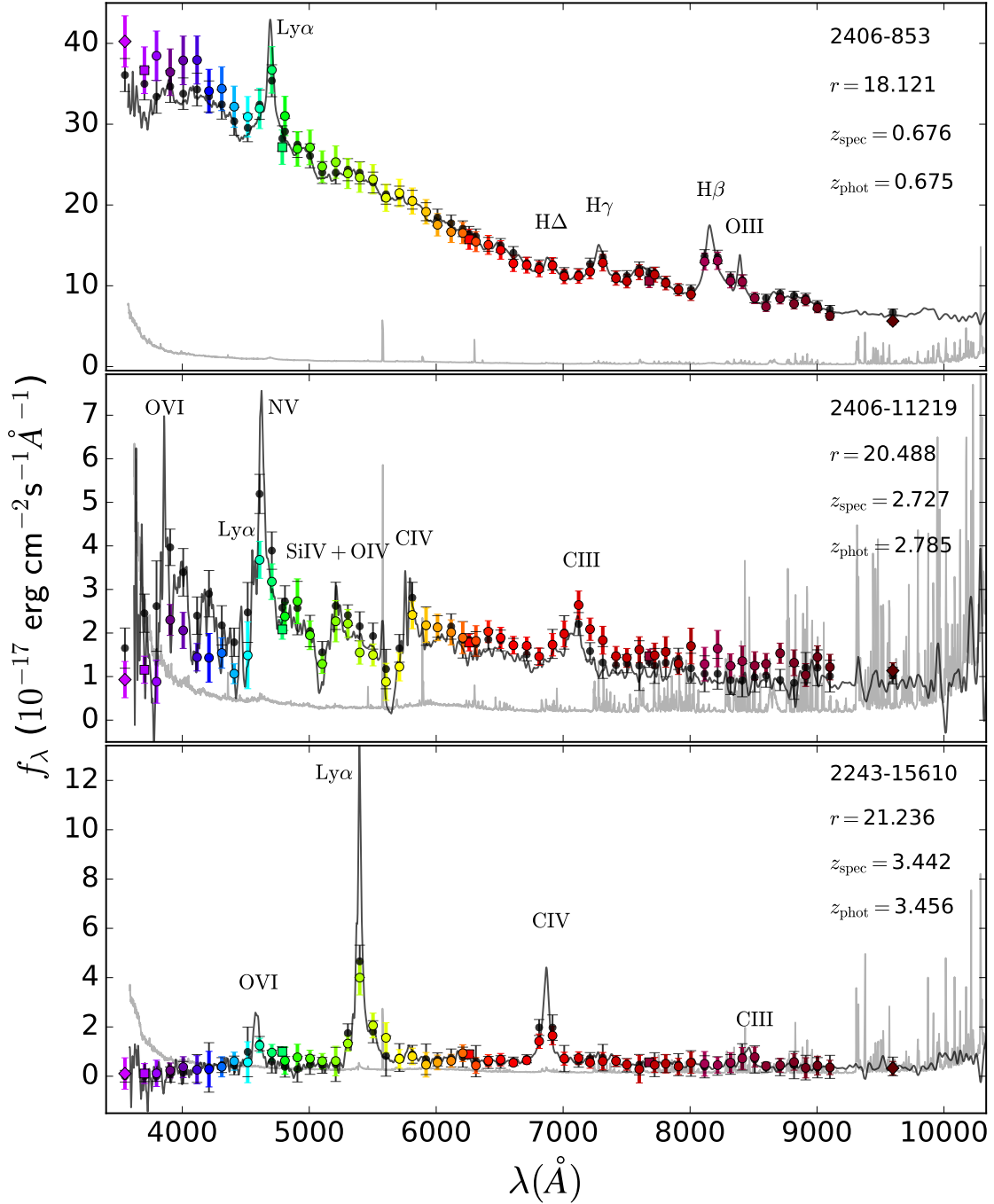


Figure 6.5: Ispectra for three miniJPAS quasars at different redshifts. Gray solid lines represent the SDSS spectrum, and light gray solid lines at the bottom are the corresponding uncertainties. The dots are a comparison between miniJPAS fluxes within $3''$, represented in black, and synthetic fluxes (generated following the procedure described in §5.2), in colors. Narrow, intermediate and broad-bands are represented by dots, diamonds and squares, respectively. The error bars are 1.5 times larger to facilitate the visualization. The main emission lines were identified. In the middle panel, we show a BAL quasar.

6.1.1 Aperture choice

There are several apertures available in the miniJPAS catalog. In order to choose the most appropriate one for the analyses with quasars, we tested the performance with

QPz using some of them. `AUTO_MAG` apertures are not recommended for point-like sources, and `PSF_CORR` magnitudes seem to have some issues of the broad-band calibrations. In Fig. 6.6 we compare the photo- z estimates for quasars using three different apertures: PETRO, APER3 and APER6.

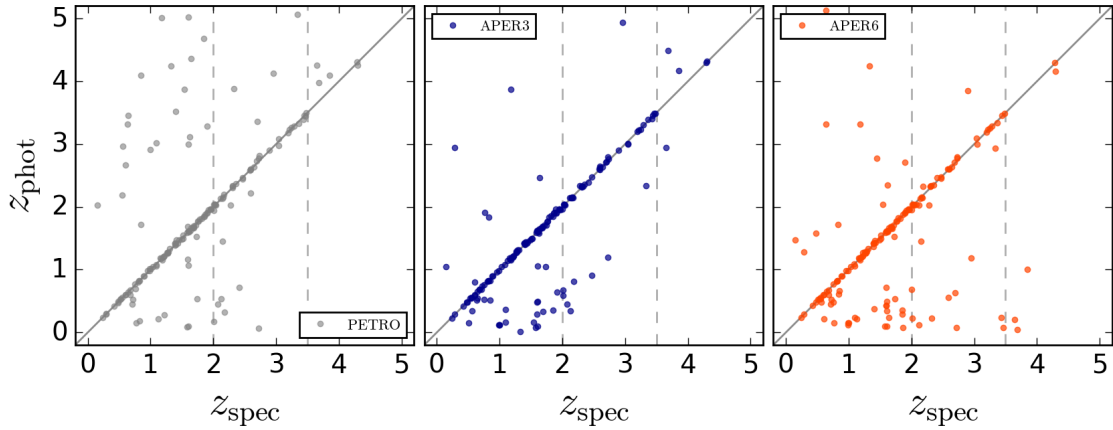


Figure 6.6: Comparison of photometric performance with QPz for different apertures: Petrosian (left panel), aperture within 3'' (middle panel) and aperture within 6'' (right panel). The diagonal line indicates $z_{\text{phot}} = z_{\text{spec}}$. The vertical dashed lines at $z_{\text{spec}} = 2$ and $z_{\text{spec}} = 3.5$ indicate the range where we expect higher precision due to the presence of several emission lines.

We show the quality of the photo- z estimates considering different apertures in Table 6.2. As the magnitude within 3'' provides the smallest redshift error ($\sigma_{\text{nmad}} = 0.0102$) and also the smallest outlier fraction ($\eta = 26.3\%$), we chose that aperture in our analyses for quasars. In principle, we could think about using even smaller apertures (i.e. within 2'' or 1'') but in that case larger corrections would be applicable due to background noise.

Table 6.2: σ_{nmad} , redshift bias and outlier fraction obtained with QPz using different apertures and for a sample of 124 quasars from miniJPAS.

Aperture	σ_{nmad}	δ_z	η
PETRO	0.0109	0.105	30%
APER3	0.0102	-0.034	26.3%
APER6	0.0147	-0.044	35%

6.1.2 Aperture corrections

Applying aperture corrections is essential when one wants to employ template-fitting methods because the chi-square minimization is sensitive to the calibration of the fluxes.

We need to correct the observed fluxes by two different types of offsets (as shown in Fig. 6.7). The first one comes from the fact that the process of data calibration of the full catalog was not optimal for both extended and point-like sources. The corrections for APER6, which are both tile and filter-dependent, are available in the catalog. However, as we have seen previously, the most appropriate aperture for quasar analyses seems to be the one within 3". Hence, we apply the following relation

$$\Delta m_{\mu, \text{APER3}} = \langle m_{\mu, \text{APER6}} + \Delta m_{\mu, \text{APER3}} - m_{\mu, \text{APER3}} \rangle \quad (6.1)$$

to obtain the corrections to the APER3 magnitude in each filter μ . Here, Δm indicates the correction for the corresponding aperture, and $\langle \rangle$ denotes an average over the tile. A zero-point (calibration) correction of 0.04 mag also has to be applied to the uncertainties in all filters.

The second type of offset comes from fine-tuning the fluxes by comparing the miniJPAS stellar observations with synthetic fluxes computed from real spectra for the same stars. We use stars and not quasars because the SDSS spectra were not obtained in the same epoch as the miniJPAS observations, and since quasars are known to suffer from variability, this could imply in larger offsets that are not necessarily related to problems in the miniJPAS calibration. Stars, on the other hand, are less subject to these effects. This correction is also dependent on tile and filter, but we verified that this is actually a second-order effect.

Lastly, we also apply a Galactic extinction correction to the data using a similar procedure as described in §5.3.

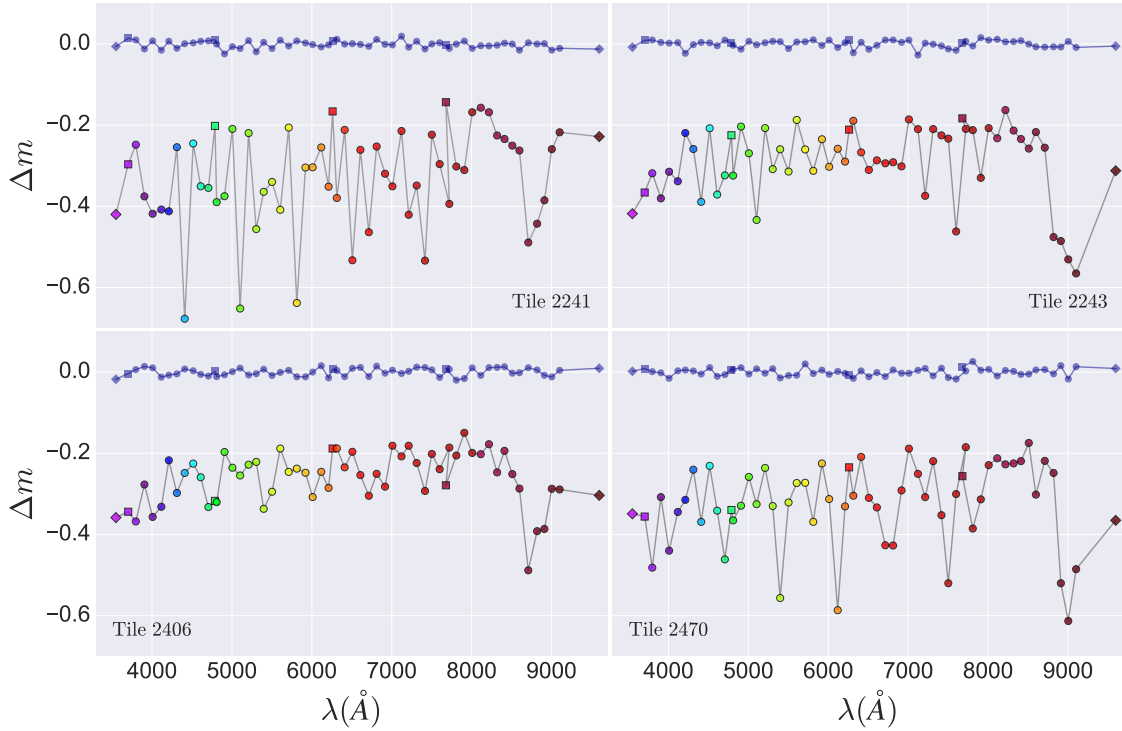


Figure 6.7: Aperture corrections applied to the miniJPAS data as a function of the tile. Colored dots represent the corrections directly related to the aperture radius (and to the fact that we are not considering total magnitudes), while blue dots represent a second-order correction due to offsets in the observed stellar fluxes when compared to synthetic fluxes. Narrow, intermediate and broad-bands are represented by dots, diamonds and squares, respectively.

6.2 Stellarity

We assume that quasars are point-like sources, but caution must be taken because this is not always the case, especially for low- z quasars, where the contribution from the host galaxy is very significant.

In Fig. 6.8 we show the stellerity parameter² ($\mathcal{P}_{\text{star}}$) for miniJPAS quasars as a function of the redshift and the magnitude in the r -band in each tile. A value of $\mathcal{P}_{\text{star}}$ closer to 1 indicates higher probability of the source being point-like.

We note that there are a few quasars with relatively high S/N that are actually extended sources ($\mathcal{P}_{\text{star}} \lesssim 0.2$).

²Even though it might not be as precise as adopting a surface-brightness or the magnitude-radius criterium.

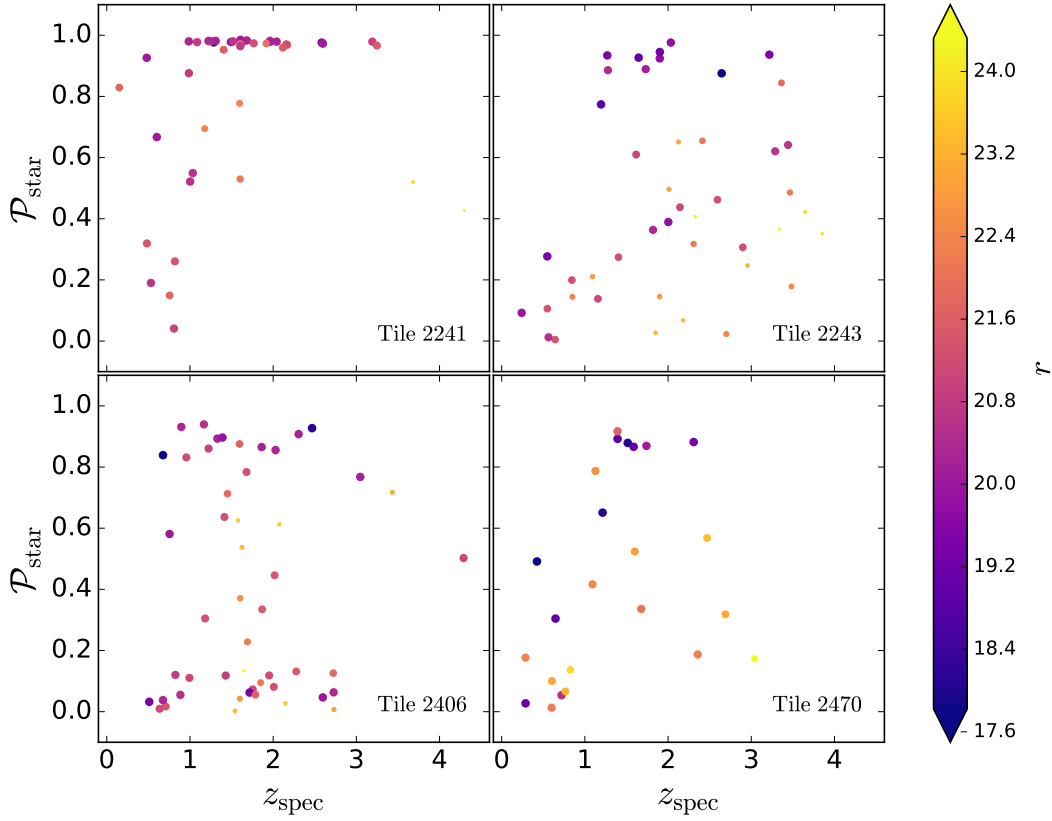


Figure 6.8: Stellerity parameter for miniJPAS quasars as a function of the redshift in each tile. The dots are colored by the r -band magnitude, and the dot sizes are related to the signal-to-noise in the r -band.

6.3 Results

QPz provides a first-approximation fit to the quasar photospectra. In Fig. 6.9 we show the resulting models for three quasars at different redshifts detected in miniJPAS.

In Fig. 6.10 we compare the performances obtained with QPz and LePhare. Considering the whole redshift range:

- for QPz we obtained: $(\sigma_{\text{nmad}}, \delta_z, \eta) = (0.0069, -0.0020, 17.7\%)$; and
- for LePhare: $(\sigma_{\text{nmad}}, \delta_z, \eta) = (0.0061, -0.0003, 22.6\%)$.

Since the continuum does not contribute to the redshift estimation and all significant information comes from emission lines, by employing a photometric system that covers the optical range in small steps of wavelength ($\sim 100 \text{ \AA}$), we are already in the limit of obtaining small values of σ_{nmad} . In this case it is, thus, the outlier fraction that gives us invaluable information about the performance of each method. As we can see, LePhare slightly outperforms QPz in most redshift bins, but it provides a larger fraction

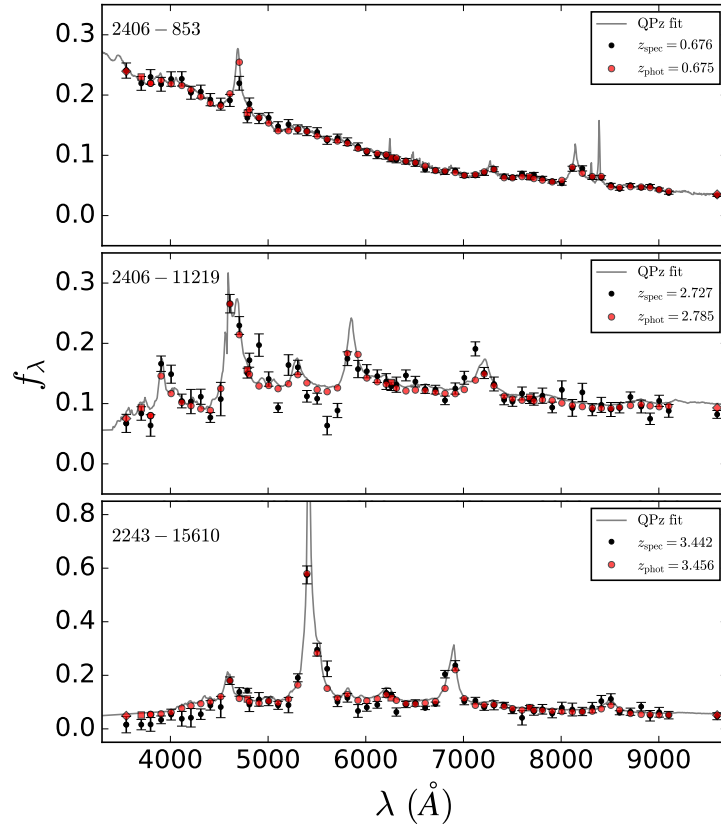


Figure 6.9: QPz model fit (gray solid line) for three miniJPAS quasars at different redshifts. The black dots indicate the miniJPAS observations within 3'' and the red dots correspond to the model flux in each band at the maximum likelihood value of redshift (i.e., z_{phot}).

of catastrophic errors. In particular, we note that most of LePhare outliers have medium to large values of S/N.

In Fig. 6.11 we also compare the spectroscopic with the photometric redshift distribution for both QPz and LePhare. From this figure it is easy to see the superiority in the performance of QPz over LePhare, also demonstrated by a two-sample K-S test. For QPz, it results in a K-S statistic of 0.08, with p-value of 81.7%, whereas for LePhare these results are 0.15 and 10.9%, respectively. Since the p-value is significantly larger for QPz, we can say that this photo-z method is able to better recover the spectroscopic redshift distribution for quasars when compared to LePhare.

In Fig. 6.12 we show some example of zPDFs obtained with QPz. For most of the quasars, the zPDFs are characterized by a single, very narrow peak. The zPDFs obtained with LePhare are very similar to those and are not shown here.

In Fig. 6.13 we show the cumulative distribution of the photometric errors obtained with QPz (blue lines) and LePhare (orange lines) as a function of r -band magnitude and

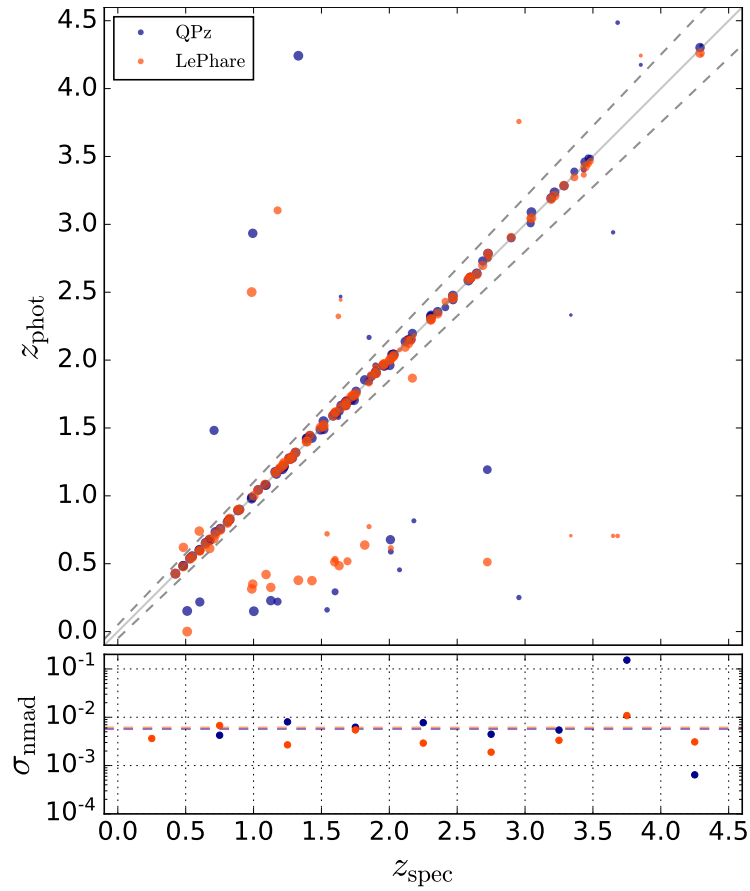


Figure 6.10: Comparison of the photo- z performance of QPz (blue dots) and LePhare (orange dots) for miniJPAS quasars. The size of the dots is related to the signal-to-noise ratio in the r -band. The solid diagonal lines indicates the relation $z_{\text{phot}} = z_{\text{spec}}$ and the dashed lines correspond to $\pm 0.05 \Delta z / (1 + z_{\text{spec}})$. On the bottom panel we show σ_{mmad} for different redshift bins.

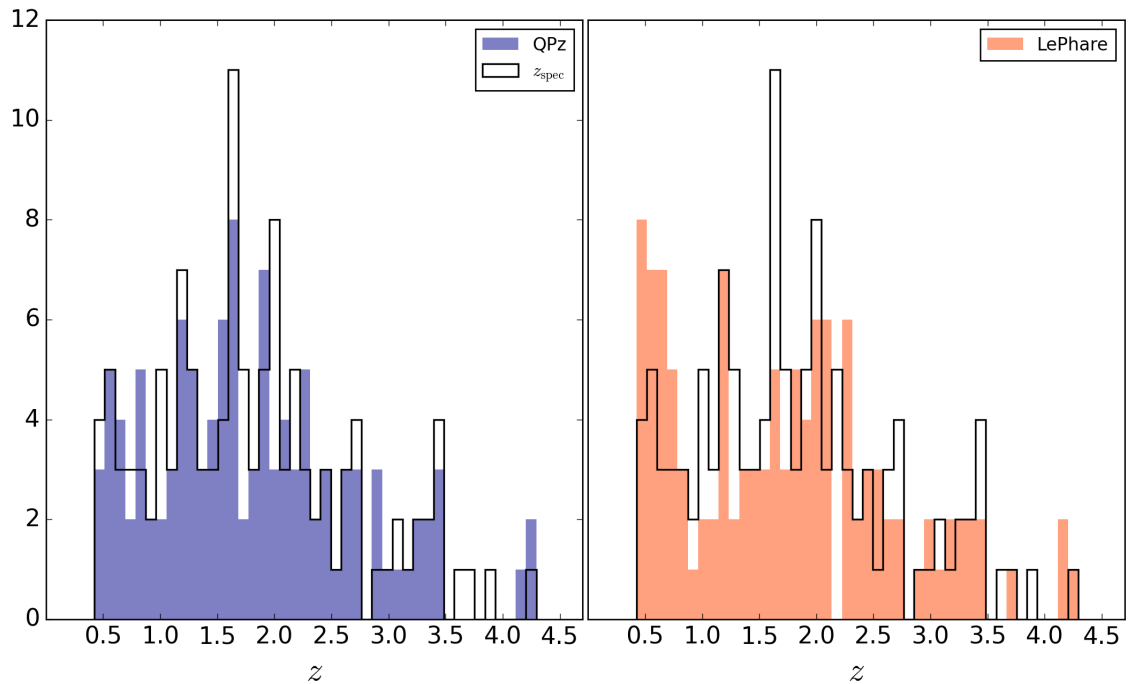


Figure 6.11: Comparison of the spectroscopic (black lines) with the photometric (solid histograms) redshift distribution for QPz (left) and LePhare (right).

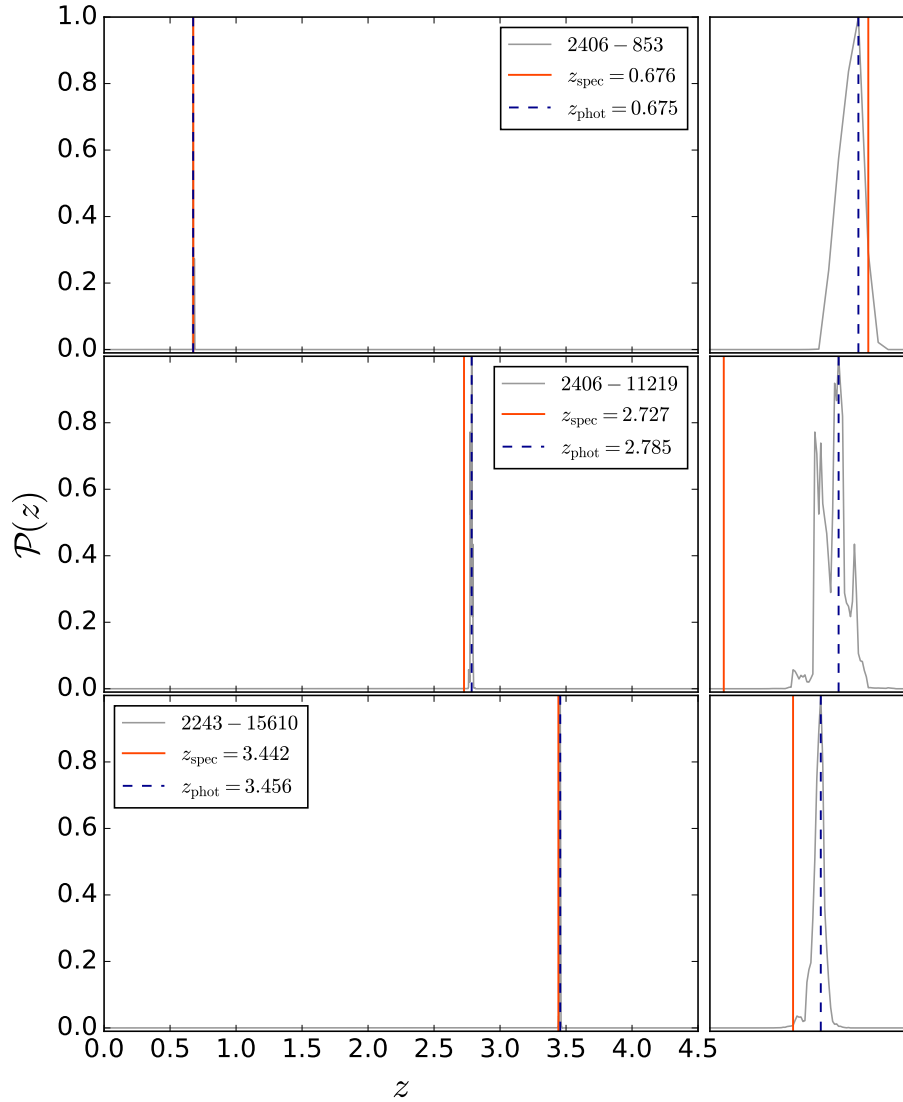


Figure 6.12: Examples of three PDFs obtained with QPz. The orange solid line indicates the spectroscopic redshift, and the blue dashed line indicates the photometric redshift. On the right panel we show the peaks in more details.

redshift. We also compare the results by considering the total sample, as well as only fractions of 75% and 50%. For the total sample, LePhare outperforms QPz in all magnitude bins, but it performs much better at low redshifts ($z \lesssim 1$).

In Fig. 6.14 we show the cumulative distribution of the outlier fraction obtained with QPz (blue lines) and LePhare (orange lines) as a function of r -band magnitude and redshift. In this case, QPz always outperforms LePhare (or have similar results at $r < 19.5$).

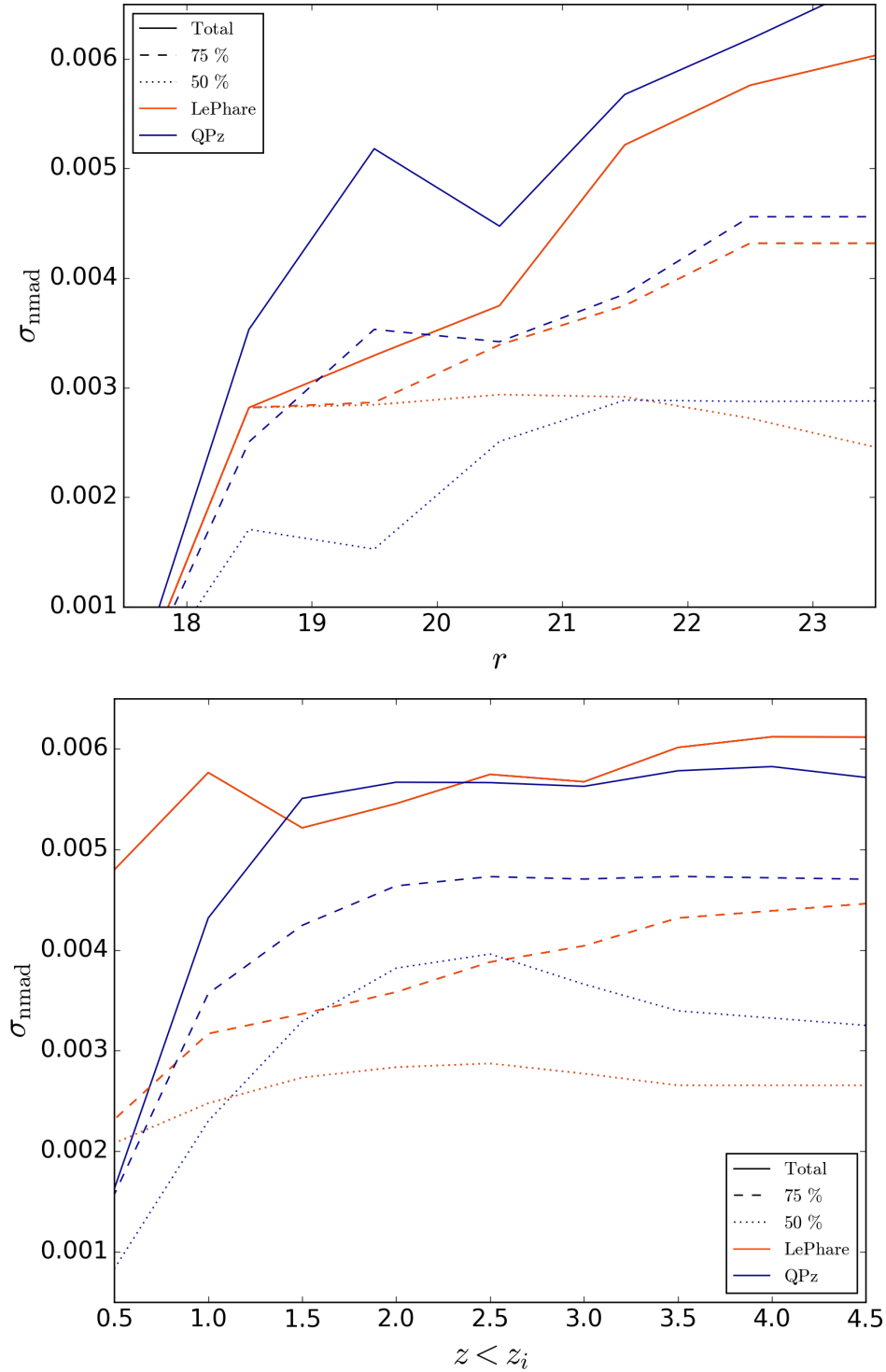


Figure 6.13: Cumulative distribution of the photometric errors as a function of magnitude (upper panel) and redshift (bottom panel) for QPz (blue lines) and LePhare (orange lines). We also show the performances for different fractions of the sample: total (solid line), 75% (dashed line) and 50% (dot-dashed line).

6.3.1 Effect of different filter configurations

At a given redshift range, what are the critical combinations of filters to detect the main quasar features? In order to provide quasar targets with nearly 100% completeness

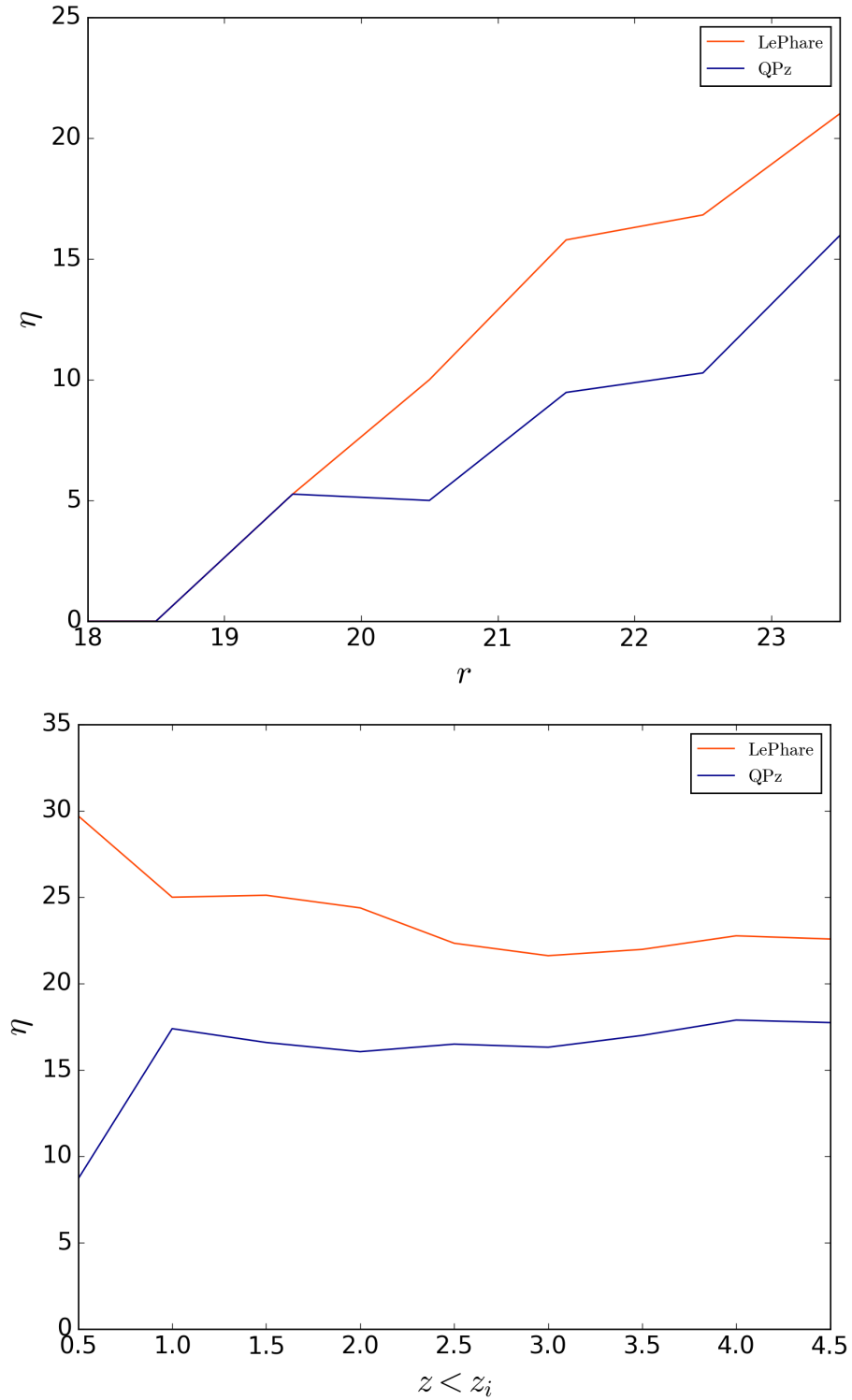


Figure 6.14: Cumulative distribution of the outlier fraction as a function of magnitude (upper panel) and redshift (bottom panel).

and purity for the WEAVE-QSO survey, J-PAS will probably observe with only half of the narrow-bands (at least in the first years of observations). In terms of the quasar identification, our first estimates with the mock catalogs show that the completeness will not be affected very much by going from 42 to 28 narrow-bands (which corresponds to using 3

and 2 trays of filters, respectively). We are also checking whether the r or i -band is more significant to detect quasars.

6.3.2 Priors from the luminosity function

Including priors from the QLF-PLE does not seem to improve the photo-zs.

6.3.3 Variability

Note that each SDSS spectrum is observed at a given epoch, while the miniJPAS sources were observed in different epochs, sometimes with weeks apart. This means that the variability may be playing a non-negligible effect on our data sets since we are employing eigenspectra derived for an early data release of the SDSS.

6.3.4 Luminosity-redshift bin eigenspectra

Although eigenspectra derived from quasars of a particular redshift and luminosity range in general do not predict quasar spectra of other ranges, the performances obtained with the luminosity-redshift bin eigenspectra were not as accurate as the ones obtained with the set of global eigenspectra. Also, the available (M_i, z) -bin eigenspectra do not span all the luminosity ranges occupied by the miniJPAS quasars, which means that sometimes it is not possible to obtain a fit.

6.3.5 Completeness of the quasar sample

Here we investigate the completeness of the quasar sample detected with miniJPAS. In Fig. 6.15 we show the density of quasars (per square degree) as a function of the redshift in three different magnitude bins: $r < 20$, $20 < r < 22$ and $r > 22$. We compare the densities estimated with the quasar luminosity function QLF-PLE (solid lines), with the 124 spectroscopically confirmed quasars detected by miniJPAS (dashed lines) and with all quasars from SDSS DR16 Superset selected within the AEGIS fields but that were not necessarily detected by miniJPAS (dot-dashed lines). The number densities obtained with miniJPAS and the Superset in the AEGIS region are always smaller than the

ones predicted by the luminosity function. Moreover, miniJPAS typically misses some quasars at $r < 22$ but the DEEP3 observations complement the densities at the faint end.

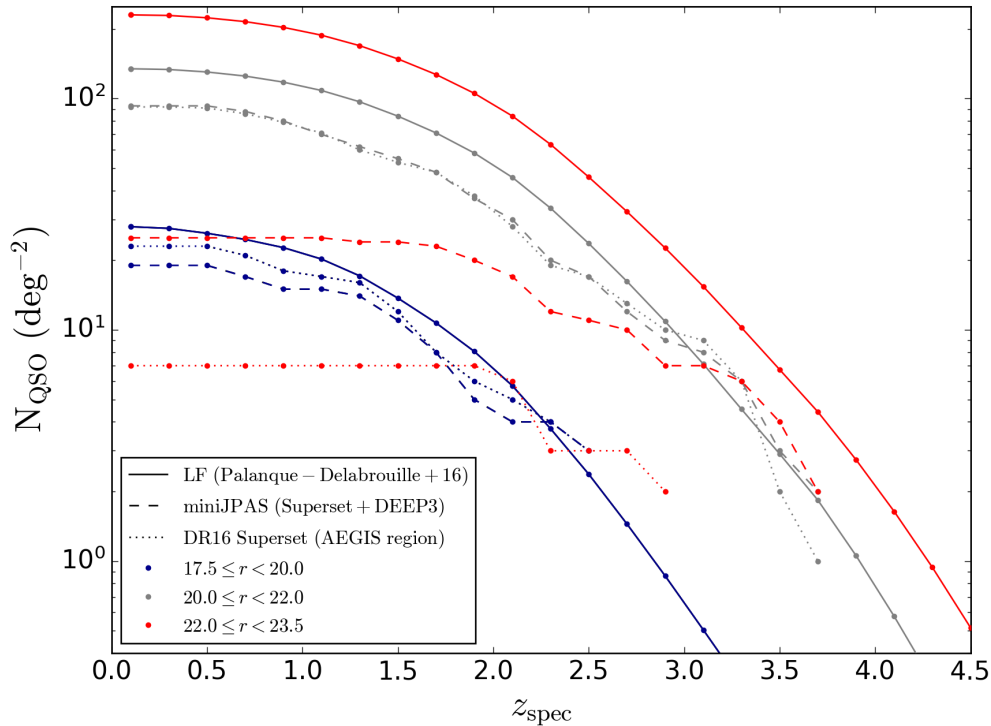


Figure 6.15: Quasar completeness as a function of redshift for different magnitude bins: $r < 20$ (blue), $20 < r < 22$ (gray) and $r > 22$ (red). We compare the density of quasars (per square degree) estimated with the luminosity function (solid lines), for the miniJPAS quasars (dashed lines) and for quasars from DR16 Superset that were detected in the AEGIS fields but not necessarily by miniJPAS (dot-dashed lines).

6.3.6 Validating the results with mocks

Since our sample of quasars with spectroscopic classification is small, we generate a mock quasar catalog over an area of 150 deg^2 to validate our results in a more statistically significant sample. In Fig. 6.16 we compare the results obtained with QPz (left panel) and LePhare (right panel). Considering the whole redshift range:

- for QPz we obtained: $(\sigma_{\text{nmad}}, \delta_z, \eta) = (0.0049, -0.003, 5.1\%)$; and
- for LePhare: $(\sigma_{\text{nmad}}, \delta_z, \eta) = (0.0047, -0.030, 21.3\%)$.

Similarly to what we had already seen with miniJPAS, QPz outperforms LePhare in all redshift and magnitude bins in terms of the outlier fraction, whereas LePhare provides slightly more precise photo-zs.

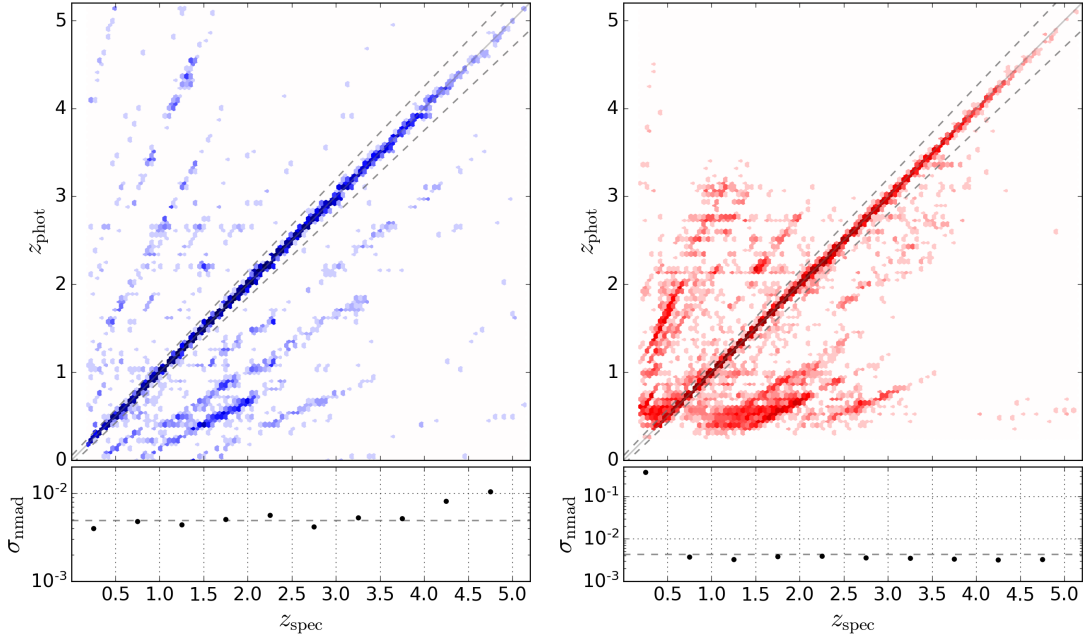


Figure 6.16: Photometric redshift performance for a mock quasar catalog. We compare the performance with QPz (left) and LePhare (right). A quality cut of $\chi^2 < 25$ is applied in both cases, which removes most of the catastrophic errors at low redshifts obtained with LePhare.

One interesting feature that appears in Fig. 6.16 is the presence of diagonal lines of the type $z_{\text{phot}} = z_* + \alpha z_{\text{spec}}$ (more clearly distinguishable for QPz). These features can be better appreciated here because there is a larger number of quasars in the mocks. This effect is a peculiarity of the quasar photo-zs and it has already been described in the literature (see, e.g., [Abramo et al. 2012](#)). Whenever two (or more) pairs of broad emission lines are separated by the same relative interval in wavelength, there is an enhanced potential for a degeneracy between the data and the template (or equivalently, eigenspectrum), and additional peaks appear in the zPDF. As the true redshift of the quasar change, the ratios between these lines remain invariant, which means that the ratios between the true and the false redshifts, $(1 + z_{\text{true}})/(1 + z_{\text{false}})$, also remain constant, giving rise to these diagonal lines. The degeneracy is broken when additional emission lines come into the filter system, which explains why some redshifts are more susceptible to this problem.

We also note catastrophic lines of constant z_{phot} – which appear more often for LePhare. In particular, LePhare is also not able to estimate the photo-zs for many sources, setting $z_{\text{phot}} = 99$ (which were removed from the plot after applying the χ^2 cut).

6.4 Final remarks on miniJPAS quasars

We have developed QPz, a method that combines eigenspectra derived from a PCA to fit the photospectra and estimate photometric redshifts for type-I quasars. QPz provides precise photometric redshifts for quasars observed with miniJPAS and, when compared to a standard template fitting method (LePhare), it seems to outperform in terms of the reduced outlier fraction. The implementation of QPz can be adapted to almost any photometric system (e.g., S-PLUS – as we demonstrate in the next chapter), and can also be employed to estimate redshifts for type-II quasars and high-resolution spectra.

S-PLUS

The Southern-Photometric Local Universe Survey (S-PLUS) is imaging approximately $9,300 \text{ deg}^2$ of the Southern Sky employing an optimized system (Cenarro et al. 2019) consisting of 5 broad-band (BB) SDSS-like filters supplemented by 7 narrow-band (NB) filters, covering the main stellar features from 3700 to 9000 Å, as shown in Fig. 7.1.

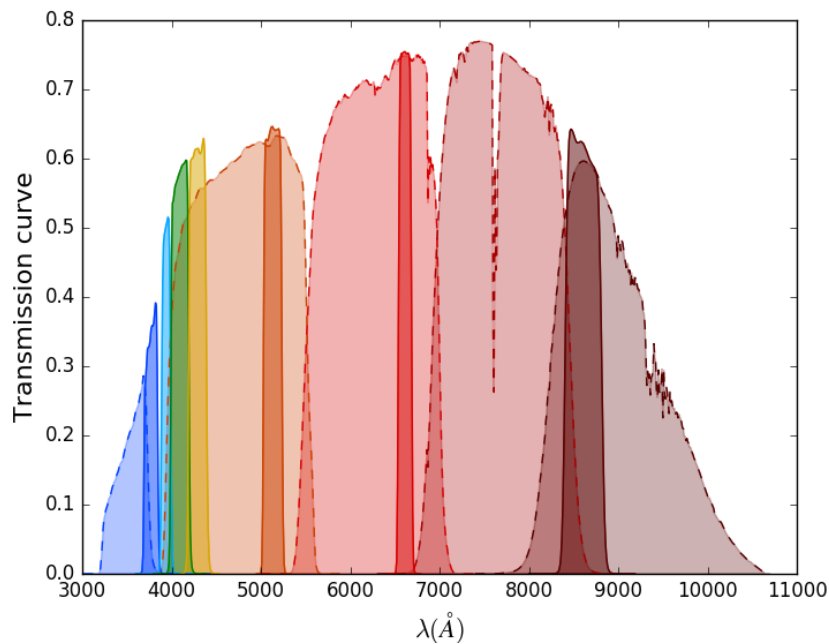


Figure 7.1: S-PLUS filter system with 5 broad-band SDSS-like filters (dashed lines) plus 7 narrow-band filters (solid lines). The y -axis shows the total efficiency taking into account the average filter transmission curves, the atmospheric transmission, the CCD efficiency, and the primary mirror reflectivity curves.

The NB filters cover prominent features in nearby galaxies (i.e. [OII], Ca H+K, H_δ , H_α , and the Mgb and Ca triplets), and reach about one magnitude deeper than SDSS, offering strong constraints on star formation histories as well as photometric redshifts for galaxies. They are, therefore, highly suitable for searching for low-metallicity and carbon-enhanced stars, the blue horizontal branch and variable stars, and for mapping the Galactic

plane. In Fig. 7.2 we show the main quasar emission lines at different redshifts in the S-PLUS photometric system. Note that for each redshift there is a different combination of filters which is more efficient to detect specific emission lines.

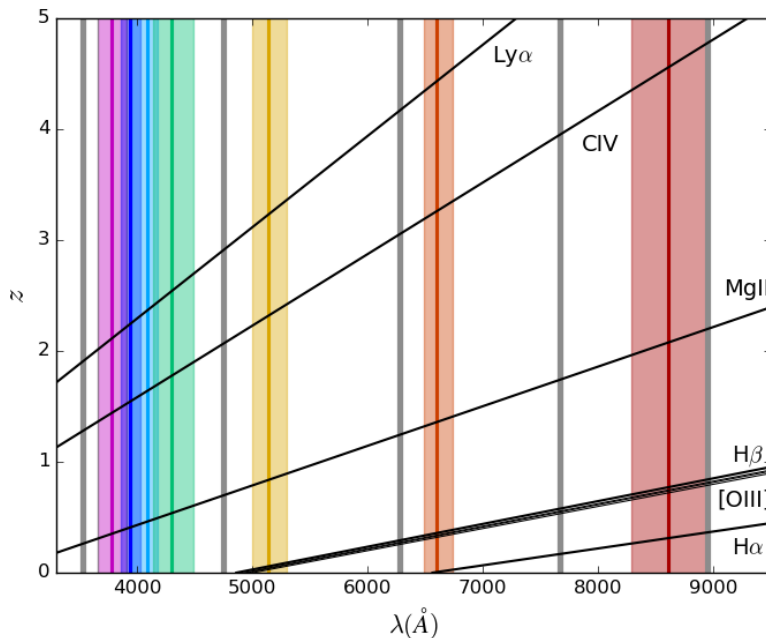


Figure 7.2: Main quasar emission lines at different redshifts in the S-PLUS filter system. Here we represent the central wavelengths of the 5 broad-band filters (gray solid vertical lines) and 7 narrow-band filters (colored solid vertical lines). The colored filled regions represent the wavelength coverage of the narrow bands (not in scale).

S-PLUS is carried out using a 2 deg² field-of-view camera mounted on a fully robotic 0.8-m diameter telescope (T80-South) on Cerro Tololo, in Chile, where it is already operational, taking science data. In particular, we already have a first data release¹ (DR1; Mendes de Oliveira et al. 2019) with more than 3M sources detected in 170 calibrated fields in the Stripe 82 region ($0 < \text{RA} < 60^\circ$; $300^\circ < \text{RA} < 360^\circ$ and $-1.4^\circ < \text{dec} < 1.4^\circ$), covering $\sim 336 \text{ deg}^2$ close to the Celestial Equator. The S-PLUS footprint is shown in Fig. 7.3. In our analyses, we considered the observations in the Stripe 82 region obtained by a preliminary version of the latest internal data release (iDR3) catalog – which contains improved reduction and calibration (in comparison to DR1).

In Fig. 7.4 we show some examples of photospectra detected with S-PLUS.

¹<http://datalab.noao.edu/splus>

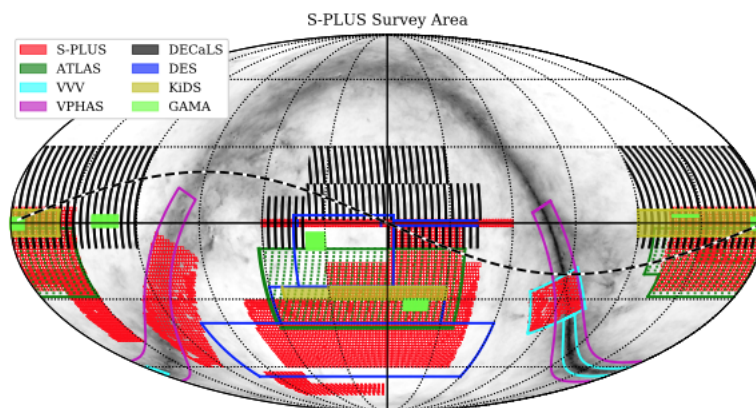


Figure 7.3: Diagram in equatorial coordinates depicting the S-PLUS footprint in red, as well as some of the main optical and near-infrared surveys in the Southern Hemisphere. The dashed-black line represents the ecliptic. The background image is the extinction map of [Schlegel et al. \(1998\)](#). Credit: [Mendes de Oliveira et al. \(2019\)](#).

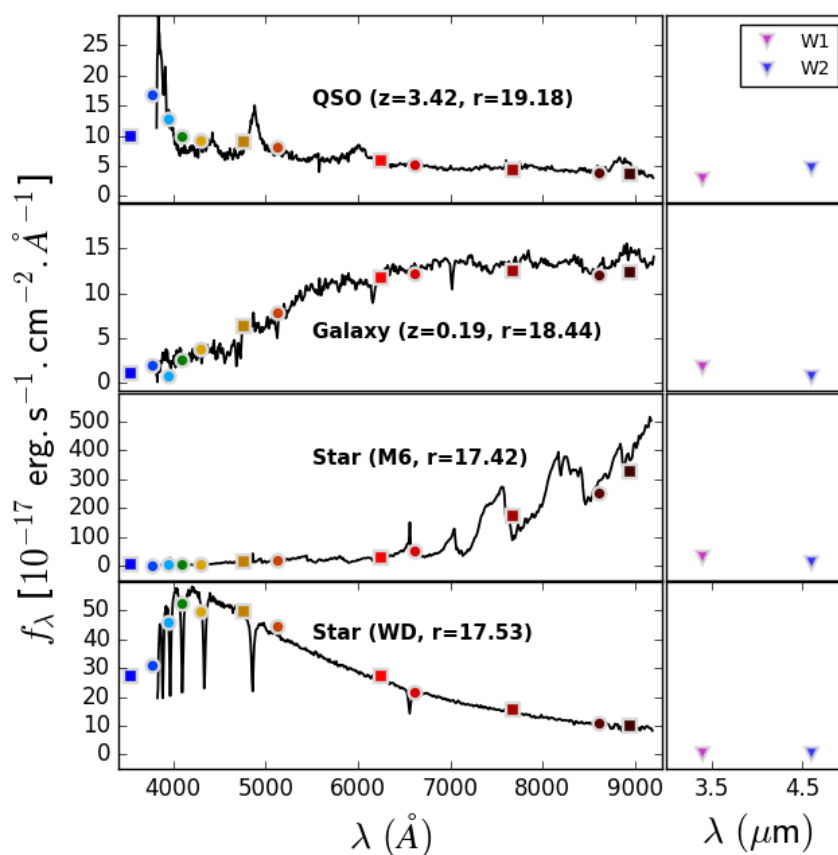


Figure 7.4: Examples of SDSS spectra (black solid lines) and their corresponding fluxes observed with the S-PLUS 12-filter system. The broad-bands are illustrated with colored squares, while the narrow-bands are depicted with colored circles. We also show with triangles the fluxes observed with WISE. Credit: C. Queiroz (Nakazono et al. 2020 submitted).

7.1 Results

After cross-matching the Stripe 82 region from the S-PLUS iDR3 with SDSS DR16 Superset, we obtain about 6 000 quasars, with $r < 22$, $Z_{\text{WARNING}} = 0$ and $\text{PhotoFlag} = 0$ (an output from SExtractor that allows us to select only the objects with reliable photometry from the S-PLUS images). The magnitude-redshift distribution of this quasar sample is shown in Fig. 7.5.

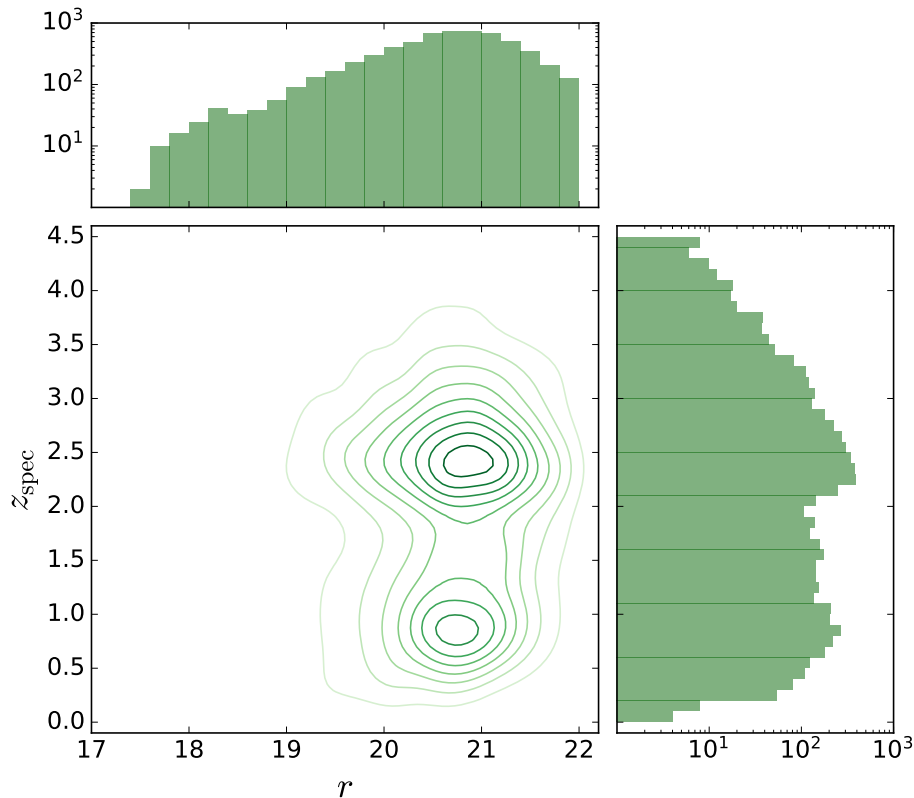


Figure 7.5: Magnitude-redshift distribution of quasars in S-PLUS. The r -band magnitude corresponds to an aperture of $3''$.

Our results with S-PLUS are still very preliminary: the photo-zs are not as accurate as the ones obtained with miniJPAS, and the fraction of catastrophic errors is large (62.8% and 53.5% for QPz and LePhare, respectively). In the case of S-PLUS, LePhare also outperforms QPz in terms of σ_{nmad} : 0.11 and 0.25, respectively. This is an ongoing analysis and we are still investigating the origin of these effects, but one possible explanation might be related to the number of eigenspectra and templates that we are employing, which might be too large for less degrees of freedom (when compared to miniJPAS). The photo-z results obtained with QPz and LePhare are shown in Fig. 7.6, and the χ^2

distributions are shown in Fig. 7.7.

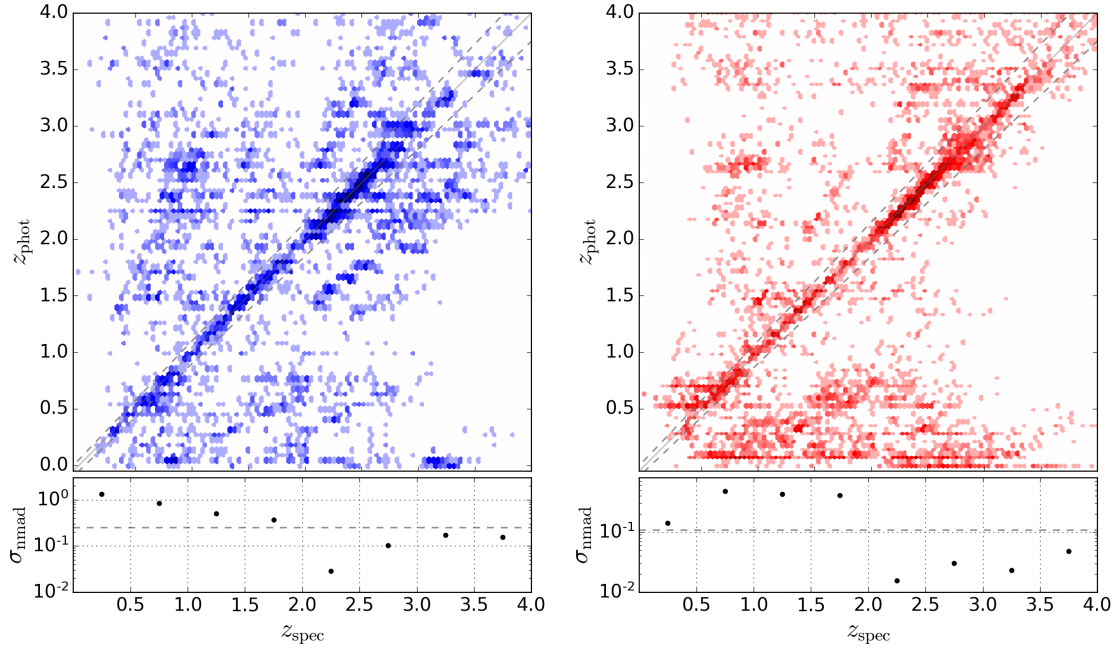


Figure 7.6: Photo- z performance of QPz (left panel) and LePhare (right panel) for S-PLUS quasars after applying a cut in χ^2 of 7 and 40, respectively. The solid diagonal lines indicates the relation $z_{\text{phot}} = z_{\text{spec}}$ and the dashed lines correspond to $\pm 0.05\Delta z/(1 + z_{\text{spec}})$. On the bottom panel we show σ_{nmad} for different redshift bins.

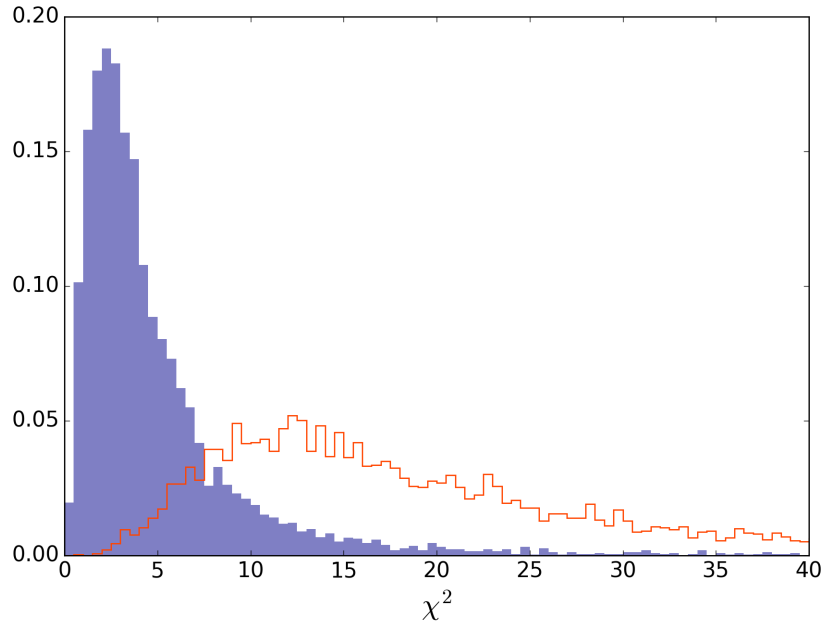


Figure 7.7: χ^2 distributions for QPz (solid blue curve) and LePhare (orange line) obtained for the S-PLUS quasars. We consider quality cuts in χ^2 of 7 and 40, respectively, for QPz and LePhare, which correspond to the thresholds where the fraction of objects is below 1%.

Considering magnitude-limited samples $r < 20.5$, $r < 19.0$ and $r < 18.0$ we obtain, respectively, $\sigma_{\text{nmad}} = (0.066, 0.055, 0.022)$ with QPz. For LePhare, we obtain $\sigma_{\text{nmad}} = (0.141, 0.572, 0.493)$.

If we apply a rigid cut in odds (> 0.9), we are able to recover a precision of $\sigma_{\text{nmad}} = 0.007$ with QPz and an outlier fraction of $\eta = 19.1\%$ but at the expense of reducing drastically the completeness (our final sample has only ~ 500 quasars). We show the photo-z scatter plot and σ_{nmad} as a function of the redshift after this cut in odds in Fig. 7.8.

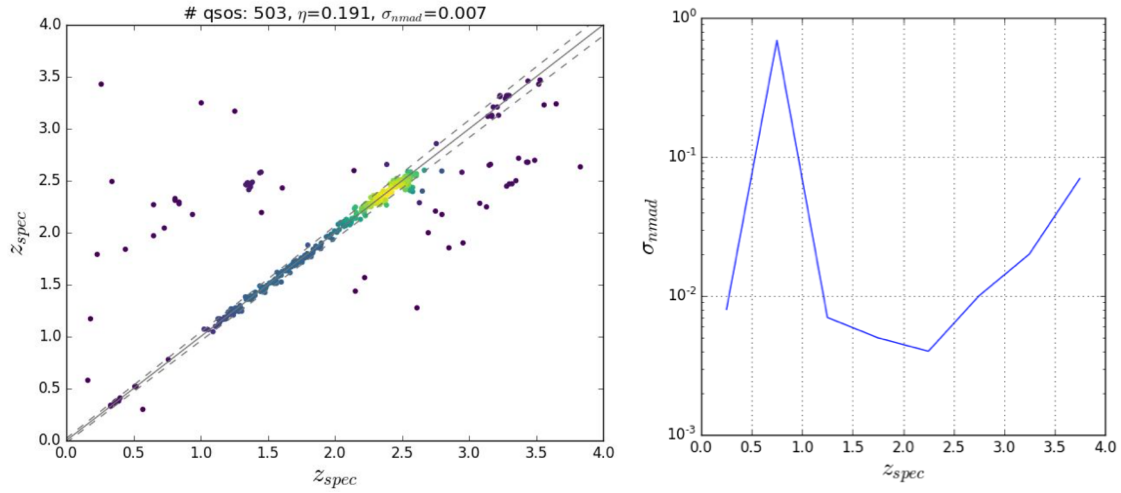


Figure 7.8: Photometric redshift (left) and σ_{nmad} (right) as a function of the spectroscopic redshift for S-PLUS quasars after a quality cut in odds (> 0.9). We obtain a value of σ_{nmad} as good as 0.004 in the range $2 < z < 2.5$.

It is interesting to quantify the benefit of including 7 additional narrow-bands to the classical u, g, r, i, z broad-band system. With QPz, when only broad-bands are considered, we obtain $\sigma_{\text{nmad}} = (0.987, 1.110, 0.563, 0.532)$ for all redshift ranges, $0 < z < 2$, $2 < z < 3.5$ and $z > 3.5$, respectively, and a large sample of outliers ($\eta > 90\%$). With LePhare, many quasars cannot be fitted, and we are still investigating the origin of this issue.

7.2 Final remarks on S-PLUS

We are still testing the best configuration of eigenspectra and templates in order to estimate photo-zs for S-PLUS quasars. We also want to assess to improvement in the photo-z estimation of including 7 narrow-bands to a system containing only 5 broad-band filters. We also have to test the most suitable aperture for quasars (which, similarly to miniJPAS, we suspect it to be $3''$ -magnitudes) and check the photometric calibration by comparing it with synthetic fluxes derived for S-PLUS stars.

CONCLUSIONS AND PERSPECTIVES

Quasars are the most luminous members of the class of AGNs. This implies that they can be detected at large distances and consequently they can act both as sources of background light to investigate the intervening matter through the Ly α forest, or directly as probes of the high redshift Universe. The present and next generation of surveys will allow us to peer deep into space with higher resolution, probing greater volumes, achieving higher completeness and improving the photometric accuracy, pointing to an exciting future with quasar research.

In particular, the J-PAS filter system containing 56 narrow-bands will allow us to identify and provide precise photometric redshifts for millions of quasars. For this reason, the availability of spectroscopically confirmed sources from SDSS and DEEP3 in the AEGIS field made the miniJPAS sample an ideal case to forecast (and foretaste) the photometric precision that we expect to find with J-PAS once it becomes fully operational.

The main contributions of this thesis are: *(i)* generation of catalogs of synthetic fluxes for quasars, galaxies and stars with realistic distributions of signal-to-noise ratio in all bands, as well as the same patterns of non-observations and non-detections found in the catalog of real data; *(ii)* development of a code (dubbed QPz) to estimate photozs for quasars detected in photometric surveys; and *(iii)* inclusion of photometric errors derived for quasars from this code in mock catalogs to study large-scale structure. In the following, we briefly describe each of these contributions.

Mock catalogs are essential to exploit data sets from surveys that are not operational yet, and they have been particularly useful within the J-PAS collaboration to classify miniJPAS sources using machine learning. In fact, since the number of sources with spectroscopic classification is not sufficient to properly train the algorithms, these mocks

are the only data sets available for the training. When we compare the performance of different classifiers on simulated and real data, we note that test sets with synthetic fluxes tend to perform better, but the current results are already very satisfactory. Additionally to the classification task, these mocks have also been used in the search for lensed quasars.

The main output of this classification process will be a list of quasar candidates for spectroscopic follow-up with the WEAVE-QSO survey. So, given a quasar candidate detected with a multi-band system, QPz finds a first-approximation best-fit model and estimate its photo-z. At each redshift the photospectrum is modelled as a linear combination of four redshifted eigenspectra — the most relevant modes of variation of broad-line quasars, combined with an extinction power law, providing, thus, the full redshift probability distribution function.

We employed this method to analyse two data sets: the miniJPAS catalog and the S-PLUS DR1. Due to the differences in the photometric systems, miniJPAS is able to identify more quasar features and, therefore, provide more accurate photo-zs. When compared to the standard template fitting code LePhare, this method is successful in providing precise photometric redshifts with a reduced fraction of catastrophic errors.

Finally, in our preliminary results with a mock catalog containing quasars and ELGs, we found that even after the inclusion of redshift errors for quasars the clustering strength for the monopole is still strong, and that the signal-to-noise ratio for the quasar-ELG cross-spectrum has more power than the spectra of ELGs alone. This is just a foretaste of the multi-tracer analyses that we will be able to do with J-PAS in the future, and the potential that J-PAS will have for studying BAOs and primordial non-Gaussianities.

Most of the results shown here are still preliminary, and there is a lot of room for improvements. In particular, *(i)* we need to assess which features found in the real catalog of miniJPAS observations are not being properly introduced in the mocks; *(ii)* we would like to perform a new PCA on the updated eBOSS sample to derive an updated version of quasar eigenspectra – and maybe extend this analysis to galaxies; *(iii)* adapt our method to fit for line widths as well, so that we can estimate physical properties for quasars from the eigenspectra (e.g., the black hole mass); and *(iv)* have a better assessment of the photo-z precision that we can extract from S-PLUS observations to understand

its potential for the quasar science.

These tools and methods for quasar identification and photo-z estimation will help us construct complete and precise photometric catalogs for quasars, which will open new doors for exploring the cosmic frontier.

Bibliography

- Abazajian K., et al., 2003, *AJ*, **126**, 2081
- Abbott B. P., et al., 2016, *Phys. Rev. D*, **93**, 122003
- Abbott B. P., et al., 2017a, *PhysRevLett.*, **119**, 161101
- Abbott B. P., et al., 2017b, *ApJ*, **848**, L12
- Abbott R., et al., 2020, arXiv e-prints, p. [arXiv:2010.14527](https://arxiv.org/abs/2010.14527)
- Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, *MNRAS*, **417**, 1891
- Abramo L. R., Leonard K. E., 2013, *MNRAS*, **432**, 318
- Abramo L. R., et al., 2012, *MNRAS*, **423**, 3251
- Abramo L. R., Secco L. F., Loureiro A., 2016, *MNRAS*, **455**, 3871
- Abramowicz M. A., Czerny B., Lasota J. P., Szuszkiewicz E., 1988, *ApJ*, **332**, 646
- Ahn C. P., et al., 2012, *ApJS*, **203**, 21
- Aihara H., et al., 2019, *PASJ*, **71**, 114
- Aird J., et al., 2010, *MNRAS*, **401**, 2531
- Alcaniz J. S., Lima J. A. S., 2004, *Brazilian Journal of Physics*, **34**, 455
- Alcock C., Paczynski B., 1979, *Nature*, **281**, 358
- Amaro-Seoane P., et al., 2017, arXiv e-prints, p. [arXiv:1702.00786](https://arxiv.org/abs/1702.00786)
- Amendola L., et al., 2013, *Living Reviews in Relativity*, **16**, 6
- Amendola L., et al., 2018, Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Rel.* 21(1):2. <https://link.springer.com/article/10.1007/s41114-017-0010-3>

- Antonucci R., 1993, *ARA&A*, **31**, 473
- Aparicio Resco M., et al., 2020, *MNRAS*, **493**, 3616
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, **310**, 540
- Assef R. J., et al., 2011, *ApJ*, **728**, 56
- Audibert A., Riffel R., Sales D. A., Pastoriza M. G., Ruschel-Dutra D., 2017, *MNRAS*, **464**, 2139
- Bañados E., et al., 2018, *ApJ*, **856**, L25
- Bacon R., Emsellem E., Combes F., Copin Y., Monnet G., Martin P., 2001, *A&A*, **371**, 409
- Baldwin J. A., 1977, *Lick Observatory Bulletin*, **748**, 1
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *Publ. Astron. Soc. Pac*, **93**, 5
- Baloković M., et al., 2014, *ApJ*, **794**, 111
- Baqui P. O., et al., 2020, arXiv e-prints, p. [arXiv:2007.07622](https://arxiv.org/abs/2007.07622)
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, **304**, 15
- Barger A. J., Cowie L. L., Capak P., Alexander D. M., Bauer F. E., Brandt W. N., Garmire G. P., Hornschemeier A. E., 2003, *ApJ*, **584**, L61
- Becker R. H., White R. L., Helfand D. J., 1995, *ApJ*, **450**, 559
- Beckmann V., Shrader C. R., 2012, *Active Galactic Nuclei*
- Begelman M. C., Volonteri M., Rees M. J., 2006, *MNRAS*, **370**, 289
- Benítez N., 2000, *ApJ*, **536**, 571
- Benítez N., et al., 2009, *ApJ*, **692**, L5
- Benitez N., et al., 2014, arXiv e-prints, p. [arXiv:1403.5237](https://arxiv.org/abs/1403.5237)
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, **575**, 587

- Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, *MNRAS*, **362**, 25
- Bland-Hawthorn J., Maloney P. R., Sutherland R. S., Madsen G. J., 2013, *ApJ*, **778**, 58
- Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, *MNRAS*, **416**, L70
- Bolton A. S., et al., 2012, *AJ*, **144**, 144
- Bolzonella M., Miralles J. M., Pelló R., 2000, *A&A*, **363**, 476
- Bondi H., 1952, *MNRAS*, **112**, 195
- Bonoli S., et al., 2020, arXiv e-prints, p. [arXiv:2007.01910](https://arxiv.org/abs/2007.01910)
- Boroson T. A., Green R. F., 1992, *ApJS*, **80**, 109
- Bovy J., et al., 2011, *ApJ*, **729**, 141
- Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C., 2000, *MNRAS*, **317**, 1014
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, **686**, 1503
- Breiman L., 2001, *Machine Learning*, **45**, 5
- Brown M. J. I., et al., 2006, *ApJ*, **638**, 88
- Busca N., Balland C., 2018, arXiv e-prints, p. [arXiv:1808.09955](https://arxiv.org/abs/1808.09955)
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, **533**, 682
- Carrasco Kind M., Brunner R., 2014, MLZ: Machine Learning for photo-Z (ascl:1403.003)
- Castander F. J., Treister E., Maccarone T. J., Coppi P. S., Maza J., Zepf S. E., Guzmán R., 2003, *AJ*, **125**, 1689
- Cattaneo A., et al., 2009, *Nature*, **460**, 213

- Cenarro A. J., et al., 2014, in Peck A. B., Benn C. R., Seaman R. L., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V. p. 91491I, doi:10.1117/12.2055455
- Cenarro A. J., et al., 2019, *A&A*, 622, A176
- Chatfield C., Collins A., 1980, Principal component analysis. In: Introduction to Multivariate Analysis. Springer, Boston, MA., p.57-81.. https://link.springer.com/chapter/10.1007/978-1-4899-3184-9_4
- Chatterjee S., Degraf C., Richardson J., Zheng Z., Nagai D., Di Matteo T., 2012, *MNRAS*, 419, 2657
- Chaves-Montero J., et al., 2017, *MNRAS*, 472, 2085
- Chaves-Montero J., Angulo R. E., Hernández-Monteagudo C., 2018, *MNRAS*, 477, 3892
- Coles P., Jones B., 1991, *MNRAS*, 248, 1
- Colless M., et al., 2001, *MNRAS*, 328, 1039
- Collister A. A., Lahav O., 2004a, *Publ. Astron. Soc. Pac*, 116, 345
- Collister A. A., Lahav O., 2004b, *Publ. Astron. Soc. Pac*, 116, 345
- Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, *AJ*, 110, 2655
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, 647, 201
- Cooper M. C., et al., 2011a, *ApJS*, 193, 14
- Cooper M. C., et al., 2011b, *ApJS*, 193, 14
- Cooper M. C., et al., 2012a, *MNRAS*, 419, 3018
- Cooper M. C., et al., 2012b, *MNRAS*, 419, 3018
- Cortes C., Vapnik V., 1995, Support-Vector Networks. *Machine Learning* 20, 273-297.. <https://link.springer.com/article/10.1023/A:1022627411411>
- Costa-Duarte M. V., et al., 2019, arXiv e-prints, p. arXiv:1909.08626
- Coulter D. A., et al., 2017, *Science*, 358, 1556

- Cowie L. L., Barger A. J., Bautz M. W., Brandt W. N., Garmire G. P., 2003, *ApJ*, **584**, L57
- Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, *MNRAS*, **349**, 1397
- Croom S. M., et al., 2005, *MNRAS*, **356**, 415
- Croom S. M., et al., 2009, *MNRAS*, **399**, 1755
- DESI Collaboration et al., 2016, arXiv e-prints, p. [arXiv:1611.00036](https://arxiv.org/abs/1611.00036)
- Dalton G., 2016, WEAVE: The Next Generation Spectroscopy Facility for the WHT. p. 97
- Davies F. B., et al., 2018, *ApJ*, **864**, 143
- Del Moro A., et al., 2016, *MNRAS*, **456**, 2105
- Di Matteo T., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, *ApJ*, **676**, 33
- Díaz M. C., et al., 2017, *ApJ*, **848**, L29
- Diemer B., 2018, *ApJS*, **239**, 35
- Dodelson S., 2003, *Modern cosmology*
- Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, *MNRAS*, **433**, 3297
- Eftekharzadeh S., et al., 2015, *MNRAS*, **453**, 2779
- Emsellem E., Dejonghe H., Bacon R., 1999, *MNRAS*, **303**, 495
- Eriksen M., et al., 2019, *MNRAS*, **484**, 4200
- Eriksen M., et al., 2020, *MNRAS*, **497**, 4565
- Event Horizon Telescope Collaboration et al., 2019a, *ApJ*, **875**, L1
- Event Horizon Telescope Collaboration et al., 2019b, *ApJ*, **875**, L2
- Event Horizon Telescope Collaboration et al., 2019c, *ApJ*, **875**, L3
- Event Horizon Telescope Collaboration et al., 2019d, *ApJ*, **875**, L4
- Event Horizon Telescope Collaboration et al., 2019e, *ApJ*, **875**, L5

- Event Horizon Telescope Collaboration et al., 2019f, *ApJ*, **875**, L6
- Fan X., et al., 2006, *AJ*, **131**, 1203
- Fan X., et al., 2019, *Bulletin of the American Astronomical Society*, **51**, 121
- Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, **426**, 23
- Feldmann R., et al., 2006, *MNRAS*, **372**, 565
- Ferrarese L., Merritt D., 2000, *ApJ*, **539**, L9
- Fiore F., et al., 2012, *A&A*, **537**, A16
- Font-Ribera A., et al., 2014, *JCAP*, **2014**, 027
- Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., 1992, *ApJ*, **398**, 476
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, **111**, 1748
- Gaia Collaboration et al., 2018, *A&A*, **616**, A1
- Gardner J. P., et al., 2006, *Space Science Reviews*, **123**, 485
- Garilli B., et al., 2014, *A&A*, **562**, A23
- Gebhardt K., et al., 2000, *ApJ*, **543**, L5
- Genzel R., Thatte N., Krabbe A., Kroker H., Tacconi-Garman L. E., 1996, *ApJ*, **472**, 153
- Genzel R., Eisenhauer F., Gillessen S., 2010, *Reviews of Modern Physics*, **82**, 3121
- Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, *ApJ*, **715**, 823
- Gezari S., et al., 2017, *ApJ*, **835**, 144
- Ghez A. M., Klein B. L., Morris M., Becklin E. E., 1998, *ApJ*, **509**, 678
- Ghez A. M., Morris M., Becklin E. E., Tanner A., Kremenek T., 2000, *Nature*, **407**, 349
- Ghez A. M., et al., 2008, *ApJ*, **689**, 1044
- Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T., 2009a, *ApJ*, **692**, 1075

- Gillessen S., Eisenhauer F., Fritz T. K., Bartko H., Dodds-Eden K., Pfuhl O., Ott T., Genzel R., 2009b, *ApJ*, **707**, L114
- Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, *A&A*, **436**, 895
- Goldstein A., et al., 2017, *ApJ*, **848**, L14
- Green P. J., MacLeod C., Anderson S. F., Eracleous M., Ruan J. J., Runnoe J. C., Graham M. J., 2017, in American Astronomical Society Meeting Abstracts #229. p. 225.03
- Gunn J. E., Peterson B. A., 1965, *ApJ*, **142**, 1633
- Harms R. J., et al., 1994, *ApJ*, **435**, L35
- Harrison C. M., 2017, *Nature Astronomy*, **1**, 0165
- Hasinger G., Miyaji T., Schmidt M., 2005, *A&A*, **441**, 417
- Hearin A. P., Watson D. F., van den Bosch F. C., 2015, *MNRAS*, **452**, 1958
- Heckman T. M., 1980, *A&A*, **500**, 187
- Heckman T. M., Best P. N., 2014, *ARA&A*, **52**, 589
- Heintz K. E., Fynbo J. P. U., Høg E., Møller P., Krogager J. K., Geier S., Jakobsson P., Christensen L., 2018, *A&A*, **615**, L8
- Ho L. C., 2008, *ARA&A*, **46**, 475
- Ho S., et al., 2015, *JCAP*, **2015**, 040
- Hopkins P. F., Richards G. T., Hernquist L., 2007, *ApJ*, **654**, 731
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, *ApJS*, **175**, 356
- Hou J., et al., 2020, *MNRAS*,
- Hoyle B., et al., 2018, *MNRAS*, **478**, 592
- Ilbert O., et al., 2006, *A&A*, **457**, 841
- Ivashchenko G., Vasylenko O., Tugay A. V., 2011, arXiv e-prints, p. arXiv:1112.6030
- Ivezic Z., et al., 2008, *Serbian Astronomical Journal*, **176**, 1

- Jiang L., et al., 2016, *ApJ*, **833**, 222
- Jones D. H., et al., 2009, *MNRAS*, **399**, 683
- Joyce A., Lombriser L., Schmidt F., 2016, *Annual Review of Nuclear and Particle Science*, **66**, 95
- Kaiser N., 1987, *MNRAS*, **227**, 1
- Kaiser N., et al., 2002, in Tyson J. A., Wolff S., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4836, Survey and Other Telescope Technologies and Discoveries. pp 154–164, doi:10.1117/12.457365
- Kauffmann G., et al., 2003, *MNRAS*, **346**, 1055
- Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, **556**, 121
- Kirkpatrick J., 2012, in American Astronomical Society Meeting Abstracts #219. p. 440.03
- Kirkpatrick J. A., Schlegel D. J., Ross N. P., Myers A. D., Hennawi J. F., Sheldon E. S., Schneider D. P., Weaver B. A., 2011, *ApJ*, **743**, 125
- Kirkpatrick A., et al., 2020, *ApJ*, **900**, 5
- Kormendy J., Gebhardt K., 2001, in Wheeler J. C., Martel H., eds, American Institute of Physics Conference Series Vol. 586, 20th Texas Symposium on relativistic astrophysics. pp 363–381 (arXiv:astro-ph/0105230), doi:10.1063/1.1419581
- Kormendy J., Ho L. C., 2013, *ARA&A*, **51**, 511
- Kormendy J., Richstone D., 1995, *ARA&A*, **33**, 581
- Kormendy J., et al., 1996, *ApJ*, **459**, L57
- Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64
- Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193
- Laurent P., et al., 2017, *JCAP*, **2017**, 017
- Lawrence A., et al., 2007, *MNRAS*, **379**, 1599

- Lee K.-G., Suzuki N., Spergel D. N., 2012, *AJ*, 143, 51
- Lee K.-G., et al., 2013, *AJ*, 145, 69
- Leistedt B., Peiris H. V., 2014, *MNRAS*, 444, 2
- Leistedt B., Peiris H. V., Mortlock D. J., Benoit-Lévy A., Pontzen A., 2013, *MNRAS*, 435, 1857
- Levi M., et al., 2019, in *Bulletin of the American Astronomical Society*. p. 57
([arXiv:1907.10688](https://arxiv.org/abs/1907.10688))
- Levin Y., Beloborodov A. M., 2003, *ApJ*, 590, L33
- Lewis A., Challinor A., Lasenby A., 2000, *ApJ*, 538, 473
- Lilly S. J., et al., 2007, *ApJS*, 172, 70
- Lima M., Hu W., 2007, *Phys. Rev. D*, 76, 123013
- Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, 390, 118
- Loh E. D., Spillar E. J., 1986, *ApJ*, 303, 154
- Lynden-Bell D., 1969, *Nature*, 223, 690
- Maarten Schmidt M., 1963, *Nature*, 197, 1040
- Mallmann N. D., et al., 2018, *MNRAS*, 478, 5491
- Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, *MNRAS*, 351, 169
- Marín-Franch A., et al., 2017, in *Arribas S., Alonso-Herrero A., Figueras F., Hernández-Monteagudo C., Sánchez-Lavega A., Pérez-Hoyos S., eds, Highlights on Spanish Astrophysics IX*. pp 670–675
- Matsuoka Y., et al., 2018, *ApJ*, 869, 150
- Matute I., et al., 2012, *A&A*, 542, A20
- Mendes de Oliveira C., et al., 2019, *MNRAS*, 489, 241

- Miyoshi M., Moran J., Herrnstein J., Greenhill L., Nakai N., Diamond P., Inoue M., 1995, *Nature*, **373**, 127
- Moles M., et al., 2008, in Knapen J. H., Mahoney T. J., Vazdekis A., eds, *Astronomical Society of the Pacific Conference Series Vol. 390, Pathways Through an Eclectic Universe*. p. 495
- Molino A., et al., 2014, *MNRAS*, **441**, 2891
- Molino A., et al., 2019a, arXiv e-prints, p. [arXiv:1907.06315](https://arxiv.org/abs/1907.06315)
- Molino A., et al., 2019b, *A&A*, **622**, A178
- Morganson E., et al., 2012, *AJ*, **143**, 142
- Morganti R., 2017, *Nature Astronomy*, **1**, 596
- Myers A. D., et al., 2006, *ApJ*, **638**, 622
- Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007, *ApJ*, **658**, 85
- Myers A. D., et al., 2015, *ApJS*, **221**, 27
- Narayan R., Yi I., 1994, *ApJ*, **428**, L13
- Nenkova M., Sirocky M. M., Ivezić Ž., Elitzur M., 2008a, *ApJ*, **685**, 147
- Nenkova M., Sirocky M. M., Nikutta R., Ivezić Ž., Elitzur M., 2008b, *ApJ*, **685**, 160
- Netzer H., 2013, *The Physics and Evolution of Active Galactic Nuclei*
- Netzer H., 2015, *ARA&A*, **53**, 365
- Newman J. A., et al., 2013, *ApJS*, **208**, 5
- Nikutta R., Hunt-Walker N., Nenkova M., Ivezić Ž., Elitzur M., 2014, *MNRAS*, **442**, 3361
- Nunes R. C., Pacucci F., 2020, *MNRAS*, **496**, 888
- Pacucci F., Natarajan P., Volonteri M., Cappelluti N., Urry C. M., 2017, *ApJ*, **850**, L42
- Padovani P., 2017, *Nature Astronomy*, **1**, 0194

- Palanque-Delabrouille N., et al., 2016, *A&A*, **587**, A41
- Pâris I., et al., 2011, *A&A*, **530**, A50
- Parkinson D., et al., 2012, *Phys. Rev. D*, **86**, 103518
- Peacock J. A., 1999, *Cosmological Physics*
- Peacock J. A., Smith R. E., 2000, *MNRAS*, **318**, 1144
- Peacock J. A., et al., 2001, *Nature*, **410**, 169
- Peebles P. J. E., 1980, *The large-scale structure of the universe*
- Penzias A. A., Wilson R. W., 1965, *ApJ*, **142**, 419
- Percival W. J., White M., 2009, *MNRAS*, **393**, 297
- Percival W. J., Verde L., Peacock J. A., 2004, *MNRAS*, **347**, 645
- Pérez-González P. G., et al., 2013, *ApJ*, **762**, 46
- Pérez-Ràfols I., Pieri M. M., 2019, arXiv e-prints, p. [arXiv:1911.04891](https://arxiv.org/abs/1911.04891)
- Pérez-Ràfols I., Pieri M. M., Blomqvist M., Morrison S., Som D., 2019, arXiv e-prints, p. [arXiv:1903.00023](https://arxiv.org/abs/1903.00023)
- Perlmutter S., Turner M. S., White M., 1999, *PhysRevLett.*, **83**, 670
- Peters C. M., et al., 2015, *ApJ*, **811**, 95
- Pieri M. M., et al., 2016, in SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 259–266 ([arXiv:1611.09388](https://arxiv.org/abs/1611.09388))
- Planck Collaboration et al., 2020a, *A&A*, **641**, A6
- Planck Collaboration et al., 2020b, *A&A*, **641**, A9
- Porciani C., Norberg P., 2006, *MNRAS*, **371**, 1824
- Porciani C., Magliocchetti M., Norberg P., 2004, *MNRAS*, **355**, 1010
- Press W. H., Schechter P., 1974, *ApJ*, **187**, 425
- Pullen A. R., Hirata C. M., 2013, *Publ. Astron. Soc. Pac*, **125**, 705

- Rau A., et al., 2009, *Publ. Astron. Soc. Pac*, 121, 1334
- Rees M. J., 1984, *ARA&A*, 22, 471
- Richard J., et al., 2019, *The Messenger*, 175, 50
- Richards G. T., et al., 2001, *AJ*, 122, 1151
- Richards G. T., et al., 2002, *AJ*, 123, 2945
- Richards G. T., et al., 2005, *MNRAS*, 360, 839
- Richards G. T., et al., 2006a, *AJ*, 131, 2766
- Richards G. T., et al., 2006b, *ApJS*, 166, 470
- Richards G. T., et al., 2009, *ApJS*, 180, 67
- Richardson J., Zheng Z., Chatterjee S., Nagai D., Shen Y., 2012, *ApJ*, 755, 30
- Riess A. G., et al., 1998, *AJ*, 116, 1009
- Robin A. C., Reyl  C., Derri re S., Picaud S., 2003, *A&A*, 409, 523
- Ross N. P., et al., 2009, *ApJ*, 697, 1634
- Ross N. P., et al., 2012, *ApJS*, 199, 3
- Ross N. P., et al., 2013, *ApJ*, 773, 14
- Sadeh I., Abdalla F. B., Lahav O., 2016, *Publ. Astron. Soc. Pac*, 128, 104502
- Saikia D. J., Jamrozy M., 2009, *Bulletin of the Astronomical Society of India*, 37, 63
- Salim S., 2014, *Serbian Astronomical Journal*, 189, 1
- Salpeter E. E., 1964, *ApJ*, 140, 796
- Salvato M., et al., 2009, *ApJ*, 690, 1250
- Schechter P., 1976, *ApJ*, 203, 297
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schmidt K. B., Marshall P. J., Rix H.-W., Jester S., Hennawi J. F., Dobler G., 2010, *ApJ*, 714, 1194

- Schneider P., 2006, *Extragalactic Astronomy and Cosmology*
- Schneider D. P., et al., 2003, *AJ*, **126**, 2579
- Schweitzer M., et al., 2008, *ApJ*, **679**, 101
- Scoville N., et al., 2007b, *The Astrophysical Journal Supplement Series*, **172**, 1
- Scoville N., et al., 2007a, *ApJS*, **172**, 1
- Seljak U., 2000, *MNRAS*, **318**, 203
- Serber W., Bahcall N., Ménard B., Richards G., 2006, *ApJ*, **643**, 68
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, **500**, 33
- Shankar F., Dai X., Sivakoff G. R., 2008, *ApJ*, **687**, 859
- Shen Y., et al., 2007, *AJ*, **133**, 2222
- Shen Y., et al., 2013, *ApJ*, **778**, 98
- Sherwin B. D., et al., 2012, *Phys. Rev. D*, **86**, 083006
- Siana B., et al., 2008, *ApJ*, **675**, 49
- Silk J., Mamon G. A., 2012, *Research in Astronomy and Astrophysics*, **12**, 917
- Simcoe R. A., Sullivan P. W., Cooksey K. L., Kao M. M., Matejek M. S., Burgasser A. J., 2012, *Nature*, **492**, 79
- Skrutskie M. F., et al., 2006, *AJ*, **131**, 1163
- Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, *JCAP*, **2008**, 031
- Smoot G. F., et al., 1992, *ApJ*, **396**, L1
- Spergel D., et al., 2015, arXiv e-prints, p. [arXiv:1503.03757](https://arxiv.org/abs/1503.03757)
- Stacy A., Bromm V., Lee A. T., 2016, *MNRAS*, **462**, 1307
- Statler T. S., King I. R., Crane P., Jedrzejewski R. I., 1999, *AJ*, **117**, 894
- Stoughton C., et al., 2002, *AJ*, **123**, 485
- Strand N. E., Brunner R. J., Myers A. D., 2008, *ApJ*, **688**, 180

- Su M., Slatyer T. R., Finkbeiner D. P., 2010, *ApJ*, 724, 1044
- Suzuki N., Tytler D., Kirkman D., O'Meara J. M., Lubin D., 2005, *ApJ*, 618, 592
- Takada M., et al., 2014, *PASJ*, 66, R1
- Taylor K., et al., 2014, *Journal of Astronomical Instrumentation*, 3, 1350010
- Tegmark M., 1997, *Phys. Rev. D*, 55, 5895
- Tegmark M., Hamilton A. J. S., Strauss M. A., Vogeley M. S., Szalay A. S., 1998, *ApJ*, 499, 555
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Urrutia T., et al., 2019, *A&A*, 624, A141
- Urry C. M., Padovani P., 1995, *Publ. Astron. Soc. Pac*, 107, 803
- Vale A., Ostriker J. P., 2006, *MNRAS*, 371, 1173
- Valiante R., Agarwal B., Habouzit M., Pezzulli E., 2017, *Publications of the Astronomical Society of Australia*, 34, e031
- Vanden Berk D. E., et al., 2001, *AJ*, 122, 549
- Vanzella E., et al., 2004, *A&A*, 423, 761
- Venemans B. P., 2017, *The Messenger*, 169, 48
- Wang F., et al., 2019, *ApJ*, 884, 30
- Werner M. W., et al., 2004, *ApJS*, 154, 1
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- White R. L., et al., 2000, *ApJS*, 126, 133
- White M., et al., 2012, *MNRAS*, 424, 933
- Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, *ApJ*, 691, 1879
- Wolf C., Meisenheimer K., Rix H. W., Borch A., Dye S., Kleinheinrich M., 2003, *A&A*, 401, 73

- Wolf C., et al., 2004, *A&A*, 421, 913
- Woltjer L., 1959, *ApJ*, 130, 38
- Wright E. L., et al., 2010, *AJ*, 140, 1868
- Xavier H. S., et al., 2014, *MNRAS*, 444, 2313
- Yèche C., et al., 2010, *A&A*, 523, A14
- Yip C. W., et al., 2004, *AJ*, 128, 2603
- York D. G., et al., 2000, *AJ*, 120, 1579
- Yuan F., Narayan R., 2014, *ARA&A*, 52, 529
- Zandivarez A., Díaz-Giménez E., Mendes de Oliveira C., Ascaso B., Benítez N., Dupke R., Sodr e L., Irwin J., 2014, *A&A*, 561, A71
- Zel'dovich Y. B., 1964, *Soviet Physics Doklady*, 9, 195
- Zheng Z., et al., 2005, *ApJ*, 633, 791
- Zheng Z., Coil A. L., Zehavi I., 2007, *ApJ*, 667, 760
- Zubovas K., King A. R., 2019, *General Relativity and Gravitation*, 51, 65
- da  ngela J., et al., 2008, *MNRAS*, 383, 565

APPENDIX A

TABLES

Table A.1: Best-fit parameters for the five and six-parameter models of quasar HOD at $\bar{z} = 0.5$. Adapted from Table 4 from [Shen et al. \(2013\)](#).

5-parameter model		6-parameter model	
$\log M_{min}$	$19.46^{+0.61}_{-0.64}$	$\log M_{cen}$	$13.57^{+4.92}_{-1.41}$
$\sigma_{\log M}$	$2.73^{+0.20}_{-0.21}$	σ_M	$0.91^{+0.82}_{-0.62}$
		$\log f_{cen}$	$-3.13^{+2.10}_{-0.46}$
$\log M_0$	$12.74^{+0.86}_{-1.05}$	$\log M_0$	$12.53^{+0.88}_{-1.02}$
$\log M_1$	$16.24^{+0.81}_{-0.51}$	$\log M_1$	$16.13^{+0.73}_{-0.40}$
α	$1.19^{+0.37}_{-0.33}$	α	$1.21^{+0.29}_{-0.33}$

Table A.2: Summary of the masses of high redshift quasars. Adapted from: [Nunes & Pacucci \(2020\)](#).

Name	M_{\bullet} (M_{\odot})	z
J1342 + 0928	$(7.8^{+3.3}_{-1.9}) \times 10^8$	7.54
J1007 + 2115	$(1.5 \pm 0.2) \times 10^9$	7.515
J1243 + 0100	$(3.3 \pm 0.2) \times 10^8$	7.07
J1120 + 0641	$(2.0^{+1.5}_{-0.7}) \times 10^9$	7.085
J0038 – 1527	$(1.33 \pm 0.25) \times 10^9$	7.021
J2348 – 3054	$(2.1 \pm 0.5) \times 10^9$	6.90
J0109 – 3047	$(1.5 \pm 0.4) \times 10^9$	6.80
J0305 – 3150	$(9.5^{+0.8}_{-0.7}) \times 10^8$	6.61
P036 + 03	$(1.9^{+1.1}_{-0.8}) \times 10^9$	6.54

Table A.3: Main emission lines of quasars. All wavelengths are given in Å at restframe. Here are only shown the spectral lines with non-zero weights for quasars observed with SDSS. Adapted from the table of spectral lines used in SDSS.

Wavelength	Line
1033.03	Ly β
1033.82	OVI
1215.24	Ly α
1240.81	NV
1399.80	SiIV+OIV
1549.48	CIV
1908.734	CIII]
2326.00	CII
2799.117	MgII
3727.092	OII
4102.89	H Δ
4341.68	H γ
4862.68	H β
4960.295	OIII
5008.24	OIII
6564.61	H α
6585.27	NII
6718.29	SII
6732.67	SII

[<http://classic.sdss.org/dr6/algorithms/linestable.html>]

Table A.4: Summary of J-PAS filters. All wavelengths are given in Å.

Filter	λ_c	$\Delta\lambda$	Filter	λ_c	$\Delta\lambda$	Filter	λ_c	$\Delta\lambda$
uJAVA	3497	509	J0550	5498	149	J0740	7414	148
uJPAS	3623	736	J0560	5596	150	J0750	7502	142
J0378	3782	157	J0570	5701	150	J0760	7602	144
J0390	3904	150	J0580	5803	148	J0770	7719	146
J0400	3996	148	J0590	5917	152	iSDSS	7725	1250
J0410	4110	144	J0600	6010	150	J0780	7811	145
J0420	4203	147	J0610	6107	150	J0790	7907	142
J0430	4303	148	J0620	6206	148	J0800	8009	140
J0440	4403	152	rSDSS	6250	1250	J0810	8124	144
J0450	4503	150	J0630	6309	150	J0820	8226	143
J0460	4603	148	J0640	6408	150	J0830	8329	148
J0470	4701	148	J0650	6506	146	J0840	8429	148
gSDSS	4750	1250	J0660	6607	151	J0850	8523	146
J0480	4799	142	J0670	6710	146	J0860	8620	148
J0490	4902	154	J0680	6812	152	J0870	8716	146
J0500	5002	152	J0690	6912	148	J0880	8810	146
J0510	5097	148	J0700	7007	148	J0890	8912	150
J0520	5202	150	J0710	7119	148	J0900	9000	154
J0530	5296	150	J0720	7207	146	J0910	9107	152
J0540	5389	152	J0730	7307	150	J1007	9316	635