



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE ASTRONOMÍA

VARIABILIDAD DE AGN EN LA ERA DE LAS GRANDES BASES DE DATOS

TESIS PARA OPTAR AL GRADO DE
DOCTORA EN CIENCIAS, MENCIÓN ASTRONOMÍA

PAULA ANDREA SÁNCHEZ SÁEZ

PROFESOR GUÍA:
PAULINA MARÍA LIRA TEILLERY

MIEMBROS DE LA COMISIÓN:
PATRICIA ARÉVALO NOORDAM
FRANCISCO FÖRSTER BURÓN
RICARDO MUÑOZ VIDAL
PATRICIO ROJO RUBKE

Este trabajo ha sido parcialmente financiado por CONICYT Beca Doctorado Nacional año 2013 grant no. 21130441, y CONICYT-PFB-06

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE DOCTORA EN CIENCIAS, MENCIÓN ASTRONOMÍA
POR: PAULA ANDREA SÁNCHEZ SÁEZ
FECHA: 2019
PROF. GUÍA: PAULINA MARÍA LIRA TEILLERY

VARIABILIDAD DE AGN EN LA ERA DE LAS GRANDES BASES DE DATOS

Los núcleos activos de galaxias o AGN por su sigla en inglés, se caracterizan por su continuo variable en el tiempo en cada banda de frecuencias en la que se han estudiado (Peterson, 2001). Sin embargo, aún no entendemos por completo los mecanismos que conducen a tales variaciones. En particular, no entendemos en detalle cómo se relaciona la variabilidad de AGN en diferentes longitudes de onda, y cómo las propiedades físicas del motor central están relacionadas con las propiedades de variabilidad del sistema.

El principal objetivo de esta tesis es entender de mejor manera la física que hay detrás de la variabilidad de los AGN, y proporcionar una técnica de selección de AGN basada en su variabilidad, para así poder obtener muestras de AGN menos sesgadas, que nos permitan entender de mejor manera como se forman y como evolucionan estos objetos.

En primer lugar, se realizó un estudio estadístico de la variabilidad infrarroja cercana (NIR) de AGN seleccionados por rayos X en el campo COSMOS, utilizando datos de UltraVISTA. En este trabajo, estudiamos las diferencias entre las propiedades de variabilidad de distintas poblaciones de AGN. Los resultados de este estudio fueron publicados en el artículo Sánchez et al. (2017), y son presentados en el capítulo 2.

En segundo lugar, estudiamos la relación que hay entre las propiedades físicas de los AGN (como la masa del agujero negro central, su luminosidad y su tasa de acreción) con las propiedades de variabilidad de estos objetos. En este estudio, encontramos que la amplitud de la variabilidad depende principalmente de la tasa de acreción del objeto. Los resultados de este análisis fueron publicados en el artículo Sánchez-Sáez et al. (2018), y son presentados en el capítulo 3.

En tercer lugar, realizamos una selección de candidatos a AGN, utilizando análisis de variabilidad y aprendizaje de máquinas. Nuestro método es capaz de seleccionar candidatos a AGN que no son normalmente encontrados por métodos tradicionales, que usan el color de los objetos para la selección. Los resultados de este estudio fueron enviados a la revista “Astrophysical Journal Supplement”, y son presentados en el capítulo 4.

Finalmente, realizamos un mapeo de reverberación del toro de polvo, utilizando datos de UltraVISTA. En este trabajo calculamos el retraso temporal entre curvas de luz asociadas a emisión del disco y del toro de distintos AGN. A partir de nuestros resultados, concluimos que para objetos más distantes ($0.5 < z < 1.2$) no se observan las mismas correlaciones entre el retraso temporal y la luminosidad de los objetos obtenidas para objetos locales, lo cual se podría explicar por la existencia de granos de polvo más grandes en el toro de polvo de AGN lejanos. Los resultados de este estudio serán incluidos en un artículo que está en preparación, y son presentados en el capítulo 5.

SUMMARY OF THE THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN
ASTRONOMY

AGN VARIABILITY IN THE ERA OF BIG DATA

Active Galactic Nuclei (AGN) are characterized by their time-variable continuum flux in every waveband in which they have been studied (Peterson, 2001). However, we do not completely understand the mechanisms that drive such variations. In particular, we are missing significant details of how AGN variability at different wavelengths is related, and how physical properties of the central engine are related to well defined variability properties of the system.

The main goal of this thesis is to better understand the physics behind AGN variability, and to provide a variability-based AGN selection technique, in order to obtain less biased AGN samples, which can allow us to understand in a better way how these objects are formed and how they evolve.

First, we performed a statistical study of the near infrared (NIR) variability of X-ray-selected AGN in the COSMOS field, using data from the UltraVISTA survey. In this work, we studied the differences between the variability properties of different AGN populations. The results of this study were published in the article Sánchez et al. (2017), and are presented in Chapter 2.

Second, we study the relationship between the physical properties of AGN (such as the black hole mass, the luminosity, and the accretion rate or the Eddington ratio) with the variability properties of these objects. In this study, we found that the amplitude of the variability depends primarily on the Eddington ratio. The results of this analysis were published in the article Sánchez-Sáez et al. (2018), and are presented in Chapter 3.

Third, we made a selection of AGN candidates using variability analysis and machine learning techniques. Our method is able to select AGN candidates that are not normally found by traditional methods, which use the color of the objects for the selection of candidates. The results of this study were submitted to the “Astrophysical Journal Supplement”, and are presented in Chapter 4.

Finally, we performed a dust reverberation mapping analysis, using data from the UltraVISTA survey. In this work we calculated the time delay between light curves associated with emission from disk and the torus, for different AGN. From our results, we conclude that for distant objects ($0.5 < z < 1.2$) the correlations between the time delay and the AGN luminosity obtained for local objects are not observed, which could be explained by the presence of dust composed by larger grains compared to the typical grain size values used in the literature. The results of this study will be included in an article that is in preparation, and are presented in Chapter 5.

If you can't love yourself, how the hell you gonna love somebody else...

Agradecimientos

Que difícil agradecer en una sola página a todas las personas que me han acompañado en estos 6 años de doctorado. Han sido muchas las experiencias vividas en este tiempo, y han sido muchos los que han estado conmigo en este proceso, por lo que podría decir que la parte más difícil de escribir en esta tesis fue esta sección de agradecimientos.

En primer lugar, quiero agradecerle a Paulina. Sin su guía esta tesis no habría sido posible. Quiero agradecerle por haber sido una guía no solo en lo académico, por apoyarme en varios de los momentos más difíciles de mi doctorado, y por haberme dado la libertad de desarrollar mis propias ideas. También quiero agradecerle a Patricia, quien si bien no fue mi supervisora oficialmente, siempre estuvo disponible para darme su opinión respecto a mi trabajo. Además, quisiera agradecer a Luis y Konrad, que fueron mis supervisores en dos de mis pasantías, y a los miembros del comité, Francisco, Pato y Ricardo, quienes ayudaron a que esta tesis fuera exitosa.

Quiero agradecerle a Chichi, la persona más importante en mi vida, por haber estado siempre conmigo en todo este proceso. No hay palabras que puedan expresar lo fundamental que fue contar con él, no solo durante este doctorado, si no que durante estos 12 años que hemos estado juntos. Si logré llevar esta tesis a término fue porque él me dio su empuje constantemente y me hizo ver que valía la pena luchar por las cosas que amo. También quiero agradecerle a mi familia, Rosa, Gabriel, Macarena, y Lito, y a mis hermosos sobrinos, por mostrarme desde pequeña que mi camino era el de la ciencia, por darme la fuerza necesaria para afrontar este doctorado, y por darme la tranquilidad que necesité muchas veces. Quiero agradecerles también a mis amigos del colegio, Susana, Pame, Adrian, Nikolee, Ely, Mauri, y Alondra, y a mis amigos Pablo, Nico, Pancho, Isa, Kike y David, que si bien no fueron parte directamente de este doctorado, fueron fundamentales en mantenerme a flote en todo este proceso.

También quiero agradecerles a todos mis compañeros de Calan, en particular a los de la oficina Pequeños Placeres: Mátias, Coni, Elise, Mari, José, Bica, Paula, Sрни, y Vachail. Siempre tendré en mi corazón los buenos y malos momentos que vivimos todos juntos, los cuales fueron muchos. Extrañaré mucho nuestras conversaciones en la oficina y nuestras juntas. Además, quisiera agradecer a Richi, Seba, Julian, Regis, Sudeep, Valentin, Blake, Jorge, Grecco, Juanpi, Patits, Nina y Matías, a todas las chicas de Cazadoras de Estrellas, y a Natalie, Marta y Alejandro. Hicieron que mi estadía en Calan fuera muy entretenida y mucho más llevadera. Quiero agradecer a todos mis compañeros de ESO, tanto estudiantes, como fellows y staffs. Hicieron que mi último año del doctorado fuera inolvidable.

Por último, quiero agradecerle Rupaul y a todas las drag queens de Rupaul's Drag Race, a quienes quizás no conozco aún personalmente, pero fueron de gran ayuda en todo este proceso. Me dieron alegrías cuando más lo necesitaba, y me mostraron que no es posible amar a nadie si no te amas a ti mismo.

Contents

List of Tables	xii
List of Figures	xiv
1 Introduction	1
1.1 Active Galactic Nuclei	1
1.2 AGN variability	2
1.2.1 Origins of the variations	4
1.2.2 Time series analysis	6
1.3 Data sets	9
1.3.1 The QUEST-La Silla AGN variability survey	9
1.3.2 UltraVISTA survey	10
1.4 This thesis	11
2 Near Infrared Variability of obscured and unobscured X-ray selected AGN in the COSMOS field	12
2.1 Introduction	12
2.2 Data	13
2.2.1 NIR data	13
2.2.2 Ancillary data	13
2.3 AGN Classification	15
2.3.1 Spectroscopic Classification	15
2.3.2 Photometric Classification	15
2.3.3 X-ray Classification	16
2.3.4 Radio Classification	16
2.4 Light Curve Construction	17
2.4.1 Calibration and Photometry	17
2.4.2 Light Curve Generation	18
2.5 Variability Analysis	23
2.5.1 The Structure Function	23
2.5.2 Damp Random Walk Process	23
2.6 Results	25
2.6.1 Variability properties	26
2.6.2 Dependency with redshift and Luminosity	29
2.6.3 Fraction of variable sources	36
2.7 Discussion and Conclusions	40

3	The QUEST–La Silla AGN variability survey: connection between AGN variability and black hole physical properties	43
3.1	Introduction	43
3.2	Data	45
3.2.1	Optical light curves	45
3.2.2	SDSS spectra	45
3.3	Variability analysis	46
3.3.1	Variability features	46
3.3.2	Sample filtering by light curves properties	46
3.3.3	Biases of the variability features	47
3.3.4	Variability of simulated light curves	47
3.3.5	Variability of the QUEST–La Silla light curves	54
3.4	Spectral analysis	56
3.4.1	Host galaxy subtraction	56
3.4.2	Spectral fitting and measurement of physical properties	56
3.4.3	Spectral properties of the selected SDSS spectra	58
3.5	Final sample definition	59
3.6	Variability parameters versus physical properties	62
3.6.1	Bivariate correlations	62
3.6.2	Principal Component Analysis	63
3.6.3	Multiple Linear Regression	66
3.6.4	Differences between variable and non–variable sources	72
3.7	Variability behavior of different classes of AGN	75
3.7.1	BAL–QSO	75
3.7.2	Radio Classification	75
3.8	Discussion and Conclusions	76
4	The QUEST–La Silla AGN Variability Survey: selection of AGN candidates through optical variability	81
4.1	Introduction	81
4.2	Light curve construction	83
4.3	Selection of AGN candidates	84
4.3.1	Random Forests	84
4.3.2	Variability features	85
4.3.3	Labeled set	86
4.3.4	Performance of the Random Forest classifier	88
4.3.5	AGN candidates from QUEST–La Silla	95
4.4	Confirmation of AGN candidates	96
4.4.1	Confirmation by ancillary data	99
4.4.2	Spectroscopic follow–up of AGN candidates	101
4.5	Comparison with previous works	105
4.6	Conclusions	107
4.7	Appendix	108
4.7.1	Catalog of observed candidates	108
4.7.2	Spectroscopic analysis of the observed candidates	109
5	Dust Reverberation Mapping of $z < 1.2$ AGN	112

5.1	Introduction	112
5.2	Data	114
5.3	Sample Selection	114
5.4	SED Analysis	115
5.5	Subtraction of the Accretion Disk Component from the rest-frame NIR emission	116
5.6	Cross-correlation Analysis	119
	5.6.1 Challenges of the Cross-correlation Analysis	119
	5.6.2 ICCF	119
	5.6.3 Lag-significance Criteria	120
	5.6.4 Cross-correlation Results	122
5.7	Correlations between Time Lags and Optical Luminosity	123
5.8	Conclusions and Future Work	128
6	Conclusions	129
	Bibliography	132

List of Tables

1	Number of well-sampled light curves before and after downsizing: clean-sample / whole data set. In brackets we show the number of variable sources.	28
2	Results of the WLS. <i>p-values</i> in brackets.	32
3.1	PCA results: Eigenvectors and Eigenvalues.	65
3.2	PCA results: Spearman correlation coefficients between the input variables and the four first principal components (PC1, PC2, PC3, and PC4). the p-values of the coefficients are given in parentheses.	65
3.3	Linear Regression α , β and ε coefficients for A as the dependent variable (for the not – C IV sample). The columns headed by physical quantities refer to their slope in the regression model (β).	66
3.4	Linear Regression α , β and ε coefficients for A as the dependent variable (for standard black hole masses). The columns headed by physical quantities refer to their slope in the regression model (β).	68
3.5	Linear Regression α , β and ε coefficients for A as the dependent variable (for corrected black hole masses). The columns headed by physical quantities refer to their slope in the regression model (β).	69
3.6	Linear Regression α , β and ε coefficients for A as the dependent variable, for spectral properties derived from Mg II. The columns headed by physical quantities refer to their slope in the regression model (β).	70
3.7	Linear Regression α , β and ε coefficients for A as the dependent variable (spectral properties per emission line). The columns headed by physical quantities refer to their slope in the regression model (β).	70
3.8	Linear Regression α , β and ε coefficients for other amplitude features as dependent variables (not – C IV sample). The columns headed by physical quantities refer to their slope in the regression model (β).	71
3.9	Linear Regression α , β and ε coefficients for γ as the dependent variable (not – C IV). The columns headed by physical quantities refer to their slope in the regression model (β).	72
4.1	Number of light curves per field.	84
4.2	List of features.	87
4.3	Scores measured in the test set for each classifier	92
4.4	Feature importance for each classifier.	93
4.5	Number of AGN candidates per field, for each classifier	96
4.6	Number of hp-AGN candidates confirmed using MILLIQUAS, for each classifier	101

4.7	Summary of the spectroscopic follow-up campaign	102
4.8	Targets observed during spectroscopic follow-up.	108
5.1	Results for the good-lag sample	124

List of Figures

1.1	Sketch of the main AGN structures seen along the equatorial and polar direction. From the center to host-galaxy scales: SMBH, accretion disk and corona, BLR, torus and NLR. Different colours indicate different compositions or densities. Credits: Ramos Almeida & Ricci (2017).	2
1.2	Optical (5100 Å) light curve of NGC 5548, taken from Peterson (2001).	3
1.3	Left: sketch of the main AGN structures seen along the equatorial and polar direction (as in Figure 1.1, credits: Ramos Almeida & Ricci 2017). Right: observed X-ray, optical and NIR light curves for MCG-6-30-15, from Lira et al. (2015). The arrows show the AGN regions associated to every light curve.	4
2.1	Observed light curve for a BL - XR II - TypeI - RQ AGN located at RA= 150.45187° and DEC= 2.144811°, with ID cid_543 from Marchesi et al. (2016a) catalog, located at z=1.298, variable in all the filter bands.	19
2.2	Observed light curve for a NL - XR II - Galaxy - RQ AGN located at RA= 150.097790° and DEC= 1.845247°, with ID cid_254 from Marchesi et al. (2016a) catalog, located at z=0.711, variable in all the filter bands.	20
2.3	Observed light curve for a XR II - Galaxy - RL AGN located at RA= 149.43076° and DEC= 1.939061°, with ID lid_2414 from Marchesi et al. (2016a) catalog, located at z=0.916, non variable in all the filter bands.	20
2.4	Mean magnitude error vs mean magnitude for all the final light curves, for the four photometric bands Y (red), J (blue), H (green) and Ks (black).	21
2.5	Histogram in logarithmic scale of the number of epochs in the light curves for the four photometric bands Y (red), J (blue), H (green) and Ks (black).	21
2.6	Top: Histogram in logarithmic scale of the rest-frame time length (t_{rest}), defined as the light curve's observed length divided by $(1+z)$, for the four photometric bands Y (red), J (blue), H (green) and Ks (black). Bottom: Cumulative distribution of the rest-frame time length. The yellow dashed vertical line marks $t_{rest} = 200$	22

2.7	Results of the Structure Function analysis for the artificial light curves with power-law PSD. In the x-axis we show the input value of γ and in the y-axis we show the value of γ computed by the different methods. The error bars in the y-axis correspond to the 15.86 and 84.14 percentiles of the output γ distributions. The blue circles correspond to the results obtained using the definition given in Kozłowski (2016), the red stars correspond to the first definition given by Schmidt et al. (2010) (eq. 1.11), and the yellow squares correspond to the Bayesian method proposed by Schmidt et al. (2010). The left panel corresponds to light curves without photometric noise. The right panel correspond to light curves with a photometric noise of 0.02 magnitudes. Light curves with “NAN” values of the γ parameter are not included.	24
2.8	Comparison of r_{in} vs r_{out} for simulated DRW light curves. The blue line shows the 1:1 relation. The red dotted line shows the region where the length of the light curve is 10 times the input τ	25
2.9	Normalized histogram of the redshift distribution of the variable and non variable sources in the Y band for the four classes defined in Table 1. The green dashed lines show the 0.3 to 1 redshift bin. From top to bottom: spectroscopic, photometric, X-ray and radio classifications.	27
2.10	Distribution of the Structure Function power law parameters A and γ , for the variable sources, in the Y and J bands, in logarithmic space. The blue sources correspond to BL AGN, the red sources to NL AGN, and the yellow to sources without spectroscopic classification. The circles correspond to XR I AGN, the triangles to XR II AGN, and the stars to sources without X-ray classification. The error of the measurements are shown with grey error bars. For most of the sources, the size of the error bars is smaller than the marker size. Along the axes we show the projected A and γ distributions for the BL AGN (blue shaded), NL AGN (red shaded), XR I AGN (blue hatched), and XR II AGN (red hatched).	30
2.11	Amplitude of the Structure Function A vs rest-frame wavelength of emission (λ) at logarithmic scale for the variable sources, in the Y (red circles) and J (blue crosses) bands. The blue and red lines show the linear regression between $\log_{10}(A)$ and $\log_{10}(\lambda)$ for the Y and J bands respectively. The shaded regions show the 95% error of the regression.	33
2.12	Top: absolute magnitude (no k-corrected) vs redshift for the Y band. The small shaded sources are non-variable. The blue sources corresponde to BL AGN, the red sources to NL AGN, and the green to sources without spectroscopic classification. The circles corresponde to XR I AGN, the triangles to XR II AGN, and the stars to sources without X-ray classification. As a reference, we include in the top x axis the value corresponding to the rest-frame wavelength of emission. Bottom: fraction of variable BL (blue) and NL (red) sources.	34
2.13	Histogram of the intrinsic luminosity (rest-frame and corrected for absorption) in the hard X-ray band, of variable sources in the Y band, considering the spectroscopic classification. BL are showed in blue and NL in red.	35

2.14	Histogram of the intrinsic luminosity (rest-frame and corrected for absorption) in the hard X-ray band, of variable NL sources in the Y band, split by their photometric classification. NL-TypeI are showed in blue, NL-TypeII in red, and NL-Gal in yellow.	36
2.15	Fraction of variable sources for the spectroscopic (a), photometric (b), X-ray (c), and radio (d) classifications, for our four photometric bands. The blue dashed line demarks the 5%, below this fraction we might expect to have several false positive variable sources. The error bars were calculated considering poisson statistic (Gehrels, 1986)	37
2.16	Fraction of variable sources for the spectroscopic/X-ray classification, and for the photomeric/X-ray classification. The blue dashed line demarks the 5%, below this fraction we might expect to have several false positive variable sources. The error bars were calculated considering poisson statistic (Gehrels, 1986)	39
3.1	Number of epochs vs. rest-frame time length of the 2345 light curves with SDSS spectra. The red vertical dashed line shows the position where $t_{\text{rest}} \equiv 200$ days, and the red horizontal dashed line shows the position where $\#\text{epochs} \equiv 40$	47
3.2	Variability features vs. light curve properties. The Spearman's rank correlation coefficient is shown as reference for every pair of variables.	48
3.3	Light curves with different number of epochs and length, but similar cadence.	49
3.4	SF parameters measured for 1000 light curves simulated from a DRW process with $\tau = 300$ days and $\text{SF}_{\text{inf}} = 0.2$ mag, with short and long samplings. Along the axes we show the histograms of every parameter.	50
3.5	SF parameters measured for light curves simulated from a DRW process with $\tau = 300$ days and different values of SF_{inf} (with the long sampling). The blue circles correspond to light curves simulated with $\text{SF}_{\text{inf}} = 0.05$, the red triangles correspond to light curges simulated with $\text{SF}_{\text{inf}} = 0.1$, and the yellow squares correspond to light curves simulated with $\text{SF}_{\text{inf}} = 0.2$. Along the axes we show the histograms of every parameter for the different SF_{inf} . The green solid line shows the position where $\gamma = 0.75$, and the magenta dashed line show the position where $\gamma = 0.5$, the expected value for a DRW process. We show with black stars the measurements done for the variable sources of the well-sampled sub-sample.	51
3.6	A measured for light curves simulated from a DRW process with $\tau = 300$ days and different values of SF_{inf} (with the long sampling). The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value of A (1:1 relation).	52
3.7	A measured for light curves simulated from a DRW process with $\tau = 300$ days and $\text{SF}_{\text{inf}} = 0.2$ mag The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value $A = 0.2$	53

3.8	γ measured for light curves simulated from a DRW process with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$ mag. The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value $\gamma = 0.5$	53
3.9	Distribution of the SF parameters A and γ for the variable and well sampled sources. Along the axes we show the histograms of every parameter. We plot in red the sources classified as BAL–QSO (sources with broad absorption–lines in their spectra), and in blue rest of the sample. The black solid line shows the expected value of γ for a DRW process. The green dashed lines show the median of the parameters. The green shaded regions show the 15.9 to 84.1 percentile range.	55
3.10	Correlations of the variable features for all the variable and well sampled sources. The diagonal shows the individual distributions. As a reference we provide the Spearman’s rank correlation coefficient for every pair of variables.	57
3.11	Correlations of the spectroscopic parameters for all the sources with M_{BH} and L_{5100} available. The diagonal shows the individual distributions. As a reference we provide the Spearman’s rank correlation coefficient for every pair of variables.	60
3.12	Comparison of M_{BH} vs. M_{BH}^C , and L/L_{Edd} vs. $(L/L_{\text{Edd}})^C$. The black dashed lines shows the 1:1 relations.	61
3.13	Bivariate correlations between the variability features and the spectral properties, for the “not – C IV sample”. The Spearman’s rank correlation coefficient is shown as reference for every pair of variables.	64
3.14	Left: Mean value of A in a grid of z_{spec} and L/L_{Edd} , for bins with 3 or more sources. Right: number of sources per bin of z_{spec} and L/L_{Edd} . The red dashed line shows the redshift from which we have available the Mg II line in the SDSS spectra (0.42).	67
3.15	Normalized histogram of the physical properties of variable (blue) and non–variable (red) sources from the QUEST–SDSS sample.	73
3.16	Normalized histogram of the light curve properties (number of epochs, t_{rest} , and mean magnitude) of variable (blue) and non–variable (red) sources from the QUEST–SDSS sample.	74
3.17	Distribution of the SF parameters A and γ , for sources classified as RL (red circles), RQ (blue stars), and sources without radio classification (None, black dots). Along the axes we show the histograms of every parameter, for the case of RL (blue) and RQ sources (red).	77
4.1	Example of four light curves from the QUEST–La Silla labeled set: two stars (blue dots, top panels) and two AGN (red dots, bottom panels).	89
4.2	Color–color diagrams of the labeled set. In the top panel we show $u - g$ versus $g - r$, in the middle panel $g - r$ vs $r - i$, and in the bottom panel $r - i$ vs $i - z$. The stars are represented by blue triangles, and the AGN are represented by circles whose colors depend on the redshift of each source. The contour plots show the distribution of AGN. In the top panel we show with yellow dashed lines the division used in Sesar et al. (2007) to identify regions of the $u - g$ versus $g - r$ diagram dominated by a particular type of source. In the middle panel, the black dashed line shows the position where $g - r = 0.6$	90

4.3	Confusion matrix from testing the RF1, RF2, and RF3 in the test set. True Label represent the classification done from SDSS spectra, and Predicted Label is the outcome of each classifier.	92
4.4	Distribution of the A_{SF} and Q31 features for the labeled set. Blue triangles correspond to non-AGN, and red circles correspond to AGN. We mark with black dots those AGN classified as variable, according to the definition used in Sánchez et al. (2017).	94
4.5	$g - r$ vs $r - i$ color-color diagrams of the unlabeled set (blue circles), and the hp-AGN sample (red stars) for the RF1 (top panel), RF2 (middle panel), and RF3 (bottom panel). The contour plots show the distribution of the hp-AGN samples for each classifier. In the top panel, we show the position of candidates observed during spectroscopic follow-up, differentiating between type 1 AGN (AGN1, yellow squares), BAL-QSO (BAL, red triangles), stars (cyan circles), and galaxies (Gal, black triangles). In the middle and bottom panels we show with letters and black squares the position of a selection of observed candidates located in the stellar locus.	97
4.6	Light curves of some RF1 candidates located in the stellar locus, observed during the spectroscopic follow-up campaign, shown in the top-right and bottom-left panels of Fig. 4.5. A and B are classified as type 1 AGN, C and D are classified as M-type stars.	98
4.7	Top panel: normalized distribution of A_{SF} for sources with detection in MILLIQUAS with well-sampled light curves in QUEST-La Silla. We show sources from the RF1 full-AGN sample (blue), the RF1 hp-AGN sample (red), and sources classified as non-AGN by RF1. Middle panel: normalized distribution of the mean Q magnitude for sources from the hp-AGN sample that are present (blue) and not present (red) in MILLIQUAS. Bottom panel: normalized distribution of A_{SF} for sources from the hp-AGN sample that are present (blue) and not present (red) in MILLIQUAS.	100
4.8	Top panel: normalized histogram of the r band magnitude of the RF1 hp-AGN sample (red) and the observed candidates (blue). Middle panel: normalized histogram of the predicted class probability (P_{RF}) of the RF1 hp-AGN sample (red) and the observed candidates (blue). Bottom panel: normalized histogram of spectroscopic redshift of AGN from the labeled sample (green) and observed candidates classified as type 1 AGN or BAL-QSO (red).	103
4.9	Rest-frame optical spectra of two type 1 AGN with evidence of continuum dominated by the host galaxy (LLAGN), observed with EFOSC2/NTT. In red we show the original spectra, and in blue we show the AGN component. The most prominent emission lines correspond to $H\alpha$	104
4.10	Rest-frame optical spectra of the observed candidates. The flux is in arbitrary units.	111
5.1	Observed light curve for a type 1 AGN located at RA= 150.194687° and DEC= 2.067890° from Marchesi et al. (2016a) catalog, located at $z=0.554$, variable in all the filter bands.	115

5.2	Examples of SEDs of the total component (light curves averages), and the four difference SED estimations (max-min, $P_{95} - P_5$, $P_{90} - P_{10}$, and σ_f^2), for (a): a source with clear variable torus component, (b): a source with unclear variable torus component, and (c): a source without variable torus component. We mark with a black dashed line the position where $\lambda_{\text{rest-frame}} = 1 \mu\text{m}$	117
5.3	Redshift (a) and rest-frame Hard band X-ray luminosities (b) distribution of sources with SED dominated by the torus (red) and SED dominated by the accretion disk (blue).	118
5.4	(a): Original Y (blue) and Ks (red) light curves, and Ks light curve after the accretion disk component subtraction (green) of a source located at RA= 149.351562° and DEC= 2.588570°, and at $z_{\text{spec}} = 0.963$. The light curves are presented with their flux normalized and shifted by a given value, for visualization purposes (b): Final time determination procedure of the same source in (a). We show the original CCCD (blue distribution), the weighted CCCD (yellow distribution), the smoothed distribution (black dashed lines), and $P(\Delta t)$ (green dashed lines). The gray dashed region shows the lag range used to define the final time lag (red solid line) and its uncertainties (red dashed lines).	121
5.5	Comparison of the time lags measured for the org-Ks ICCF (x-axis) and the ad-sub-Ks ICCF (y-axis), that satisfied the three conditions of Section 5.6.3. The blue dashed line shows the 1:1 relation.	122
5.6	Final radius (R_{DRM} ; in parsecs) versus the bolometric luminosity (L_{Bol} ; in erg s^{-1}), obtained from the the Hard band X-ray luminosity, from Marchesi et al. (2016a). The green solid line show the relation obtained by Koshida et al. (2014). The blue dashed line show the result obtained by Bentz et al. (2013) for the BLR ($\text{H}\beta$). The red dot-dashed line shows the sublimation radius of a pure-graphite dust torus with $T_{\text{sub}} \sim 1800 \text{ K}$, and a grain size of $a \sim 0.05 \mu\text{m}$. And the grey solid line shows the sublimation radius of a pure-graphite dust torus with $T_{\text{sub}} \sim 1900 \text{ K}$, and a grain size of $a \sim 0.55 \mu\text{m}$	126
5.7	Final lag (τ ; in days) vs different variable component luminosity in the V band (ΔL_V ; in $\text{erg s}^{-1} \text{ Hz}^{-1}$) measured using different techniques: a) $P_{95} - P_5$, b) $P_{90} - P_{10}$, c) Min-Max, d) σ_f . The blue dashed lines show the weighted least-squares linear regression (WLS).	127

Chapter 1

Introduction

1.1 Active Galactic Nuclei

The nuclei of some galaxies produce large amounts of electromagnetic radiation that cannot be explained by stellar activity. These are known as Active Galactic Nuclei or AGN. Soon after the discovery of the first AGN it was hypothesised that this radiation must be powered by the release of gravitational energy during the accretion of material onto a Supermassive Black Hole (SMBH), with a typical mass $M \sim 10^6\text{--}10^8 M_\odot$, located at the centre of its host galaxy (Lynden-Bell, 1969).

Current understanding of AGN physics requires the presence of an accretion disk that funnels matter onto the black hole (BH). As a result, the most basic traits of a given AGN should be well characterised by the BH mass (M_{BH}); the mass accretion rate (\dot{M}), that is often described by the ratio of the AGN bolometric luminosity (L) to the Eddington luminosity (L_{Edd}), also known as the Eddington ratio (L/L_{Edd}); and its spin or angular momentum (J), that is usually represented by the dimensionless spin parameter $a_* = J/(GM^2/c)$, with values between -1 and 1, and which determines the innermost limit of the accretion disc. For those cases where accretion takes the form of an optically thick flow (e.g., Shakura & Sunyaev 1973), these three properties, M_{BH} , \dot{M} and a_* jointly determine the AGN bolometric luminosity.

AGN are commonly classified in the optical range by the presence or absence of broad permitted emission lines ($\text{FWHM} \gtrsim 2000 \text{ km s}^{-1}$), into Broad Line AGN (or type I) and Narrow Line AGN (or type II), respectively. The unified model is one of the most successful approaches to explain this dichotomy. It postulates that an obscuring dusty torus around the central engine is responsible for the different classes of AGN, which occur when we observe the source at different angles (Antonucci & Miller, 1985). The most promising models include a clumpy torus and disk winds (see Netzer 2015; Ramos Almeida & Ricci 2017 and references therein), as they would explain the spectral energy distribution (SED) observed from the infrared to the X-ray bands, and the existence of at least some ‘‘Changing look’’ AGN, which correspond to AGN that suffer changes of their broad emission lines, or changes of their level of obscuration (Tohline & Osterbrock, 1976; Shappee et al., 2014; Denney et al., 2014;

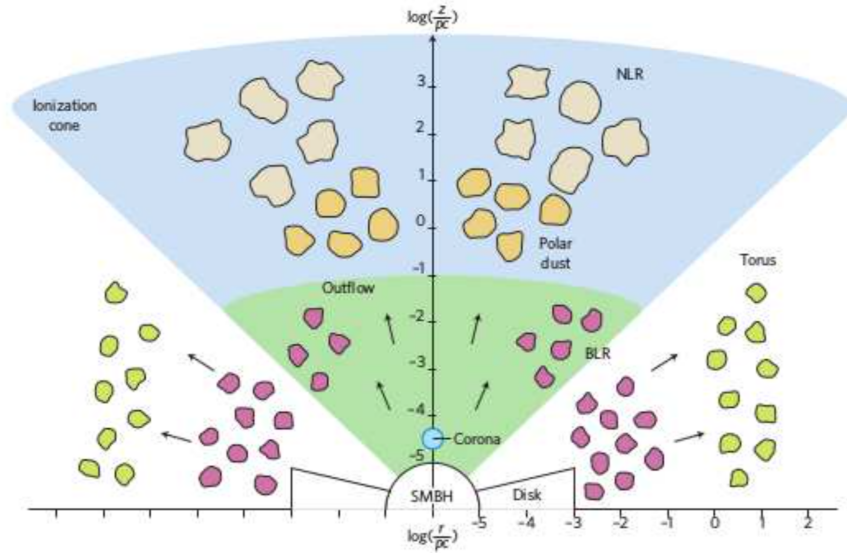


Figure 1.1: Sketch of the main AGN structures seen along the equatorial and polar direction. From the center to host–galaxy scales: SMBH, accretion disk and corona, BLR, torus and NLR. Different colours indicate different compositions or densities. Credits: Ramos Almeida & Ricci (2017).

LaMassa et al., 2015; Ricci et al., 2016). Figure 1.1 shows a graphical representation of the different AGN structures, taken from Ramos Almeida & Ricci (2017), where from the very center to host–galaxy scales the main AGN structures are the corona (which emits mostly in the X–rays), the accretion disk (which emits mostly in the UV/optical/NIR range), the broad line region (BLR, which emits mostly in the UV/optical/NIR range), the dusty torus (which emits mostly in the infrared regime) and the narrow line region (NLR, which emits mostly in the UV/optical/NIR range).

Some objects might require more drastic modifications to the unified model in order to explain their properties, like those that seem to lack a BLR, called by some authors “True type II” AGN (Panessa & Bassani, 2002; Elitzur & Netzer, 2016) or Weak Emission Line Quasars (WLQ) (Diamond-Stanic et al., 2009; Shemmer et al., 2009; Wu et al., 2012; Shemmer & Lieber, 2015; Luo et al., 2015). Moreover, the relation between the optical and X–ray obscuration is still a debatable matter (Merloni et al., 2014; Burtscher et al., 2016; Marchesi et al., 2016b).

1.2 AGN variability

AGN show time–variable emission in every waveband in which they have been studied. The variations appear to be aperiodic and have variable amplitude (see Figure 1.2). The characteristic time–scales of the variability range from hours to years, and for a single object, the shortest time–scales are associated with shorter emission wavelengths. This can be understood in the context of the current AGN structure models, where ultraviolet (UV) and

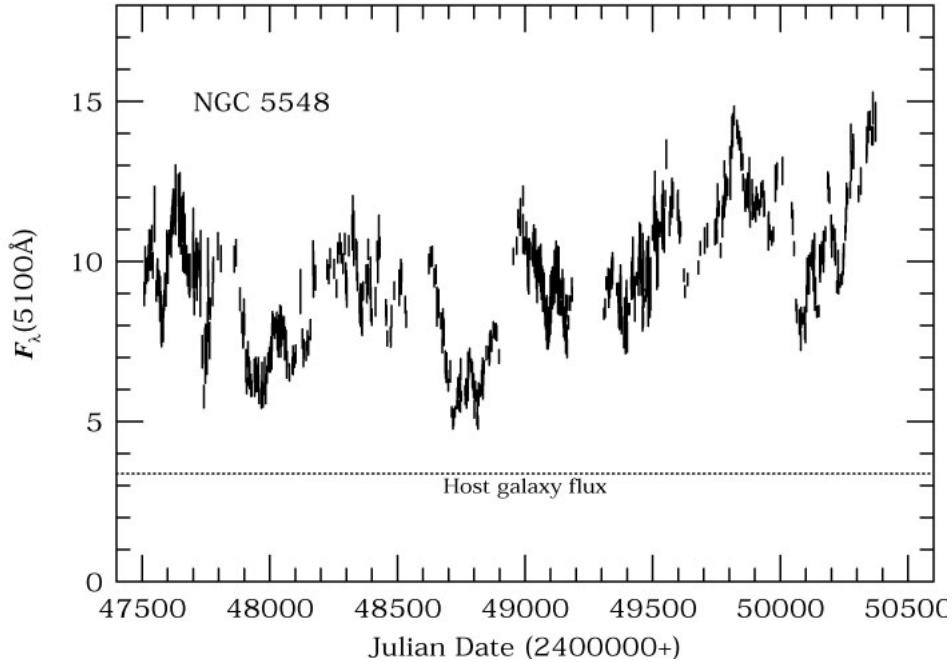


Figure 1.2: Optical (5100 Å) light curve of NGC 5548, taken from Peterson (2001).

optical emission are originated in an accretion disk around a SMBH, while non-thermal X-ray emission is produced in a inner hot plasma component (corona), which is geometrically much smaller and compact than the accretion disk (see Figure 1.3), and therefore able to show more rapid variability. Intensive monitoring of nearby AGN suggests that short term variability from the UV to the near infrared (NIR) could be driven by the rapid changes in the X-ray or far UV flux, which illuminates the accretion disk producing the short term UV/optical/NIR variations. This explains the lags seen between continuum bands, found in reverberation mapping (RM) analyses (e.g Lira et al. 2011, 2015; Edelson et al. 2015). However, it has been noticed that at time-scales of months or years, the amplitude of the UV/optical variability can be larger than the amplitude of the X-ray variability, which implies that X-ray reprocessing is not the main source of the UV/optical variations, and intrinsic variability from the accretion disk is required (Krolik et al., 1991; Arévalo et al., 2008; Lira et al., 2015; Edelson et al., 2015).

There is significant evidence of a strong correlation between the X-ray, UV, and optical variability (Uttley et al., 2003; Arévalo et al., 2008, 2009; Breedt et al., 2009, 2010; McHardy et al., 2016; Troyer et al., 2016; Buisson et al., 2017). Also, correlations between the NIR and optical bands have recently been established (Suganuma et al., 2006; Koshida et al., 2009; Lira et al., 2011; Koshida et al., 2014; Lira et al., 2015), as can be seen in the light curves shown in Figure 1.3. These results show that the rest-frame optical emission comes primarily from the accretion disk, and the infrared emission comes mostly from the dusty torus (Lira et al., 2011, 2015), while the spectral energy distribution (SED) of an AGN around the rest-frame wavelength $\sim 1\mu m$ samples simultaneously two emission components, the accretion disk and the hottest part of the dusty torus. This has been confirmed both photometrically and spectroscopically (Glass, 1992; Landt et al., 2011). Besides, Lira et al. (2015) found evidence of the presence of NIR emission from the accretion disk in the J and

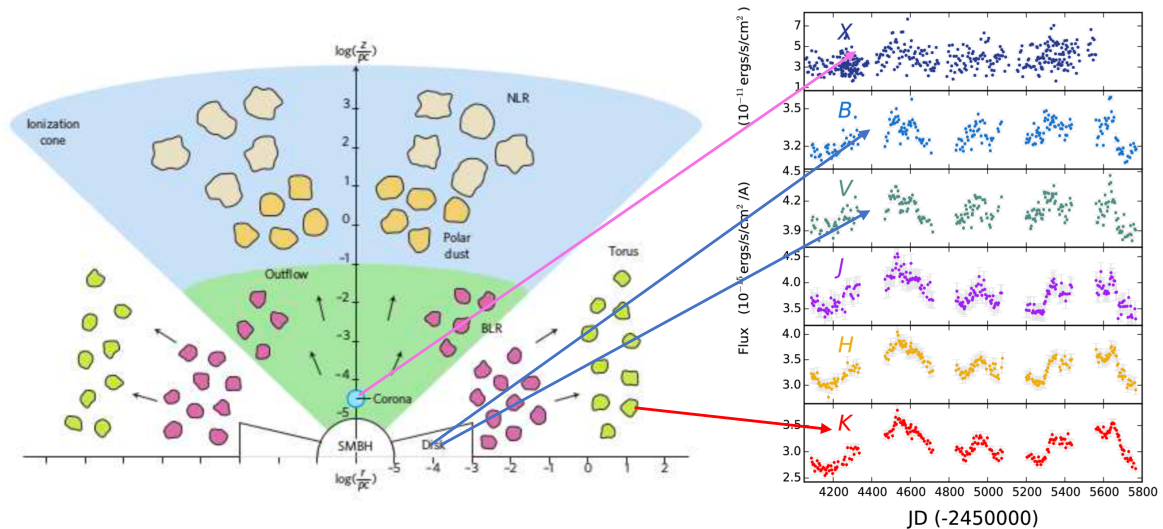


Figure 1.3: Left: sketch of the main AGN structures seen along the equatorial and polar direction (as in Figure 1.1, credits: Ramos Almeida & Ricci 2017). Right: observed X-ray, optical and NIR light curves for MCG-6-30-15, from Lira et al. (2015). The arrows show the AGN regions associated to every light curve.

H bands observations of MCG-6-30-15 by cross correlating optical and NIR light curves.

In order to study variability properties of statistically significant AGN samples it is necessary to probe a wide range of time-scales. Hence, long and intensive campaigns of a large number of targets are crucial. Fortunately, in recent years surveys covering a significant part of the sky, revisiting the same regions on time-scales from days to years, and containing a large sample of serendipitous objects –blind surveys– are now becoming available as predecessors of the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008). LSST will revolutionize time-domain astronomy, providing for the first time the opportunity to study variable objects for a long-period of time (~ 10 years), at very faint magnitudes ($r \sim 24.5$ for single images), and with a large total covered area ($>18,000$ deg²). Simulations performed by the LSST AGN Science Collaboration predict detection of over 10^7 AGN to beyond $r \sim 24$ (LSST Science Collaboration et al., 2009). This is a huge increase in the number of sources available for variability analysis since current studies typically only probe to limiting magnitudes of $r \sim 21$, with total number of sources between 10 and 10^5 (e.g., Cristiani et al. 1997; Vanden Berk et al. 2004; MacLeod et al. 2010; Peters et al. 2015; Simm et al. 2016; Caplar et al. 2017; Li et al. 2018).

1.2.1 Origins of the variations

Even though variability is one of the defining characteristics of AGN, we do not completely understand the mechanisms that drive such variations. Some physical models used to explain them include: accretion disk instabilities (like thermal instabilities in the accretion disk, magneto-rotational instabilities, or changes in the torque conditions at the innermost circular stable orbit; ISCO), and obscuration by orbiting or infalling dusty clouds (see Stern et al.

2018; Ross et al. 2018 and references therein).

There are five characteristic time-scales important to understand the physics behind AGN variability; the light crossing time (t_{lc}), which corresponds to the time at which any light variation propagates; the gas orbital time-scale (t_{orb}), which corresponds to the time scale of the orbital period of the gas in the accretion disk; the accretion disk thermal time-scale (t_{th}), which corresponds to the time-scale on which thermal instabilities are produced (or the time-scale on which the disk heats or cools); the accretion disk front time-scale (t_{front}), which corresponds to the time-scale on which the thermal instabilities are propagated (or the time-scale on which the cooling and heating fronts propagate); and the viscous time-scale (t_v), which corresponds to the time-scale over which the effects of variations in the accretion rate will propagate through the disk (e.g., Edelson & Nandra 1999; Peterson 2001; Kelly et al. 2009; Stern et al. 2018; Ross et al. 2018). These time-scales are defined as:

$$t_{lc} = 1.1 \times \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{100 R_S} \right) \text{ days}, \quad (1.1)$$

$$t_{orb} = 104 \times \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{100 R_S} \right)^{3/2} \text{ days}, \quad (1.2)$$

$$t_{th} = 4.6 \times \left(\frac{\alpha}{0.01} \right)^{-1} \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{100 R_S} \right)^{3/2} \text{ years}, \quad (1.3)$$

$$t_{front} = 20 \times \left(\frac{h/R}{0.05} \right)^{-1} \left(\frac{\alpha}{0.03} \right)^{-1} \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{150 R_S} \right)^{3/2} \text{ years}, \quad (1.4)$$

$$t_v = 400 \times \left(\frac{h/R}{0.05} \right)^{-2} \left(\frac{\alpha}{0.03} \right)^{-1} \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{R}{150 R_S} \right)^{3/2} \text{ years}, \quad (1.5)$$

where M_{BH} is the BH mass, R is the emission distance from the central BH, R_s is the Schwarzschild radius ($2GM_{\text{BH}}/c^2$), α is the standard disk viscosity parameter (Shakura & Sunyaev, 1973), and h is the accretion disk height.

As can be seen from the Equations 1.1, 1.2, 1.3, and 1.5, for $M_{\text{BH}} \sim 10^8 M_{\odot}$, t_{lc} must be related with very rapid variability (of the order of hours or few days), as those typically observed during RM campaigns, t_{orb} and t_{th} must be related with variability with time-scales of months or few years, and t_{front} and t_v must be related with long time-scale AGN variability (tens or hundreds of years). Moreover, from these equations, we must expect a strong correlation between the time-scale of the AGN variability and the M_{BH} , which has been found by previous analyses (e.g., McHardy et al. 2006; Kelly et al. 2009; MacLeod et al. 2010). This tell us that the source of rapid optical variability must be the reprocessing of highly variable high-energy component by the accretion disk, and the source or variability with larger time-scales must be intrinsic to the accretion disk.

1.2.2 Time series analysis

AGN variability seems to be well described as a stochastic process (Kelly et al., 2009, 2014; Graham et al., 2014; Simm et al., 2016). Power spectral density (PSD) analysis is a useful tool to analyze the physical processes involved in the stochastic variability of AGN. The PSD measures the variability power per temporal frequency ν . However, the irregular sampling of ground-based light curves complicates the PSD analysis. Previous analysis show that AGN power spectra are well described by a broken power law with a PSD $\propto 1/\nu^2$ after the break (Collier & Peterson, 2001; Czerny et al., 2003; Kelly et al., 2009), which is consistent with Damped Random Walk (DRW) or autoregressive processes. Kelly et al. (2014) showed that AGN light curves can also be well described by continuous-time autoregressive moving average (CARMA) models, which fully account for irregular sampling and measurement errors. From these models, we can describe the whole range of possible AGN PSD shapes, by using different parameters in the CARMA models.

In order to have a complete description of the AGN variability, we can use different variability features. Here we described the four statistics mostly used during this thesis: P_{var} and the excess variance (σ_{rms}), which are two parameters useful to describe the amplitude of the variability; and the structure function (SF) and the DRW modelling, which are useful to describe the shape of the variability between two observations separated by a given time (i.e. the structure of the variability).

The P_{var} parameter

P_{var} (e.g., Paolillo et al. 2004; Young et al. 2012; Lanzuisi et al. 2014; Cartier et al. 2015; Sánchez et al. 2017; Sánchez-Sáez et al. 2018). corresponds to the probability that the source is intrinsically variable; it considers the χ^2 of the light curve, and calculates the probability $P_{var} = P(\chi^2)$ that a χ^2 lower or equal to the observed value could occur by chance for an intrinsically non-variable source:

$$\chi^2 = \sum_{i=1}^{N_{obs}} \frac{(x_i - \bar{x})^2}{\sigma_{err,i}^2} \quad (1.6)$$

where x_i is the magnitude at each epoch, $\sigma_{err,i}$ is its error, \bar{x} is the mean magnitude, and N_{obs} is the number of epochs in which the object was detected. If the source were intrinsically non-variable, the value of χ^2 would be $\sim (N_{obs} - 1)$, the number of degrees of freedom in the data. If the value of P_{var} is large, $1 - P_{var}$, the probability that the variability observed is due to Poisson noise alone and the source is intrinsically non-variable is low.

The Excess Variance σ_{rms}^2

The normalized excess variance σ_{rms}^2 (Nandra et al., 1997; Turner et al., 1999; Allevato et al., 2013; Lanzuisi et al., 2014; Cartier et al., 2015; Sánchez et al., 2017; Sánchez-Sáez et al.,

2018), is a measure of the variability amplitude. Here we adopt the definition:

$$\sigma_{rms}^2 = \frac{1}{N_{obs}\bar{x}^2} \sum_{i=1}^{N_{obs}} [(x_i - \bar{x})^2 - \sigma_{err,i}^2] \quad (1.7)$$

And its error due to Poisson noise is:

$$err(\sigma_{rms}^2) = \frac{S_D}{\bar{x}^2 N_{obs}^{1/2}}, \quad (1.8)$$

$$S_D^2 = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \{[(x_i - \bar{x})^2 - \sigma_{err,i}^2] - \sigma_{rms}^2 \bar{x}^2\}^2 \quad (1.9)$$

For a non-variable object, the excess variance can be negative: if variability is not detected, due to large errors, the value of $err(\sigma_{rms}^2)$ can be greater than σ_{rms}^2 . The excess variance has to be treated with extreme care (Allevato et al., 2013), since it can be biased by the structure of the variability of the source, the sampling and the length of the light curve.

The Structure Function

The Structure Function (SF) describes the variability of a source by quantifying the amplitude of the variability as a function of the time lapse between compared observations (τ) (Cristiani et al., 1996; Giveon et al., 1999; Vanden Berk et al., 2004; de Vries et al., 2005; Rengstorf et al., 2006; Schmidt et al., 2010; Palanque-Delabrouille et al., 2011; Graham et al., 2014; Cartier et al., 2015; Kozłowski, 2016; Caplar et al., 2017; Sánchez et al., 2017; Sánchez-Sáez et al., 2018). Kozłowski (2016) provides a good summary of the different definitions of the Structure Function used in the literature. In particular he proposed to define it as: $SF_{obs}(\Delta t) = 0.741 \times IQR$, where IQR is the interquartile range between 25% and 75% of the sorted $y(t) - y(t + \Delta t)$ distribution. He also provides a method to measure the Structure Function taking into account the photometric noise: $SF_{true}^2(\Delta t) = 0.549(IQR^2(\Delta t) - IQR^2(n))$, where $IQR(n)$ is the interquartile range between 25% and 75% of the sorted Δm for $\Delta t < 2$ days.

The Structure Function of AGN is generally well described by a broken power law. The time scale where the break happens is known as the decorrelation time scale. Emmanoulopoulos et al. (2010) discussed the different limitations of the use of the decorrelation time scale as a physically meaning parameter, which can be severely affected by the light curve sampling. Therefore, we limit our analysis to the regime where we can describe the Structure Function as a single power law (i.e. before the break):

$$SF(\tau) = A \left(\frac{\tau}{1\text{yr}} \right)^\gamma \quad (1.10)$$

where A corresponds to the mean magnitude difference on a one year time-scale and γ is the logarithmic gradient of this change in magnitude.

The γ parameter is directly related to the PSD slope. When $SF \propto t^\gamma$ then $PSD \propto 1/f^{1+2\gamma}$, therefore, the value of γ changes depending of the type of process involved in the variation. For example, for a white noise process $\gamma = 0$, and for a random walk process $\gamma = 0.5$.

Schmidt et al. (2010) provides two methods to define the Structure Function, the first one is similar to other definitions used in the literature:

$$SF(\tau) = \frac{1}{N_{bin}} \sum_{i,j} \left(\sqrt{\frac{\pi}{2}} |\Delta m_{ij}| - \sqrt{\sigma_i^2 + \sigma_j^2} \right) \quad (1.11)$$

Where the average $|\Delta m_{ij}|$ is taken over all the epoch pairs i, j whose lag in time $\Delta t_{ij} = t_i - t_j$ falls inside the bin $[\tau - \frac{\Delta\tau}{2}, \tau + \frac{\Delta\tau}{2}]$. N_{bin} corresponds to the number of pairs inside the bin, Δm_{ij} is the difference in magnitude between the two epochs ($m_i - m_j$), and σ_i and σ_j are the magnitude errors for each epoch, respectively.

The second method proposed by Schmidt et al. (2010) is a Bayesian approach, where they model the Structure Function with a power-law using a Markov Chain Monte Carlo (MCMC) method. In this method, a list with all the possible epoch pairs is constructed. Then the Structure Function is modeled considering a likelihood $\mathcal{L}(A, \gamma)$, and priors $p(A)$ and $p(\gamma)$ defined as:

$$\mathcal{L}(A, \gamma) = \prod_{ij} \frac{1}{\sqrt{2\pi V_{\text{eff},ij}^2}} \exp \left(-\frac{\Delta m_{ij}^2}{2V_{\text{eff},ij}^2} \right) \quad (1.12)$$

$$V_{\text{eff},ij}^2 = \left[A \left(\frac{\Delta t_{ij}}{1\text{yr}} \right)^\gamma \right]^2 + (\sigma_i^2 + \sigma_j^2) \quad (1.13)$$

$$p(A) \propto \frac{1}{A}, \quad p(\gamma) \propto \frac{1}{1 + \gamma^2} \quad (1.14)$$

The main advantage in determining the Structure Function with a Bayesian approach compared with traditional definitions, is that it can avoid problems given by the sampling of the light curve and the selection of the bin size and shape, since these parameters are inferred directly from the data. This method is also less susceptible to windowing effects, given by the finite length of the light curve. Moreover, from the posterior distribution of the parameters, we can determine the mean value and the 1σ errors of the measurements. In our particular case, the values of γ and A were constrained to be in the ranges $\gamma \in [0, 10]$ and $A \in [0, 1]$ mag/year. Besides, we only considered epoch pairs with a maximum separation of 1 year in the rest-frame, in order to avoid the regime of the Structure Function after the break.

Damp Random Walk

Kelly et al. (2009) proposed that a continuous-time first-order autoregressive process or CAR(1) can be a good descriptor of this kind of variation. This model is also called “Damped Random Walk”, since it is represented by a stochastic differential equation which includes a damping term that pushes the signal back to its mean:

$$dX(t) = -\frac{1}{\tau}X(t)dt + \sigma_{DRW}\sqrt{dt}\varepsilon(t) + bdt, \quad \tau, \sigma_{DRW}, t > 0 \quad (1.15)$$

here, $X(t)$ is the AGN light curve, represented by the observed magnitude, τ is the “relaxation time” of the process or the characteristic time for the time series to become roughly uncorrelated (related with the decorrelation time mentioned in the previous section), $\varepsilon(t)$ is a white noise process with zero mean and variance equal to 1, and σ_{DRW} is the amplitude of the variability on short time-scales compared with τ . The mean value of the process is $b\tau$ and its variance is $\tau\sigma_{DRW}^2/2$. The long time-scale variability (SF_{inf}) is calculated as $\sigma_{DRW}\sqrt{\tau/2}$.

From the fit of a DRW process, we can derive the power spectral density of the light curve, avoiding the windowing effect that appears when the PSD is derived directly from the data:

$$\text{PSD}(f) = \frac{2\sigma_{DRW}^2\tau^2}{1 + (2\pi\tau f)^2} \quad (1.16)$$

where f is the frequency measured in days^{-1} .

There are two main regimes for the $\text{PSD}(f)$ of a DRW process. For short time-scales (compared to the relaxation time, i.e., $f \lesssim (2\pi\tau)^{-1}$), the power spectrum falls off as $1/f^2$. On longer time-scales than the relaxation time, the power spectrum flattens to white noise. Therefore, τ can be considered as the characteristic time-scale of the variability. (Kelly et al., 2009).

Kelly et al. (2014) provides a public PYTHON package to fit CARMA models called *carma_pack*¹. The package includes the option to model DRW processes.

1.3 Data sets

This thesis is based in the following photometric data sets:

1.3.1 The QUEST-La Silla AGN variability survey

Between 2010 and 2015 we carried out “The QUEST–La Silla AGN variability survey” (hereafter QUEST–La Silla), using the wide-field QUEST camera mounted on the 1m ESO-Schmidt telescope at La Silla Observatory (Cartier et al., 2015, 2016). The survey used a

¹https://github.com/brandonkelly/carma_pack

broadband filter, the Q band, similar to the union of the g and r SDSS filters. Our survey includes the COSMOS, ECDF-S, ELAIS-S1, XMM-LSS and Stripe 82 fields. These are some of the most intensively observed regions in the southern sky. Our QUEST fields are much larger than the nominal fields, but we will still adopt the same names, with a surveyed area of $\sim 14 \text{ deg}^2$ per field, with the exception of the XMM-LSS field, which covers an area of $\sim 38 \text{ deg}^2$. One of the advantages of our survey over other surveys is the intense monitoring used, observing the fields every possible night. Individual images reached a limiting magnitude between $r \sim 20.5$ and $r \sim 21.5$ mag for an exposure time of 60 seconds or 180 seconds, respectively.

The aims of our survey are: 1) to test and improve variability selection methods of AGN, and find AGN populations missed by other optical selection techniques (Schmidt et al., 2010; Butler & Bloom, 2011; Palanque-Delabrouille et al., 2011); 2) to obtain a large number of well-sampled light curves, covering time-scales ranging from days to years; 3) to study the link between the variability properties (e.g., characteristic time-scales and amplitudes of variation) with physical parameters of the system (e.g., black-hole mass, luminosity, and Eddington ratio).

Cartier et al. (2015) presented the technical description of the survey, the full characterisation of the QUEST camera, and a study of the relation of variability with multi-wavelength properties of X-ray selected AGN in the COSMOS field.

We reduced the data from the QUEST-La Silla survey using our own customized pipeline, following the same procedure described by Cartier et al. (2015), which includes dark subtraction, flat-fielding, and astrometric and photometric calibration. To calibrate the photometry, we used public photometric SDSS catalogs (Gunn et al., 1998; Doi et al., 2010) for the COSMOS, Stripe 82 and XMM-LSS fields, and public catalogs from the first year of the Dark Energy Survey (DES; Abbott et al. 2018) for the ELAIS-S1 and ECDF-S fields.

1.3.2 UltraVISTA survey

UltraVISTA (McCracken et al., 2012) is an ultra-deep, near infrared survey on the COSMOS field, using the VIRCAM instrument on the VISTA telescope of the European Southern Observatory (ESO) (Emerson et al., 2006; Dalton et al., 2006; Sutherland et al., 2015). The aim of this survey is to study the evolution of galaxies out to $z \sim 4$, and achieve a comprehensive view of AGN and the most massive galaxies up to the epoch of re-ionization. UltraVISTA is repeatedly imaging the field in 5 bands (YJHKs and NB118), covering an area of 1.5 deg^2 . The most recent UltraVISTA data release (DR3) corresponds to the first 5 years of observations. The advantages of these data compared to previous surveys is that UltraVISTA provides good quality and good resolution images (with a mean seeing of $\sim 0.8''$) at several epochs, and light curves with a length of almost 5 years and with good sampling. Besides, the depth of the images allowed us to cover a wide redshift range, therefore allowing the access to optical and near infrared rest-frame emission.

1.4 This thesis

Despite significant efforts to understand the physics behind AGN variability, we do not completely understand the mechanisms that drive such variations. In particular, we are missing significant details of how AGN variability at different wavelengths is related (Arévalo et al., 2008; Lira et al., 2011, 2015); how physical properties of the central engine (e.g., luminosity, black hole mass, hardness ratio, optical colors, etc) are related to well defined variability properties of the system (e.g., characteristic time-scale, variability amplitude, etc); and how distinct is the variability behaviour of different AGN populations. Moreover, given the ubiquity of variability and the large number of variable sources to be found with LSST, it is critical to characterize AGN variability and define reliable selection criteria before LSST enters in operations.

In this thesis I aim to better understand the nature of AGN variability. The specific goals of this thesis are:

- Study the changes in variability across different classes of AGN.
- Explore the connection between AGN optical variability and black hole physical properties.
- Improve AGN variability selection methods, finding populations missed by other selection techniques.
- Study whether we can use the time lag between optical and NIR continuum emission of AGN as standard candle for cosmology.

This thesis is organized as follows. In Chapter 2, I present our statistical study of NIR variability of X-ray-selected AGN in the COSMOS field, using UltraVISTA data. This is the largest sample of AGN light curves in YJHKs bands, making it possible to have a global description of the nature of AGN for a large range of redshifts and for different levels of obscuration. The results of this work are presented in Sánchez et al. (2017). In Chapter 3, I present our statistical analysis of the connection between AGN variability and physical properties of the central SMBH, using data from the QUEST–La Silla survey. The results of this work are presented in Sánchez-Sáez et al. (2018). In Chapter 4, I present our variability-based active galactic nuclei (AGN) selection technique. We used data from the QUEST—La Silla survey to construct light curves for 208583 sources over ~ 70 deg². We used different variability features to characterize our light curves, and we implemented a Random Forest (RF) algorithm to classify our objects as either AGN or non-AGN according to their variability features and optical colors. The results of this work are presented in Sánchez-Sáez et al. 2019, submitted to *Astrophysical Journal Supplement* (ApJS). In chapter 5 I present our Dust Reverberation Mapping analysis, done to determine the characteristic optical–NIR time lags for sources with $0.5 < z < 1.2$, to study whether the emission received in the NIR of high redshift sources is consistent with emission from the dusty torus or the accretion disk, and to determine whether we can use the time lag between optical and NIR continuum emission of AGN as standard candle for cosmology. The results of this work will be included in a forthcoming paper. Finally, in chapter 6 I present the summary and most relevant conclusions of this thesis.

Chapter 2

Near Infrared Variability of obscured and unobscured X-ray selected AGN in the COSMOS field

The material presented in this chapter is based on the work published in Sánchez et al. (2017)

2.1 Introduction

To understand the physics behind AGN we need multi-wavelength variability studies, from which we can determine the contribution of the distinct components of the emission. But multi-wavelength projects are expensive and difficult to accomplish. Several authors have studied the optical, UV, and X-ray variability of AGN for a significant number of sources (e.g., Cristiani et al. 1996; Nandra et al. 1997; Turner et al. 1999; Giveon et al. 1999; Vanden Berk et al. 2004; de Vries et al. 2005; Rengstorf et al. 2006; Schmidt et al. 2010; MacLeod et al. 2011; Palanque-Delabrouille et al. 2011; Lanzuisi et al. 2014; Graham et al. 2014; Cartier et al. 2015; Simm et al. 2016; Caplar et al. 2017). However, very little is known about the variability in the near infrared range. The study of AGN variability in the NIR is particularly difficult, since the contamination from the host galaxy in this wavelength region can be large (Hernán-Caballero et al., 2016). Therefore, for most sources, the NIR variability of the central source is overshadowed by the emission from the galaxy. Neugebauer et al. (1989) studied the NIR variability for individual quasars using a sample of 108 optically selected sources. They showed that only half of their sources have a high probability of been variable. Enya et al. (2002a,b,c) analyzed the variability of 226 AGN, in the J, H and K' bands. Their work suggests that most AGN are variable in the near-infrared. However, both studies have the following limitations: small number of epochs, limited redshift coverage and only some specific classes of AGN were considered.

In this work, we constructed light curves using YJHKs data from the UltraVISTA survey (see Section 1.3.2) of known AGN, selected from public catalogs of the COSMOS field (Lusso

et al., 2012; Muzzin et al., 2013; Marchesi et al., 2016a; Laigle et al., 2016), classifying them according to their X-ray, optical, and radio properties. We used these light curves to understand the differences in the variability behavior of the different classes of AGN, by implementing statistical tools widely used by the AGN community, like the Structure Function, damp random walk modelling, and the excess variance.

The chapter is structured as follows: in section 2.2, we briefly introduce the data set and the public catalogs used. In section 2.3, we give a definition of the different AGN classes considered in this work. In section 2.4 we describe the steps needed to get the light curves. In section 2.5 we explain all the statistical tools used for the variability analysis. In section 2.6 we present the results of the different analyses, considering the whole data set and the different classifications. In section 2.7 we discuss the physical implications of our work and we summarize the main results. The photometry reported is in the AB system. We adopt the cosmological parameters $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

2.2 Data

2.2.1 NIR data

Our work is based on the NIR imaging data from the UltraVISTA survey (McCracken et al., 2012), which has repeatedly imaged the COSMOS field during five years in the *YJHKs* bands (see Section 1.3.2). The data considered in this work were taken between December 2009 and June 2014.

The final product of the UltraVISTA survey are stacked images and their corresponding weight maps for each filter band. However for our purpose, we worked with the individual OBs stacks, which have a total exposure of 0.5 or 1 hour, produced by the Cambridge Astronomy Survey Unit (CASU)¹, and provided by the survey team, from which we have constructed our light curves. The reduction process done by CASU includes dark subtraction, flat-fielding, sky-subtraction, astrometric and photometric calibration, therefore, we did not implement any reduction steps², with the exception of photometric calibration (see section 2.4.1). The pixel scale of the images is 0.34"/px, and the average 5σ magnitude limits for the single images in each bands are 23.3, 23.1, 22.2, and 22.1 for the Y, J, H and Ks bands, respectively.

2.2.2 Ancillary data

We take advantage of the huge amount of ancillary data available for the COSMOS field, ranging from X-rays to radio waves. In our analysis we use four public catalogs.

¹<http://casu.ast.ac.uk/surveys-projects/vista/technical/data-processing>

²see Emerson et al. (2004); Irwin et al. (2004); Lewis et al. (2010); Milvang-Jensen et al. (2013) for more details of the CASU processing steps

The first one is a public Ks–selected catalog of the COSMOS/UltraVISTA field using the UltraVISTA data (Muzzin et al., 2013). This catalog provides photometry for 30 bands, covering the range 0.15–24 μm , besides other parameters related to the quality of the UltraVISTA photometry. For our work, the *contamination* parameter resulted particularly important, as indicates whether an object’s photometry has been contaminated by a nearby bright star. When the value of this parameter is zero there is no contamination.

The second one is the catalog of optical and infrared counterparts of the Chandra COSMOS–Legacy Survey (Marchesi et al., 2016a). This catalog contains 4016 X–ray sources from the 4.6 Ms Chandra program on 2.2 deg^2 of the COSMOS field (Civano et al., 2016), with 3877 sources having an optical/IR counterpart. The catalog provides X–ray fluxes measured in three bands (Soft: 0.5–2 keV, Hard: 2–10 keV, and Full: 0.5–10 keV), hardness ratios, intrinsic neutral hydrogen (N_H) column densities, luminosity distances, identification and photometry of the counterparts in i, Ks and 3.6 μm , spectroscopic redshift and classification for 1770 sources, photometric redshift and classification for 3885 sources, among other measurements. We complement the information of this catalog, with the new catalog of the 1855 extragalactic sources in the Chandra COSMOS–Legacy survey catalog having more than 30 net counts in the 0.5–7 keV band (Marchesi et al., 2016b). This catalog provides new values of N_H and the photon index (Γ) computed through spectral fitting. We use the values of N_H reported by Marchesi et al. (2016b) when their are available, otherwise we use the values reported by Marchesi et al. (2016a).

Besides, we used the COSMOS2015 catalog (Laigle et al., 2016), which provides photometric redshifts and stellar masses for more than half a million objects over 2 deg^2 in the COSMOS field, including photometry for several bands, with a wavelength coverage from the ultraviolet to the radio regimes. In particular we used SuprimeCam B band photometry (Taniguchi et al., 2007, 2015), and VLA photometry at 1.4 GHz (Schinnerer et al., 2004, 2007, 2010; Bondi et al., 2008; Smolčić et al., 2014).

Finally, Lusso et al. (2012) provides measurements of Bolometric luminosities and Eddington ratios for X–ray selected broad–line (382 sources) and narrow–line (547 sources) AGN from the XMM–Newton survey in the COSMOS field (Brusa et al., 2010). The bolometric luminosities are computed from the integrated SED.

The X–ray selection currently is the least biased but most expensive method to identify AGN, thus we considered as base for our analysis the Marchesi et al. (2016a) catalog, as this allow us to work with sources securely classified as AGN. We cross–matched this catalog with that of Muzzin et al. (2013), in order to secure the quality of the UltraVISTA photometry for every source, saving only the sources having a *contamination* parameter equal to zero. Then, we cross–matched the catalog with the Laigle et al. (2016) catalog, in order to obtain the B band and radio photometry. Finally, the resultant catalog was cross–matched with the Lusso et al. (2012) catalog to obtain the bolometric luminosities for 718 of our sources. The final cross–matched catalog (hereafter the clean–AGN catalog) contains 3050 sources.

2.3 AGN Classification

Most of the variability studies available in the literature are based on objects classified as type I or unobscured AGN (Cristiani et al., 1996; Nandra et al., 1997; Turner et al., 1999; Enya et al., 2002a,b,c; Vanden Berk et al., 2004; de Vries et al., 2005; Schmidt et al., 2010; Lanzuisi et al., 2014; Graham et al., 2014; Simm et al., 2016). Since we have well-sampled NIR light curves, we can expand our analysis and try to look for possible differences in variability between objects classified as obscured or unobscured according to different criteria. Using the information available from three public catalogs described in section 2.2.2, we classified the sources of the clean-AGN catalog in the following way:

2.3.1 Spectroscopic Classification

Marchesi et al. (2016a) made use of the master spectroscopic catalog available for the COSMOS collaboration (M. Salvato et al., in preparation), to classify spectroscopically 1770 sources. For details on the source of the spectroscopic redshifts see Marchesi et al. (2016a). They classified the objects according to the following criteria:

1. *Broad line* (BL): sources with at least one broad ($\text{FWHM} \geq 2000 \text{ km s}^{-1}$) emission line in their spectra (632 sources).
2. *Not Broad line* (NL): sources that do not present broad lines (they show narrow emission lines or absorption lines). These sources have not been separated between star-forming galaxies or type II AGN because most of the sources have low S/N spectra, or are in an observed wavelength range which does not allow to use emission line diagnostic diagrams to separate Type II AGN and star-forming galaxies (1049 sources).
3. *Stars*: sources spectroscopically identified as stars (89 sources).

From these sub-samples, we have 563 BL and 952 NL in the clean-AGN catalog.

2.3.2 Photometric Classification

Marchesi et al. (2016a) also provide photometric redshifts and classification for 3885 objects. They used the method described by Salvato et al. (2011), which adjusts templates to the sources multiwavelength SEDs. The templates are divided in (Salvato et al., 2009): ‘unobscured AGN’, which corresponds to a type I AGN or type I QSO template (894 sources), ‘obscured AGN’, which corresponds to a type II AGN or type II QSO template (365 sources), ‘galaxy’, which corresponds to an elliptical, spiral, or starburst galaxy (2475 sources), and ‘star’ (121 sources).

From these sources, in the clean-AGN catalog we have 688 Unobscured AGN, 295 Obscured AGN, 1944 Galaxies, and 76 Stars.

2.3.3 X-ray Classification

Typically, AGN are classified according to their X-ray obscuration by using the Hardness Ratio: $HR = (H - S)/(H + S)$ (where H are the hard-band counts and S the soft-band counts, respectively), or by the intrinsic hydrogen (N_H) column density. Since our sample has a wide dynamic range in redshift, we decided to use the N_H provided by Marchesi et al. (2016b) (computed from spectral fitting) for the brightest sources, and by Marchesi et al. (2016a) (computed from the HR-z curve) for the faintest sources, instead of the HR , which is not corrected by redshift. We divided the sources between obscured and unobscured according to:

1. *X-ray Unobscured AGN (XR I AGN)*: objects with $0 \leq N_H < 10^{22} \text{cm}^{-2}$
2. *X-ray Obscured AGN (XR II AGN)*: objects with $N_H \geq 10^{22} \text{cm}^{-2}$

In the clean-AGN catalog we have 2114 type *I* AGN and 936 type *II* AGN.

2.3.4 Radio Classification

The COSMOS2015 catalog (Laigle et al., 2016) provides photometry in the SuprimeCam B band (Taniguchi et al., 2007, 2015), and VLA fluxes at 1.4 GHz (Schinnerer et al., 2004, 2007, 2010; Bondi et al., 2008; Smolčić et al., 2014). We used this information to classify our sources as radio-loud and radio-quiet using the ratio:

$$R = \frac{L_\nu(5 \text{ GHz})}{L_\nu(4440 \text{ \AA})} \quad (2.1)$$

Where $L_\nu(5 \text{ GHz})$ is the radio luminosity of the source measured at 5 GHz and $L_\nu(4440 \text{ \AA})$ is the B band luminosity. Since the emission in the radio and optical regimes comes from the same source, it does not matter if we use directly the flux instead of the luminosity. We needed to apply a K-correction to the photometry provided by the COSMOS2015 catalog, since the values needed to calculate R have to be in the rest-frame. To do this, we considered that the radio and optical emissions follow a power-law like $F \propto \nu^{-0.8}$ and $F \propto \nu^{-0.44}$ respectively. Therefore, the final flux values used to determine R are:

$$\begin{aligned} F_\nu(5 \text{ GHz})_{\text{rest}} &= F_\nu(1.4 \text{ GHz})_{\text{obs}} \left(\frac{1.4}{5}\right)^{0.8} (1+z)^{-0.2} \\ F_\nu(4440 \text{ \AA})_{\text{rest}} &= F_\nu(4440 \text{ \AA})_{\text{obs}} (1+z)^{-0.56} \end{aligned} \quad (2.2)$$

In the COSMOS2015 catalog, not all sources have a detection at 1.4 GHz: for those sources without a detection reported, we assume that the measured flux corresponds to the 4σ upper limit of $45 \mu\text{Jy}$ reported in Schinnerer et al. (2010). We then use the following criteria to classify our sources:

- Radio-loud (RL): the source has a real detection at 1.4 GHz and $R \geq 10$
- Radio-quiet (RQ): the source has a $R < 10$, including upper limits and real detections at 1.4 GHz.

In the clean-AGN catalog we have 355 Radio-loud AGN and 566 Radio-quiet AGN.

2.4 Light Curve Construction

2.4.1 Calibration and Photometry

For the construction of the light curves we used the time-resolved UltraVISTA images. These images were reduced using the CASU pipeline. The CASU stacks consist of 16 images, one per detector. Detector 16 is known for its unstable gain, therefore we did not consider it in our analysis. We further point spread function (PSF) homogenized the images and applied a photometric re-calibration to match the UltraVISTA DR3³ catalogs.

We perform a PSF homogenization, taking all the images of a certain filter band to a common seeing value, to avoid false variability detection, due to differences in seeing. Each image was convolved with a Gaussian of width equal to $\sqrt{(\sigma_0^2 - \sigma^2)}$, where σ_0 is the width corresponding to some of the worst seeing conditions, and σ is the width of each individual image. Since the behavior of the seeing in the four filters is similar, we selected a fixed value of $\sigma_0 \sim 1.0''$ (or 2.95 pixels) for all the data sets, and discarded all the images with a seeing worst than this value. The mean seeing value of the images is $\sim 0.82''$ for J, H, and Ks, and $\sim 0.89''$ for the Y band. We discarded the 17%, 10%, 8%, and 9% of the total images available for the Y, J, H, and Ks bands respectively.

With these new images, we proceeded with the source detection and photometry, using the public package SExtractor (Bertin & Arnouts, 1996). We generate catalogs for every image in the four NIR bands. The catalogs contain for every source: position, aperture magnitude and flux using an aperture of 2", star classification (CLASS_STAR, given by the shape of the source), and the FLAG parameter, which informs if an object is saturated or has been truncated at the edge of the image.

It is well known that SExtractor underestimates the photometric errors when there are correlated pixels and when the objects are dimmer than the background. Since the PSF homogenisation produce correlated pixels, we calculated the photometric errors using a method similar to the one implemented by Gawiser et al. (2006). We estimated the photometric errors by placing 2000 random circular apertures of a certain size, and measuring the number of counts (F) inside the apertures, taking care to not overlap the apertures with sources detected by SExtractor. We then calculated the standard deviation of the number of counts in the apertures. We repeated this procedure changing the size of the apertures between 1 to 14 pixels in aperture. We then modeled the standard deviation of F as $\sigma_N = \sigma_1 a N^b$, where $N = n_{pix}^{1/2}$, n_{pix} is the number of pixels in the aperture, σ_1 is the standard deviation for an

³http://www.eso.org/sci/observing/phase3/data_releases/uvista_dr3.pdf

aperture of 1 pixel, and a and b are the parameters of the model. For every single image, we calculated the values of a and b , and then calculated the final photometric error for every source as $\sigma_{phot} = \sqrt{\sigma_N^2 + F/\text{gain}}$, with σ_N the number of pixels used in the aperture of 2". The typical values of σ_1 are 1.0, 1.1, 3.5, and 2.0 for the Y, J, H and Ks bands, respectively. The values of a and b are ~ 1.0 and ~ 1.6 , respectively.

Finally, we produced the calibrated catalogs. We used the catalogs generated by SExtractor, and cross-matched them with the DR3 catalog for the corresponding photometric band. We selected all the sources classified as stars by SExtractor ($\text{CLASS_STAR} \geq 0.9$) and with good quality in their photometry ($\text{FLAG} = 0$), and took the difference between their measured magnitudes and the magnitudes according to the DR3 catalogs for the 2" aperture. We then used a linear fit to model these residuals, following a procedure similar to Cartier et al. (2015), to prevent possible non-linearities in the detector, that might be produced after the PSF homogenization, with $m - m_{dr3} = \alpha + \beta \times m$, where m is the magnitude in our catalogs for a certain band, and m_{dr3} the magnitude provided by the DR3 catalogs. Finally, the calibrated magnitudes (m_{cal}) are computed as $m_{cal} = m - (\alpha + \beta \times m)$, and the calibrated errors as $\sigma_{cal} = \sqrt{\sigma_{phot}^2 + \text{var}}$, where var is the variance of $(m - m_{dr3}) - (\alpha + \beta \times m)$. The typical values of α are -0.55, -0.81, -1.42, and -1.91 for the Y, J, H and Ks bands, respectively. For β we normally have a value of ~ 0.0002 , ~ 0.001 , ~ 0.005 , and ~ 0.009 for the Y, J, H and Ks bands, respectively. Therefore, the non-linearities of the detectors are negligible.

2.4.2 Light Curve Generation

We constructed light curves for all the sources in the clean-AGN catalog with a detection in the UltraVISTA single images. To construct the light curves, we cross-matched the clean-AGN catalog with every calibrated catalog, for which we knew their associated Julian dates, using a radius of 1". We discarded the outskirts of the images, considering only the regions with a distance greater or equal to 0.015 degrees to the border. We then constructed light curves for each source, saving only epochs where the SExtractor FLAG parameter was equal to zero. This prevents false detections of variability due to bad photometry. As a rule of thumb, we only saved those light curves with more than three epochs.

After this preliminary construction, we cleaned every light curve following a σ -clipping procedure. First, all the epochs with magnitude error bigger than twice the mean magnitude error were rejected; second, we fitted an order five polynomial to the light curves (which might properly trace the long term variations of the light curve), and rejected all the epochs with a distance from the polynomial bigger than 2σ , with $\sigma^2 = (\sigma_{epoch}^2 + \text{std})$, where σ_{epoch} is the magnitude error in each epoch, and std is the standard deviation of the whole light curve. Then, we only saved those light curves which ended with three epochs or more after the cleaning process. Finally, we transformed every light curve to the AGN rest-frame: $t_{rest} = t_{obs}/(1+z)$. Where t_{rest} is the light curve time at the rest-frame in days and t_{obs} is the observed time. Every variability feature (see section 2.5) was computed in the rest-frame of the AGN. We generated 1715, 1895, 1835, and 2107 light curves with magnitudes up to the 5σ limit in the Y, J, H and Ks bands respectively. Figures 2.1 and 2.2 show examples of variable light curves in the four bands for objects classified as a BL AGN and NL AGN. The

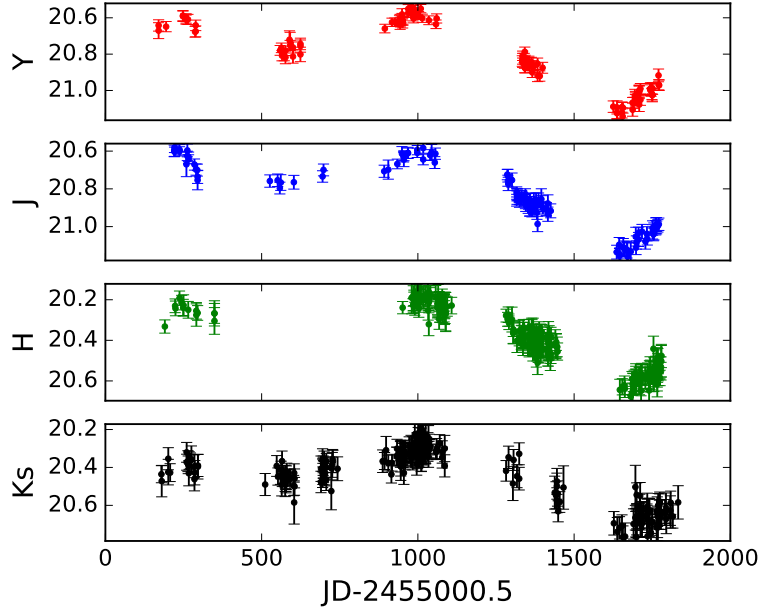


Figure 2.1: Observed light curve for a BL - XR II - TypeI - RQ AGN located at RA= 150.45187° and DEC= 2.144811°, with ID cid_543 from Marchesi et al. (2016a) catalog, located at $z=1.298$, variable in all the filter bands.

light curves are plotted in the observed frame. Figure 2.3 is an example of a non-variable light curve. Figure 2.4 shows the mean error in magnitudes vs the mean magnitude for the final light curves. From the figure we can see that the Y band has the best quality in the photometry, and the Ks band has the worst. This is expected, since as we move to redder bands, the sky brightness increases.

The UltraVISTA survey is split into two sets of strips called deep and ultra-deep (see McCracken et al. 2012). In the first year of UltraVISTA both sets of strips were observed, thereby providing a nearly homogeneous coverage of the field. In the following four years only the ultra-deep strips were observed. This clearly influences both the number of epochs and the total length of the light curves for the objects in this study. Additionally, each strip consists of 3 pointings (pawprint positions), where each OB obtained images were jittered around one such pawprint position. Therefore, we have light curves with a number of epochs that ranges from 3 to 365. Furthermore, the UltraVISTA project did not observe the COSMOS field in every photometric band by the same number of epochs. The band with the best sampled light curves is the Ks band, with a mean number of 92 epochs, followed by the H band, with 85 epochs, the J band with 48 epochs and finally the Y band with 45 epochs, in average. Figure 2.5 shows the number of epochs in the light curves. Most of the light curves have less than 50 epochs. Figure 2.6 shows the rest-frame time length (t_{rest}), defined as the length of the light curve observed length divided by $(1+z)$. From the figure we can see that half of the light curves have a length of less than 200 days.

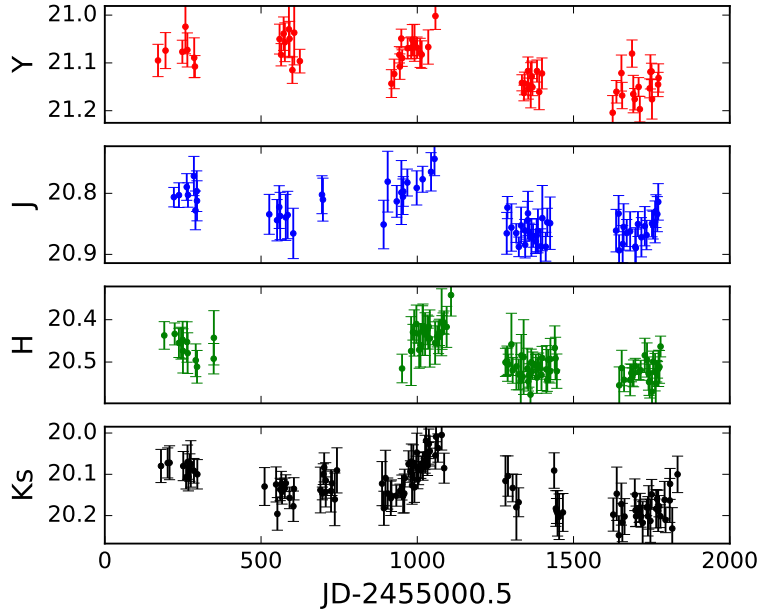


Figure 2.2: Observed light curve for a NL - XR II - Galaxy - RQ AGN located at RA= 150.097790° and DEC= 1.845247°, with ID cid_254 from Marchesi et al. (2016a) catalog, located at $z=0.711$, variable in all the filter bands.

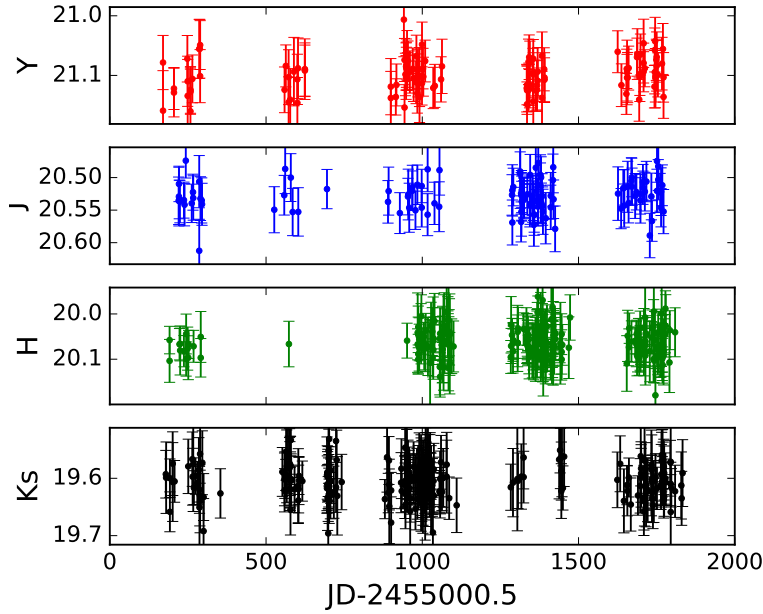


Figure 2.3: Observed light curve for a XR II - Galaxy - RL AGN located at RA= 149.43076° and DEC= 1.939061°, with ID lid_2414 from Marchesi et al. (2016a) catalog, located at $z=0.916$, non variable in all the filter bands.

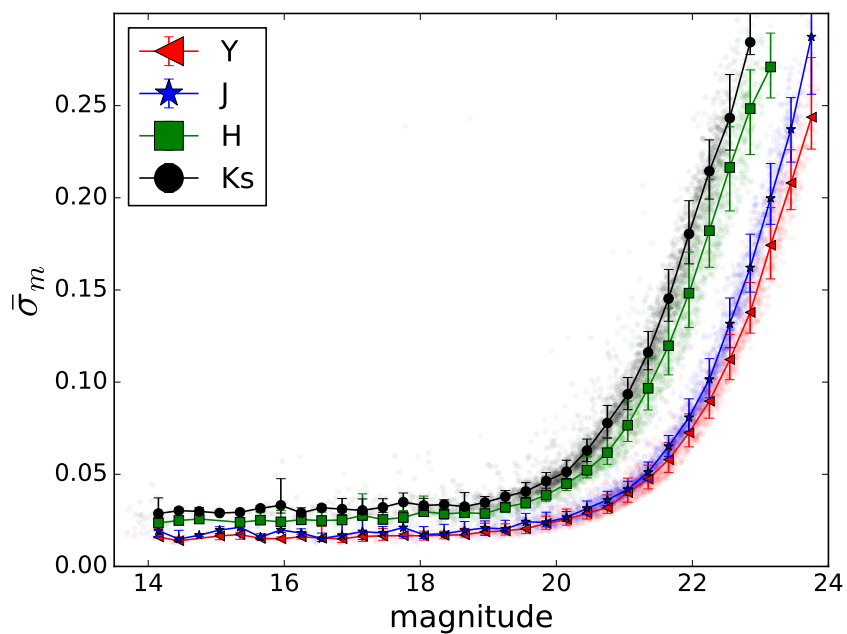


Figure 2.4: Mean magnitude error vs mean magnitude for all the final light curves, for the four photometric bands Y (red), J (blue), H (green) and Ks (black).

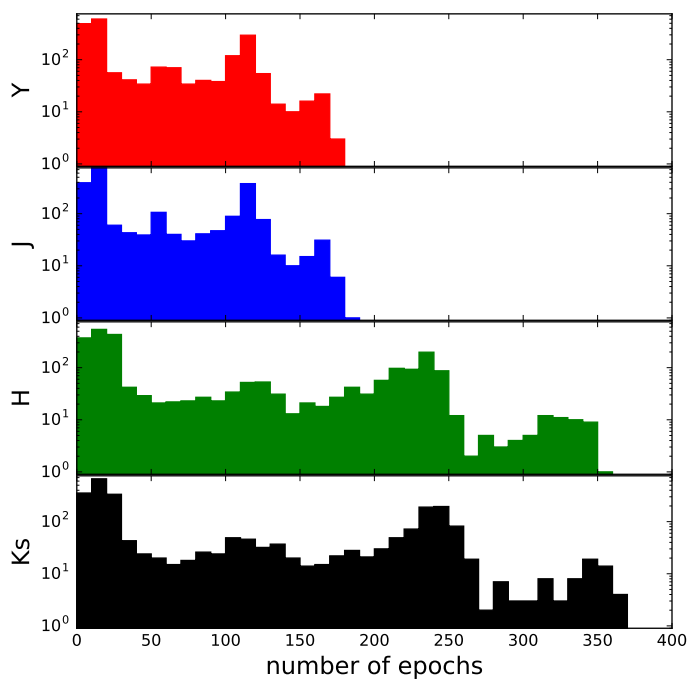


Figure 2.5: Histogram in logarithmic scale of the number of epochs in the light curves for the four photometric bands Y (red), J (blue), H (green) and Ks (black).

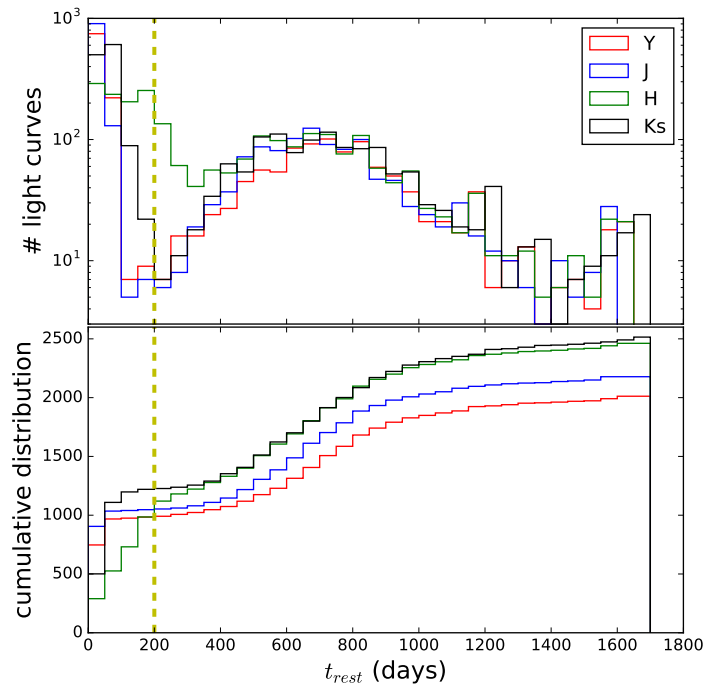


Figure 2.6: Top: Histogram in logarithmic scale of the rest-frame time length (t_{rest}), defined as the light curve's observed length divided by $(1 + z)$, for the four photometric bands Y (red), J (blue), H (green) and Ks (black). Bottom: Cumulative distribution of the rest-frame time length. The yellow dashed vertical line marks $t_{rest} = 200$.

2.5 Variability Analysis

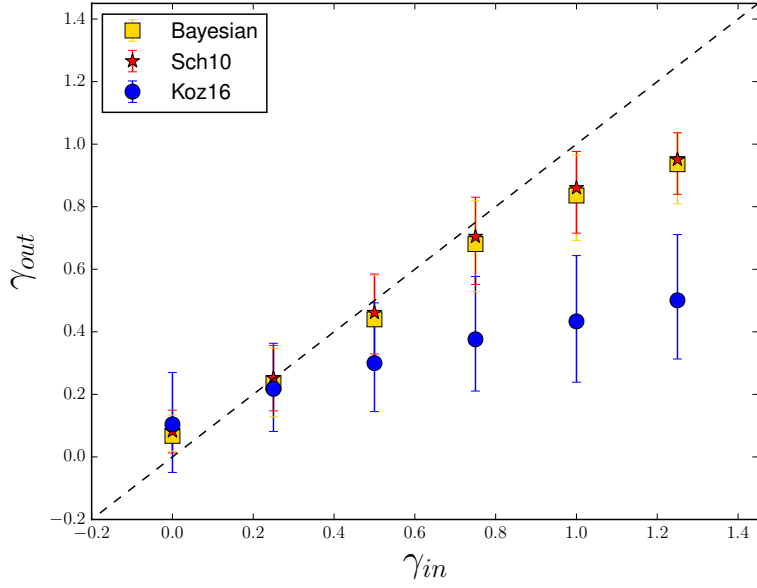
To characterize the variability of the sources we used the variability features described in Section 1.2.2: P_{var} , σ_{rms} , the SF and the DRW process. In the following sections we tested the performance of the SF and DRW statistics.

2.5.1 The Structure Function

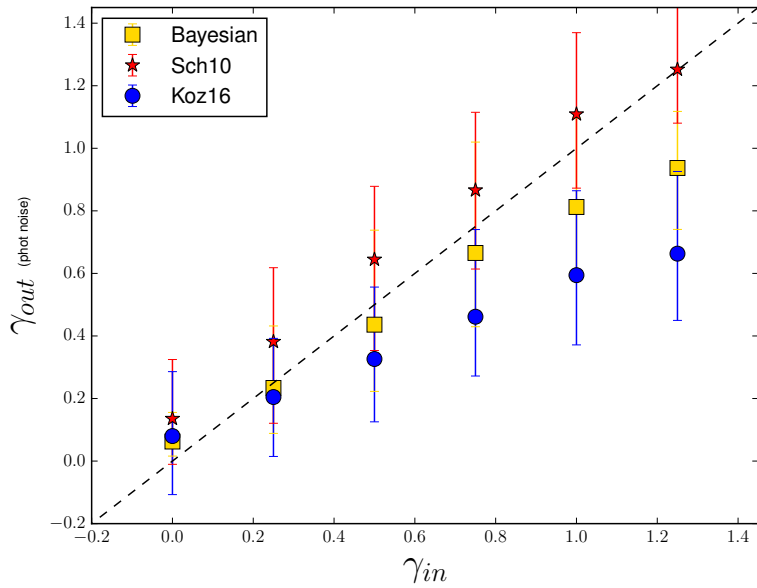
In order to test which definition of the SF presented in section 1.2.2 (standard and Bayesian methods of Schmidt et al. 2010, and the definition provided by Kozłowski 2016) is more suitable for the analysis of our data, we simulated artificial light curves with a power-law PSD, and with a sampling representative of the UltraVISTA light curves in the Y band. To simulate the light curves we used the algorithm proposed by Timmer & Koenig (1995). Then, we analyzed which method recovers the slope of the PSD more accurately. We simulated light curves with $\gamma = [0, 0.25, 0.5, 0.75, 1, 1.25]$. We generated 1000 light curves without photometric noise and 1000 light curves with a photometric noise of 0.02 magnitudes (representative of our light curves in the Y band) for every value of γ . Figure 2.7 shows the results of this analysis. The left panel of the figure shows the results for the light curves without photometric noise. For all these light curves we were able to obtain measurements of the γ parameter by using the three methods. However it can be seen that the best results are obtained with the two methods proposed by Schmidt et al. (2010). The right panel of Figure 2.7 shows the results for the light curves with photometric noise. We only were able to obtain measurements of the Structure Function parameters for all light curves with the method of Kozłowski (2016) and with the Bayesian method of Schmidt et al. (2010). For the case of the analytic definition of the Structure Function proposed by Schmidt et al. (2010) (Eq.1.11), $\sim 70\%$ of the light curves return “NAN” values of γ . These light curves are not considered in the results shown with red stars in the right panel of Figure 2.7. We think this is a consequence of a too high subtraction of the noise term in Eq. 1.11, as previously pointed out by Kozłowski (2016). Therefore, since the Bayesian method proposed by Schmidt et al. (2010) gives the best results and is more stable under the presence of photometric noise, we decided to use this method in the rest of our analysis. All the results presented below for the Structure Function were computed using the Bayesian method.

2.5.2 Damp Random Walk Process

Kozłowski (2017a) presents an analysis of the limitations of the DRW to model AGN light curves. He demonstrated that it is necessary to have light curves with at least 10 times the length of the “relaxation time” in order to have accurate variability parameters derived from the DRW analysis. We tested whether this effect is present when we use *carma_pack* to fit DRW models to our light curves. We simulated DRW light curves with a sampling representative of the light curves in the Y band, following the approach proposed by Kelly et al. (2009) using different values of τ , and then we fitted these light curves using *carma_pack*. We compared the output τ obtained from the method with the input τ used to generate the



(a)



(b)

Figure 2.7: Results of the Structure Function analysis for the artificial light curves with power-law PSD. In the x-axis we show the input value of γ and in the y-axis we show the value of γ computed by the different methods. The error bars in the y-axis correspond to the 15.86 and 84.14 percentiles of the output γ distributions. The blue circles correspond to the results obtained using the definition given in Kozłowski (2016), the red stars correspond to the first definition given by Schmidt et al. (2010) (eq. 1.11), and the yellow squares correspond to the Bayesian method proposed by Schmidt et al. (2010). The left panel corresponds to light curves without photometric noise. The right panel corresponds to light curves with a photometric noise of 0.02 magnitudes. Light curves with “NAN” values of the γ parameter are not included.

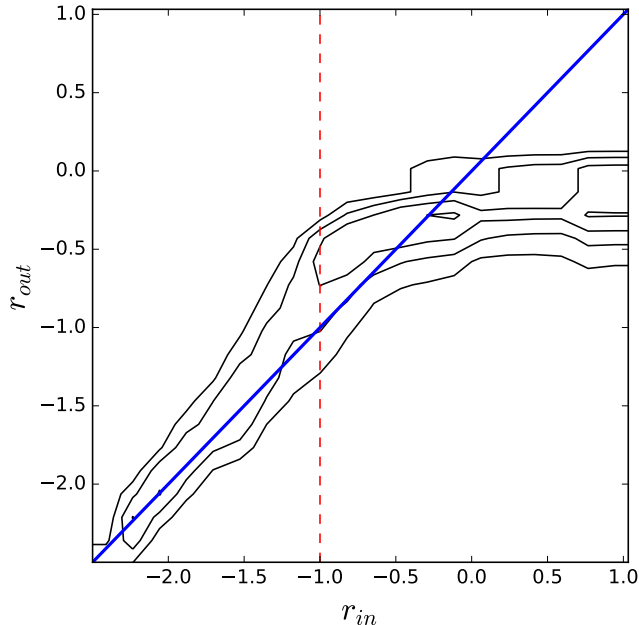


Figure 2.8: Comparison of r_{in} vs r_{out} for simulated DRW light curves. The blue line shows the 1:1 relation. The red dotted line shows the region where the length of the light curve is 10 times the input τ .

light curves. We define the parameter $r = \log_{10}(\tau/t_{lc})$, where t_{lc} is the length of the light curve. Figure 2.8 shows a comparison of r_{in} calculated using the input τ and r_{out} calculated using the output τ . In the figure we confirm the results of Kozłowski (2017a). For light curves whose input τ is less than 10 times the length of the light curve, the output “relaxation time” is close to the original value. For the remaining light curves, the length is too short to give confident results. Since our light curves have a typical length of 5 years, we would be able to detect accurately values of τ lower than ~ 180 days. After correcting by redshift, the length of most of our light curves is too short to measure τ accurately. Therefore we decided not to include the DRW analysis in our results.

2.6 Results

We define a source as intrinsically variable, when its light curve has $P_{var} \geq 0.95$ and $(\sigma_{rms}^2 - err(\sigma_{rms}^2)) > 0$. We found that 13.47%, 11.13%, 5.4% 6.22% of the total number of sources in the Y, J, H and Ks bands, respectively, are variable. However, as seen in Figures 2.5 and 2.6, there are several light curves with poor sampling. Therefore, in order to homogenize our analysis, we only considered well-sampled light curves, that is, light curves with at least 20 epochs and with a rest-frame time length (t_{rest}) greater or equal to 200 days. Besides, we only considered sources with either spectroscopic or photometric redshifts available (we use the best redshift reported by Marchesi et al. 2016a), and with a total Ks magnitude (equivalent to SExtractor ‘AUTO_MAG’) brighter than 22.0. We call this

downsized sample the “clean-sample”.

In this section we present the results of our variability analysis. Table 1 summarizes the number of light curves available for each object type in each NIR band before and after downsizing. The first value of every entrance in Table 1 corresponds to the results for the clean-sample. The second value gives the numbers for the whole data set, before we downsized the sample. We found that 27.4%, 21.7%, 9.1% and 11.5% of the sources from the clean-sample in the Y, J, H and Ks bands, respectively, are variable. As can be seen from the Table 1, we only miss a small fraction ($\sim 10\%$) of variable sources after the sample is downsized, while the relative fraction of variable sources doubles.

The Y and J bands have the best quality in the photometry (see Figure 2.4), and the larger fraction of variable sources. We will focus on the analysis of the clean sample in these bands in the following sections, unless otherwise noticed.

Figure 2.9 shows the redshift distribution of the variable and non-variable sources for the different classifications in the Y band. From the figure we can see that there are not important differences in the distributions of the variable and non variable sources, except for the radio classification, where we can see that the non-variable sources are clustered at lower redshifts. This is produced by the spectroscopic classification of the sources with radio classification, since most of the non-variable sources are NL and are located at low redshifts, and most of the variable sources are BL and are located at higher redshifts.

2.6.1 Variability properties

Figure 2.10 shows the distribution of the power law parameters of the Structure Function A and γ , for the variable sources in the Y and J bands in logarithmic scale. We mark the sources according to their spectroscopic and X-ray classifications. Histograms at the top and right hand side of the plots better represents the normalized distributions of variable sources of different AGN populations.

The mean and 1σ errors of the Structure Function parameters for the variable sources are: $A_Y = 0.15^{+0.16}_{-0.07}$ and $\gamma_Y = 0.62^{+0.42}_{-0.32}$ for the Y band, and $A_J = 0.13^{+0.12}_{-0.09}$ and $\gamma_J = 0.63^{+0.56}_{-0.29}$. Clearly several sources have values of γ consistent with DRW process. However, there are some sources with $\gamma > 1.0$ (36 and 43 for the Y and J bands, respectively), which implies deviations from a DRW process ($\gamma = 0.5$). These results are consistent with previous analysis, which have found that CARMA models with higher orders (and not a simple DRW model) better describe AGN light curves (e.g., Kasliwal et al. 2015, 2017; Simm et al. 2016).

From Figure 2.10 we can see that the distribution of the A parameter has a noticeable difference when we compare the sources classified as BL and NL, in particular in the Y band. On the other hand, when we compare the XR I and XR II sources the difference is less evident. In order to have a more quantitative comparison of the Structure Function parameters distributions, we performed a two-sample Anderson-Darling test (Pettitt, 1976) for the A and γ parameters considering the spectroscopic and X-ray classification. Since the Anderson-Darling test does not take into account the errors of the parameters, we only

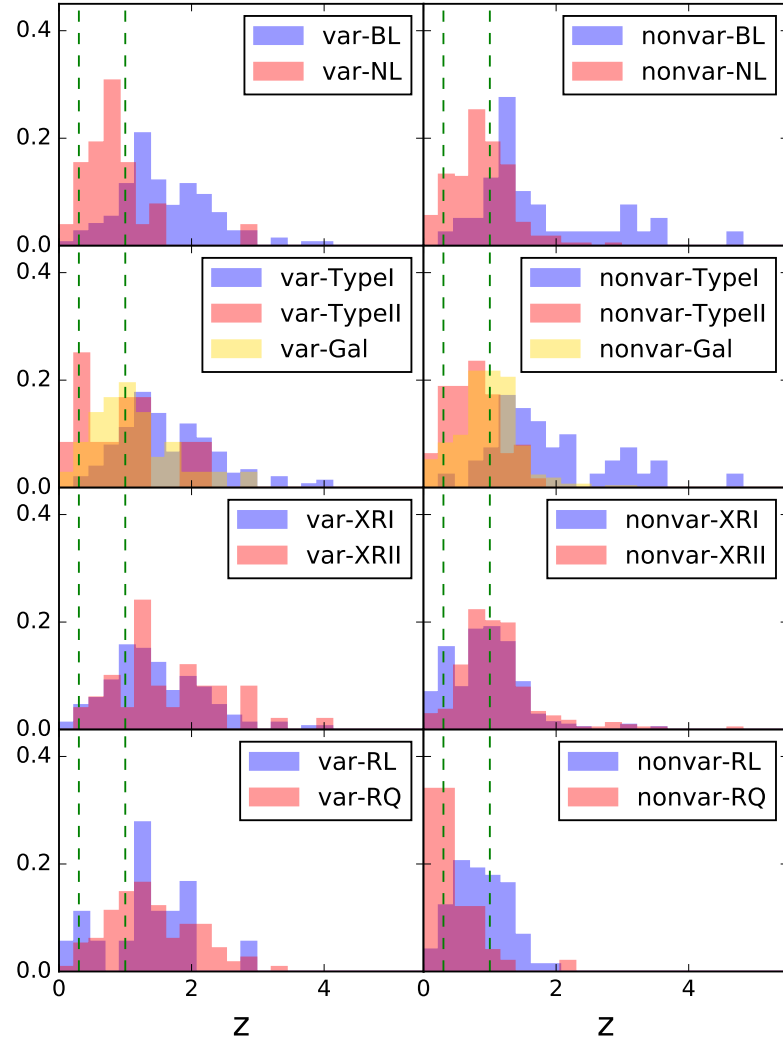


Figure 2.9: Normalized histogram of the redshift distribution of the variable and non variable sources in the Y band for the four classes defined in Table 1. The green dashed lines show the 0.3 to 1 redshift bin. From top to bottom: spectroscopic, photometric, X-ray and radio classifications.

Table 1: Number of well-sampled light cuves before and after downsizing: clean-sample / whole data set. In brackets we show the number of variable sources.

Classification		Y	J	H	Ks
Spectroscopic	BL	196 / 432 (152 / 162)	203 / 447 (129 / 141)	207 / 425 (64 / 67)	223 / 470 (79 / 83)
	NL	319 / 677 (28 / 31)	343 / 703 (29 / 33)	405 / 717 (12 / 13)	388 / 781 (19 / 24)
Photometric	Type I	205 / 465 (159 / 170)	211 / 488 (138 / 151)	216 / 442 (67 / 70)	237 / 503 (81 / 87)
	Type II	91 / 198 (13 / 13)	96 / 207 (14 / 16)	115 / 212 (5 / 5)	118 / 241 (11 / 12)
	Galaxy	478 / 1048 (39 / 46)	564 / 1196 (36 / 42)	637 / 1176 (16 / 24)	663 / 1357 (25 / 32)
X-ray	XR I	368 / 793 (153 / 161)	395 / 837 (133 / 149)	450 / 814 (63 / 65)	455 / 907 (81 / 87)
	XR II	293 / 636 (50 / 58)	334 / 731 (49 / 53)	351 / 702 (20 / 25)	396 / 837 (32 / 37)
Radio	RL	106 / 232 (18 / 18)	116 / 243 (16 / 18)	140 / 269 (8 / 9)	136 / 287 (10 / 11)
	RQ	185 / 408 (118 / 128)	183 / 392 (109 / 121)	231 / 436 (58 / 65)	210 / 439 (71 / 77)
Total		777 / 1715 (213 / 231)	874 / 1895 (190 / 211)	971 / 1835 (88 / 99)	1021 / 2107 (117 / 131)

considered in the test those variable sources with a measured parameter having a signal to noise ratio higher than 3, considering as the error of the parameter the average of the lower and upper errors given by the Bayesian analysis. According to the test, for the BL and NL, the distributions of the A parameter are different at a 99.5% significance level in both Y and J bands, with a p_{value} of 8.99×10^{-6} and 1.69×10^{-4} for the Y and J respectively. For the XR I and XR II sources, no statistically significant difference is found for the A parameter, with a p_{value} of 1 and 0.87 for the Y and J, respectively. For the case of the γ parameter, no statistically significant differences are found in any filter for the spectroscopic and X-ray classifications. The p_{value} are 0.052 and 0.076 for the Y and J bands in the spectroscopic classification, and 0.48 and 0.16 for the Y and J bands in the X-ray classification.

Since the dynamic range in redshift considered in this analysis is wide, we repeated the analysis selecting the smallest bin of redshift where we can ensure the presence of at least 6 sources belonging to every population (BL, NL, XRI and XR II), in order to have confident results from the Anderson–Darling test. This requirement is accomplished by the bin of redshifts between 0.3 and 1 (see Figure 2.9). In this bin we expect to observe emission coming from the accretion disk in both Y and J bands. The results of the Anderson–Darling test are the same. The distributions of the A parameter for the spectroscopic classification are different at a 99.5% significance level in both Y and J bands, with a p_{value} of 0.01 and 0.03 for the Y and J, respectively. For the case of the A parameter in the X-ray classification (p_{value} of 0.22 and 0.1) and the γ parameter in both classifications (spectroscopic: p_{value} of 0.41 and 0.40; X-ray: p_{value} of 0.39 and 0.23), no statistically significant differences are found in any filter.

Figures 2.1 and 2.2 show examples of light curves variable in the four photometric bands. We plotted the light curves of these two sources because they are particularly interesting. In Figure 2.1 the source is classified as BL - XR II. On the other hand, in Figure 2.2, the source is classified as NL - XR II. Even though this source is NL and obscured in the X-ray regimes, we can detect its variation in the four photometric bands. Its variability parameters in the Y filter band are: $A = 0.05$ (mag/year) and $\gamma = 0.69$. This source could potentially be an example of “Changing look” AGN.

2.6.2 Dependency with redshift and Luminosity

Redshift will obviously change the rest-frame emission observed by each band. Any correlation with redshift, therefore, needs to take this into account. Besides, previous analysis have shown evidence of an anti-correlation between optical/UV luminosity with the variability amplitude (e.g., Uomoto et al. 1976; Hook et al. 1994; Trevese et al. 1994; Cristiani et al. 1997; Wilhite et al. 2008; Kelly et al. 2009; MacLeod et al. 2011; Meusinger & Weiss 2013; Simm et al. 2016; Caplar et al. 2017). In other words, a more luminous (and probably larger) system varies, at a given fractional amplitude, on larger time scales.

In order to test any possible correlation of the Structure Function parameters with luminosity, we used bolometric luminosities (L_{BOL} ; the total AGN luminosity) from Lusso et al. (2012), which were determined for type 1 and type 2 AGN through SED-fitting. We also test any correlation with the intrinsic Hard X-ray luminosity (HL_{int} ; the AGN X-ray lumi-

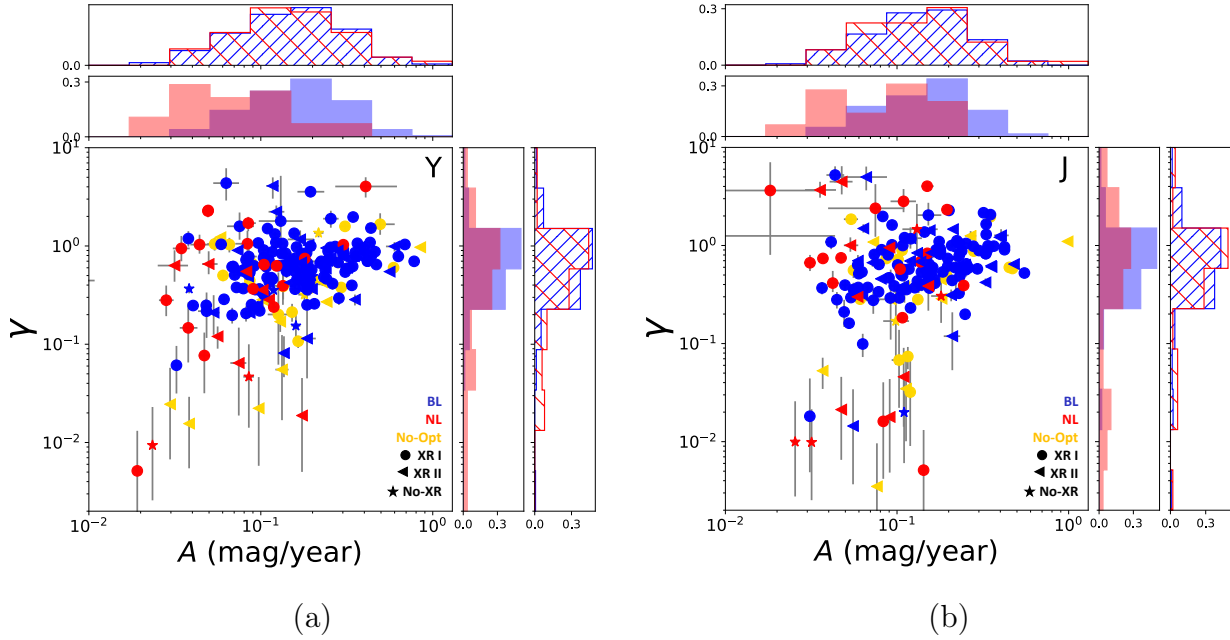


Figure 2.10: Distribution of the Structure Function power law parameters A and γ , for the variable sources, in the Y and J bands, in logarithmic space. The blue sources correspond to BL AGN, the red sources to NL AGN, and the yellow to sources without spectroscopic classification. The circles correspond to XR I AGN, the triangles to XR II AGN, and the stars to sources without X-ray classification. The error of the measurements are shown with grey error bars. For most of the sources, the size of the error bars is smaller than the marker size. Along the axes we show the projected A and γ distributions for the BL AGN (blue shaded), NL AGN (red shaded), XR I AGN (blue hatched), and XR II AGN (red hatched).

osity in the 2-10 keV range) (i.e., rest-frame luminosity corrected by absorption), using the Chandra X-ray data from Marchesi et al. (2016a). The value of HL_{int} was computed using the rest-frame observed luminosity (HL_{obs}) and the luminosity absorption correction (k_{corr}) provided by Marchesi et al. (2016a): $HL_{int} = HL_{obs}/k_{corr}$. We tested the correlations with luminosity and redshift for those variable sources with measured values of both L_{BOL} and HL_{int} , which corresponds to 139 and 123 sources in the Y and J bands, respectively.

We calculated the Spearman rank-order correlation coefficient r_s for $\log(A)$ vs $\log(1+z)$ and for $\log(\gamma)$ vs $\log(1+z)$ in the Y and J bands. We find clear evidence of positive correlation between A and $(1+z)$. The values of the coefficient for the correlation between A and $(1+z)$ are $r_s = 0.47$ ($p_{value} = 5.5 \times 10^{-9}$) and $r_s = 0.54$ ($p_{value} = 1.4 \times 10^{-10}$) for Y and J respectively. For the case of $\log(\gamma)$ vs $\log(1+z)$, the correlation is not evident. The value of the Spearman rank-order correlation coefficients are $r_s = 0.08$ ($p_{value} = 0.38$) and $r_s = 0.1$ ($p_{value} = 0.29$) for Y and J, respectively.

For the case of L_{BOL} , the Spearman coefficient showed some evidence of a positive correlation between $\log(A)$ and $\log(L_{BOL})$, contrary with what was expected from previous analysis. The values of the coefficient are $r_s = 0.3$ ($p_{value} = 3.7 \times 10^{-4}$) and $r_s = 0.3$ ($p_{value} = 6.6 \times 10^{-4}$) for Y and J, respectively. However, this result might be affected by the wide dynamic range in redshift considered in the sample ($z \sim 0.3 - 4$). Thus, in order to disentangle whether the positive correlation is driven by redshift or by luminosity, further analysis is needed. Moreover, the correlation analysis showed no evidence of correlation between $\log(\gamma)$ and $\log(L_{BOL})$. The coefficients are $r_s = 0.08$ ($p_{value} = 0.32$) and $r_s = -0.04$ ($p_{value} = 0.65$) for Y and J, respectively.

Finally, for the case of HL_{int} , the correlation coefficient showed a weak (or non) positive correlation between $\log(A)$ and $\log(HL_{int})$, with $r_s = 0.19$ ($p_{value} = 0.024$) and $r_s = 0.15$ ($p_{value} = 0.093$) for Y and J, respectively. On the other hand, for $\log(\gamma)$ and $\log(HL_{int})$, the correlation is negligible, with $r_s = 0.01$ ($p_{value} = 0.9$) and $r_s = 0.01$ ($p_{value} = 0.87$) for Y and J, respectively. Similar to the case of L_{BOL} , we need further analysis to say if the correlation between $\log(A)$ and $\log(HL_{int})$ is affected by the wide dynamic range in redshift.

To test whether the positive correlation of the amplitude of the Structure Function with luminosity can be due by a positive correlation with redshift, we calculated the correlation of A with both L_{BOL} and redshift in the logarithmic space, i.e., we computed

$$\log_{10}(A) = a \log_{10}(L_{BOL}/10^{45} \text{ erg s}^{-1}) + b \log_{10}(1+z) + c$$

for the same sources considered in the previous analysis. For this purpose, we computed the Weighted Least Squares linear regression (WLS), considering as weights the inverse of the variance of $\log_{10}(A)$, calculated as $\sigma^2(\log_{10}A) = (0.434 * (A_{loerr} + A_{uperr})/2A)^2$. A summary of the regression for the Y and J filter bands can be found in Table 2. From the table, we can see that the correlation between $\log(A)$ and $\log(L_{BOL})$ is in fact negative, and it is statistically significant but weak, and the correlation between $\log(A)$ and $\log(1+z)$ is positive and significant. Therefore, the positive correlation between $\log(A)$ and $\log(L_{BOL})$ obtained using the Spearman coefficient was actually an effect of the positive correlation with redshift. Thus, whenever we perform a correlation analysis for the luminosity, we need to consider the redshift of the source as a second independent variable.

In Table 2 we can also see the results for the regression of A with both HL_{int} and redshift in the logarithmic space. In this case, we computed

$$\log_{10}(A) = a\log_{10}(HL_{int}/10^{43}\text{erg s}^{-1}) + b\log_{10}(1+z) + c$$

The results of the WLS analysis show that the correlation with HL_{int} is not statistically significant in any band. Therefore the weak positive correlation observed in the previous analysis might be produced by the positive correlation with redshift.

Table 2: Results of the WLS. p_{values} in brackets.

Filter	luminosity	a luminosity	b redshift	c intercept
Y	L_{BOL}	-0.11 ± 0.05 (0.035)	1.38 ± 0.29 (0.0)	-1.22 ± 0.1 (0.0)
Y	HL_{int}	-0.05 ± 0.03 (0.075)	1.14 ± 0.26 (0.0)	-1.12 ± 0.1 (0.0)
J	L_{BOL}	-0.13 ± 0.05 (0.018)	1.85 ± 0.26 (0.0)	-1.42 ± 0.09 (0.0)
J	HL_{int}	0.003 ± 0.03 (0.9)	1.52 ± 0.24 (0.0)	-1.38 ± 0.09 (0.0)

The positive correlations with redshift are consistent with observing bluer regions of the AGN SED as redshift increase, as bluer emission is expected to vary with larger amplitude. Figure 2.11 shows A vs rest-frame wavelength of emission (λ) at logarithmic scale for the variable sources, in the Y and J bands. There is a clear anti-correlation between A and λ , and the result of the linear regression for the Y and J bands are consistent at 95% level. However, there is a large dispersion in the correlation. This can be related to other properties of AGN that may affect the amplitude of the variability aside from the emission wavelength, like the bolometric luminosity (which, as we already demonstrated, anti-correlates with the amplitude of the variability), the black hole mass, the accretion rate, among other physical properties. As well as the length and quality of the light curves, among observational factors.

The lack of correlation between γ and redshift might indicate that the structure of the variability is independent of wavelength. This is consistent with previous analysis of optical and NIR light curves that claim that light curves of AGN observed at different wavelengths have the same structure or shape in time scales of months to years, but showing time lags between them, due to the distance between the emitting regions, and showing a decrement in the amplitude of the variation (Lira et al., 2011, 2015).

Figure 2.12 shows the absolute magnitude for the variable sources (no k-corrected) vs redshift (bottom axis) and vs rest-frame wavelength of emission, for the Y band (top axis). In the figure we mark those sources according to their spectroscopic and X-ray classifications, as before. At the bottom of the figure we show the fraction of variable BL and NL sources in logarithmic scale. The number of variable NL sources detected at $0.5\mu\text{m} < \lambda_{rest} < 1\mu\text{m}$ is low ($\lesssim 10\%$). However, since this wavelength range is expected to be dominated by emission from the accretion disk and not directly observable in most of the obscured systems, we

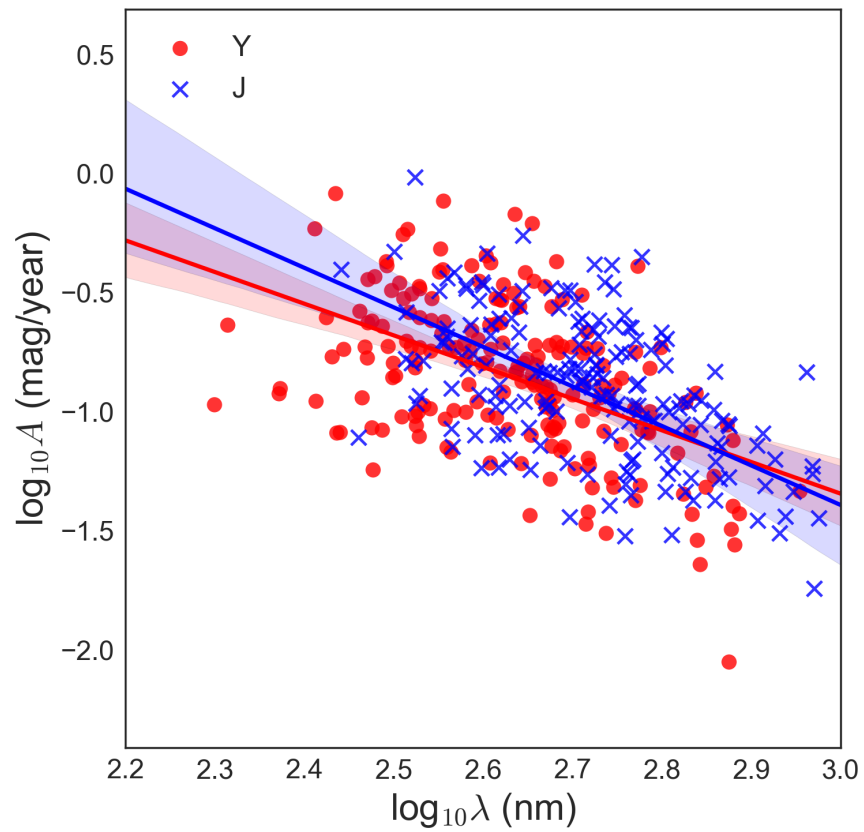


Figure 2.11: Amplitude of the Structure Function A vs rest-frame wavelength of emission (λ) at logarithmic scale for the variable sources, in the Y (red circles) and J (blue crosses) bands. The blue and red lines show the linear regression between $\log_{10}(A)$ and $\log_{10}(\lambda)$ for the Y and J bands respectively. The shaded regions show the 95% error of the regression.

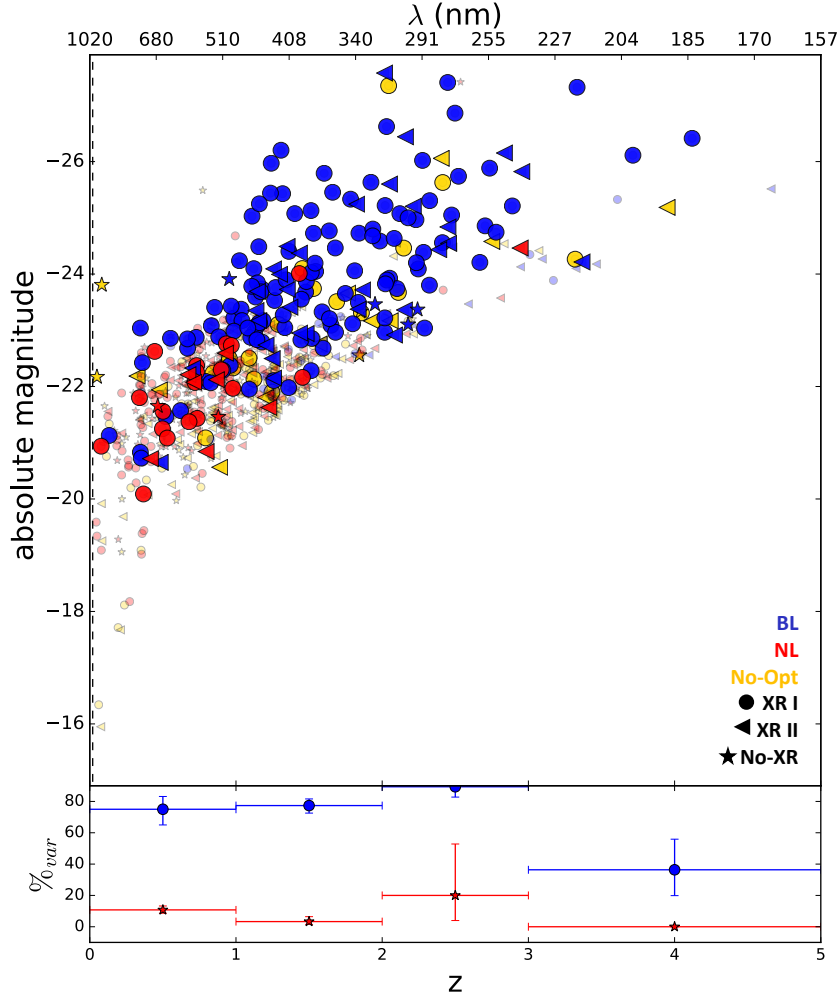


Figure 2.12: Top: absolute magnitude (no k -corrected) vs redshift for the Y band. The small shaded sources are non-variable. The blue sources correspond to BL AGN, the red sources to NL AGN, and the green to sources without spectroscopic classification. The circles correspond to XR I AGN, the triangles to XR II AGN, and the stars to sources without X-ray classification. As a reference, we include in the top x axis the value corresponding to the rest-frame wavelength of emission. Bottom: fraction of variable BL (blue) and NL (red) sources.

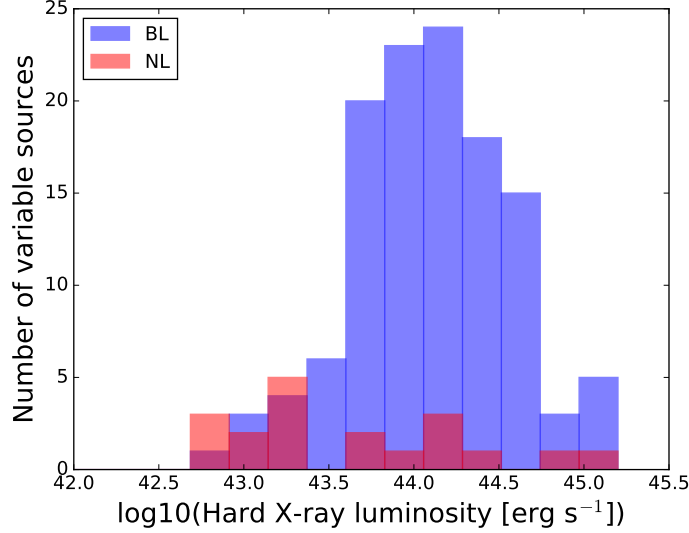


Figure 2.13: Histogram of the intrinsic luminosity (rest-frame and corrected for absorption) in the hard X-ray band, of variable sources in the Y band, considering the spectroscopic classification. BL are showed in blue and NL in red.

would not expect to detect high variability from NL sources in this regime. These sources, however do not show the same variability properties as their BL counterparts, and cluster at significantly smaller values of A and σ in Figure 2.10.

As torus emission is not expected around rest-frame $0.5\mu\text{m}$, the most likely explanation is that these variable NL sources correspond to BL AGN but where the host galaxy might be damping the variability signal (hence, yielding smaller values of the parameter A), and masking the presence of weak broad lines in the spectra. One way to check this is by looking at the intrinsic X-ray luminosity in the hard band (rest-frame luminosity and corrected for absorption), for variable BL and NL AGN. This is presented in Figure 2.13. It clearly shows that most of the NL variable sources have low luminosity ($HL_{int} < 10^{44} \text{ erg s}^{-1}$). We therefore discard “True type II” AGN as a possibility to explain our low luminosity variable NL sources, since we might expect to observe that their variability properties, like the Structure Function parameters, are similar to the properties of normal BL sources, which is not seen for most of our variable NL sources. For the case of the few bright NL variable sources, we cannot discard that they are “True type II” AGN.

To have a better understanding of the nature of these variable NL sources, we plot in Figure 2.14 a histogram of HL_{int} for the NL variable sources, split by their photometric classification. From the figure we can see that most of the low luminosity sources were adjusted by a galaxy template in Marchesi et al. (2016a). This result supports our idea that most of the low luminosity variable NL sources correspond to BL sources whose emission is overshadow by their host galaxy. However, a few sources have a photometric classification of Type I. These sources can be either “True type II” AGN or normal BL whose optical spectrum does not sample the region where the broad lines are present, since the SED analysis reveals a continuum emission that is consistent with continuum emission of BL sources.

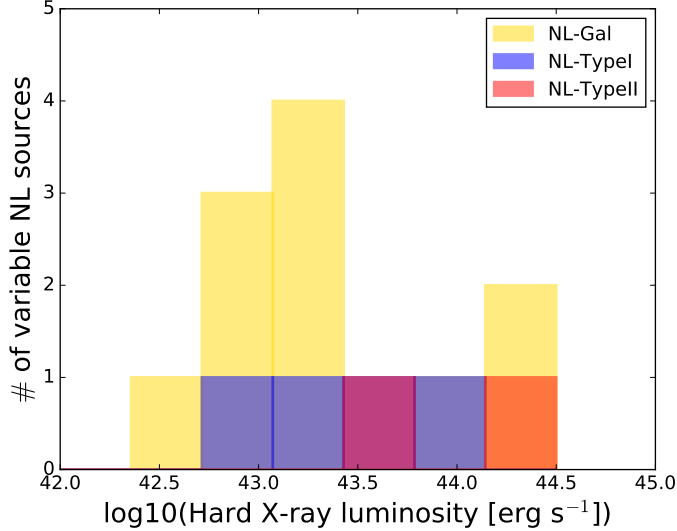
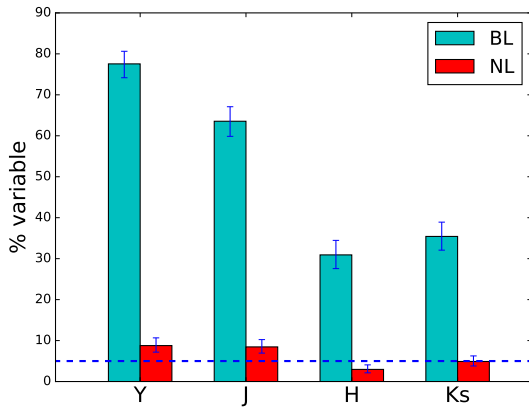


Figure 2.14: Histogram of the intrinsic luminosity (rest-frame and corrected for absorption) in the hard X-ray band, of variable NL sources in the Y band, split by their photometric classification. NL-TypeI are showed in blue, NL-TypeII in red, and NL-Gal in yellow.

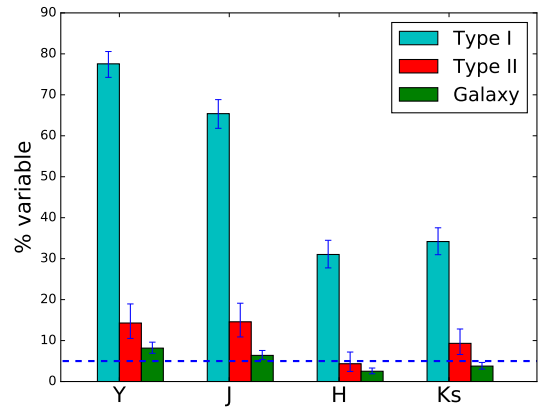
2.6.3 Fraction of variable sources

In this section we show the results of our variability analysis considering the fraction of variable sources. The results shown in this section are for the four photometric bands. Figure 2.15 (a) shows the fraction of variable sources for BL AGN and NL AGN (also see Table 1). We can see that BL AGN have a higher fraction of sources with detected variability in comparison with NL AGN. Besides, we can see a decrement in the fraction of variable BL AGN as we move to redder bands. This can be related to the increment of the photometric errors from the Y to Ks bands. However, we also have to consider that for centrally driven variations it is expected that the amplitude of the variability is lower at longer wavelengths, and therefore we might expect to have a reduction in the fraction of variable sources detected as we move to longer wavelengths. This is in fact observed in BL AGN, where the fraction of variable sources reduces systematically from the Y to the H band. The increase observed in the Ks band could be accounted by the presence of the torus, which has a large solid angle as seen by the innermost region of the disk, hence boosting its variability.

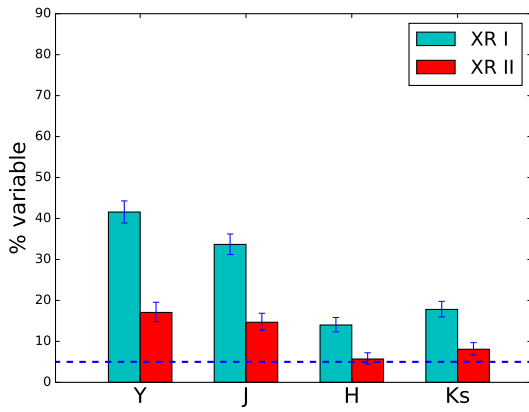
For the case of NL AGN we observe that the fraction of variable sources is above the 5% only for the Y and J bands. In the context of the unified model, we might expect a very low probability to detect variable sources for the Y and J bands, and an increment in the fraction for the H and Ks bands, since the level of obscuration produced by the dust is higher in the optical range than in the NIR. However, some optical variability might be expected for NL sources considering that the obscuring material is a distribution of moving clumps or clouds. Besides, as mentioned in section 2.3.1, the classification of NL sources considers the lack of visible broad emission lines. For the case of BL sources located at redshift ~ 1 , depending of the spectral coverage, the typical broad line components might be out of the spectra, and therefore be classified as NL. For these sources, we might expect to detect optical variability.



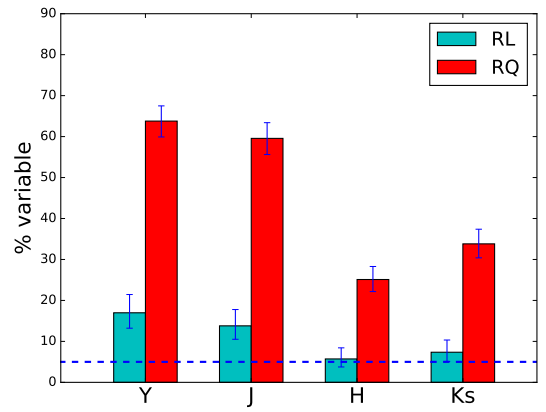
(a)



(b)



(c)



(d)

Figure 2.15: Fraction of variable sources for the spectroscopic (a), photometric (b), X-ray (c), and radio (d) classifications, for our four photometric bands. The blue dashed line demarks the 5%, below this fraction we might expect to have several false positive variable sources. The error bars were calculated considering poisson statistic (Gehrels, 1986)

For the case of the photometric classification, we show in Figure 2.15 (b) the fraction of variable sources in the four photometric bands (also see Table 1). We can see that the highest fraction of variable sources is for the Type I objects. They have a similar fraction of variable sources than BL AGN. The fraction of Type II variable sources is lower than Type I sources, but slightly larger than that of NL AGN. Sources classified as Galaxies show the lowest fraction of variability. It is important to notice that for most of sources best-fitted with a galaxy template by Marchesi et al. (2016a), the X-ray luminosity is $> 10^{42}$ erg/s, and therefore the sources are most likely AGN, although their optical-IR SED does not clearly show this. As before, variability is revealing unobscured AGN in sources where a SED analysis predicts otherwise.

We show in Figure 2.15 (c) the fraction of variable sources for the X-ray classification (also see Table 1). From the figure we can notice that XR I objects have a larger fraction of variable sources than XR II objects. Another obvious result is the lower fraction of X-ray classified variable sources when compared with those with an optical classification. This is not a luminosity effect. In fact, the mean of the X-ray luminosity of the X-ray sample is $10^{43.4 \pm 0.7}$ erg s $^{-1}$, while the same value for the sub-sample with optical spectroscopy corresponds to $10^{43.5 \pm 0.7}$ erg s $^{-1}$. We also checked whether this result is related to the method used to separate obscured and unobscured X-ray sources. We repeated the analysis separating the sources by their Hardness ratio. We considered a source as unobscured in X-rays if $HR < -0.2$. We obtained similar results than the ones showed in Figure 2.15 (c). Therefore this result is not produced by the definition of X-ray obscuration used.

To understand in a better way this difference, we plot in Figures 2.16 (a) and (b) the fraction of variable sources, considering the X-ray and spectroscopic classification, and the X-ray and photometric classification, respectively. From the figures, we can see that the difference in the fraction of variable sources is more closely related with the optical obscuration of the sources than with the X-ray obscuration. Sources unobscured in the optical range are the ones with the largest fraction of variable objects, irrespective of their X-ray classification. In fact, sources classified as unobscured in the X-rays but obscured in the optical range (NL - XR I or Type II - XR I) have a lower fraction of variable objects than sources classified as obscured in the X-rays and unobscured in the optical range (BL - XR II or Type I - XR II). As a significant fraction of XR I sources have a NL classification (136/368, 142/395, 175/450 and 159/455, in the Y to K bands, respectively), this explains the low fraction of variable sources classified as unobscured in X-rays. We repeated the analysis separating the sources by their HR. Again, we obtain similar results, and therefore our result are not biased by the definition of X-ray obscuration. We also repeated the analysis for the redshift bin $z \sim 0.3 - 1.0$, and the results were consistent with what we found for the whole sample.

These results might be related to the differences in the origin of the obscuration in the optical and X-ray regimes. Marchesi et al. (2016b) analyzed the X-ray spectral properties for a sub-sample of the Chandra COSMOS-Legacy Survey catalog, and showed that most of the sources classified as BL - XR II have $L_{2-10\text{keV}} > 10^{44}$ erg s $^{-1}$. They conclude that the existence of these objects suggests that optical and X-ray obscuration can be caused by different mechanisms, and that the X-ray obscuration might be due to dust-free material surrounding the inner part of the nuclei. Merloni et al. (2014) presented a detailed discussion on the nature of sources that have inconsistent classification in the optical and X-ray regimes.

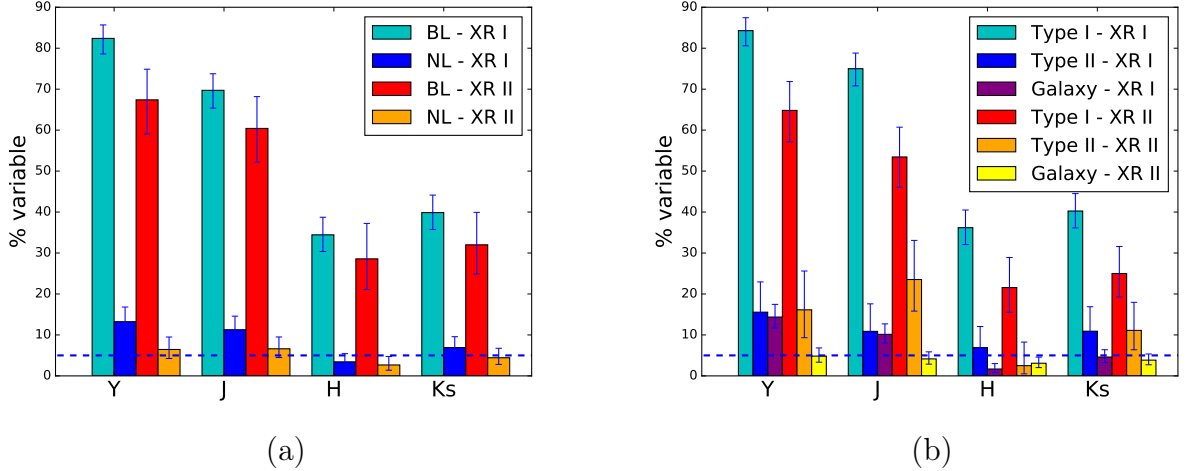


Figure 2.16: Fraction of variable sources for the spectroscopic/X-ray classification, and for the photometric/X-ray classification. The blue dashed line demarks the 5%, below this fraction we might expect to have several false positive variable sources. The error bars were calculated considering poisson statistic (Gehrels, 1986)

They found that sources classified as BL - XR II (type-12 in Merloni et al. 2014) tend to have higher luminosities. Additionally, they showed that the main differences between sources classified as BL - XR II and sources classified as BL-XRI (or type-11) happen in the X-ray regime. They demonstrated that the optical spectra and the SED of these two populations do not show substantial differences, and proposed that the excess absorption in the X-ray regime presented in the type-12 sources could be produced by dust-free material within (or inside) the broad line region. Our variability analysis strengthens this idea, since we do not observe significant differences in the variability features and in the fraction of variable sources between the BL - XRI and BL - XR II populations.

Marchesi et al. (2016b) showed that most of the sources classified as NL - XR I have $L_{2-10\text{keV}} < 10^{43}\text{erg s}^{-1}$, and they expected a fraction of these sources to be “True type II” AGN. Merloni et al. (2014) found that the sources classified as NL - XR I (or type-21) tend to have lower luminosities, $L_X < 10^{44}\text{erg s}^{-1}$, and that their composite spectrum and SED reveals evidence of host galaxy dilution. They proposed that a minority of these objects could be “True type II” AGN. Our variability analysis agrees with the findings of Merloni et al. (2014), since of the 18 sources classified as variable NL - XR I in the Y band, 10 have photometric classification of galaxy, 4 are Type II, and 4 are Type I. Therefore, these 10 NL - XR I - Galaxy sources can be a clear example of BL sources with host galaxy dilution, but whose variability can be detected. The 4 NL - XR I-Type I sources can be examples of “True type II” AGN or BL AGN whose optical spectrum does not cover the region where the broad line are present, since their Structure Function features are in agreement with those of BL sources ($A = [0.12, 0.13, 0.18, 0.12]$, $\gamma = [0.63, 0.37, 0.76, 0.67]$), and $z = [1.43, 0.98, 0.73, 0.74]$. Finally, the 4 sources classified as variable NL-Type II, probably are examples of optically obscured variable sources.

Figure 2.16 (d) shows the fraction of variable sources for the Radio Classification. In this case, the Radio Quiet sources show a considerably higher fraction of variable sources than

the Radio Loud sources. A 72% of the RL and a 60% of the RQ sources are classified as NL AGN. The large fraction of RL objects with a NL classification is not surprising. RL objects represent a distinctive AGN population not only because of their radio properties. They are also characterized by massive hosts, large BH masses and very low accretion rates (Heckman & Best 2014, and references therein). These traits seem to suggest that these are systems at the end of their life cycles of actively growing BH masses. The very low Eddington ratios in turn make accretion highly inefficient and the optically thick, geometrically thin disks usually invoked in most AGN would be replaced by an advection-dominated or radiatively inefficient accretion flow (ADAFs/RIAFs, Narayan & Yi 1994, 1995; Blandford & Begelman 1999). The absence of a classical accretion disk, and most likely of a classical BLR, would explain the spectroscopic classification of the RL sources and of the observed lack of variability. In this scenario, the lack of significant variability would not be due to obscuration but to the intrinsic nature of these extremely low accreting sources.

2.7 Discussion and Conclusions

In this Chapter we present a characterization of AGN variability in the near infrared regime, using data from the UltraVISTA survey (McCracken et al., 2012). UltraVISTA repeatedly imaged the COSMOS field in 5 bands (YJHKs and NB118), covering an area of 1.5 deg^2 , to achieve a very deep final image. The survey provides excellent quality, high spatial resolution data (with a mean seeing of $\sim 0.8''$) at different epochs, and has allowed us to analyze NIR variability within a time span of almost five years with good sampling. Besides, the depth of the images allows us to cover a wide redshift range, accessing the optical and NIR rest-frame emission. We used four public catalogs (Lusso et al., 2012; Muzzin et al., 2013; Marchesi et al., 2016a; Laigle et al., 2016) to construct a catalog of X-ray selected AGN, with information about their bolometric luminosities and spectroscopic, X-ray and radio properties. We used these catalogs to analyze the differences in the variability properties of sources classified according to their obscuration in the X-ray and optical range, and according to their radio properties.

Our variability analysis is based mostly on sources classified as variable. When we downsized the sample by selecting only those sources with well-sampled light curves (with more than 20 epochs and a length larger than 200 days in the rest-frame), we only missed a small fraction of variable sources ($\sim 10\%$). While providing a much more significant fraction of variable sources. As we showed in Table 1, the removed light curves are mostly non-variable.

According to the unified model, we might expect to detect a low fraction of NL variable sources in the optical range, since the obscuring material is expected to be a non-homogeneous distribution of moving clumps, thus, in some cases, optical emission might be detected for NL sources. It also predicts that as we move to the infrared regime, we should observe re-processed emission coming from the dusty torus, and therefore we should be able to detect damped variability for both BL and NL AGN (see Figure 1.3). The photometric bands used in our analysis (YJHKs) allowed us to access optical and near infrared rest-frame emission, depending on the redshift of the source and the band considered. These two predictions are verified by our analysis in section 2.6.3 and Figure 2.15. For the case of the Ks

band, for sources located at redshifts lower than 1.15, we observed emission coming from the NIR in the rest-frame, and therefore, we should expect to detect variability for obscured and unobscured sources, therefore increasing the fraction of variable sources. This is seen in our data.

Previous variability analysis have mostly been focused on unobscured sources, however there are some cases where optical variability has been reported for type II AGN. Choi et al. (2014) used SDSS data to select AGN candidates by variability. They found that contrary to the AGN unification model prediction, two of their six type II candidates showed a non-negligible amount of optical variability. Cartier et al. (2015) used data from the QUEST–La Silla AGN variability survey to study optical variability of BL and NL AGN. They show that 80% of the BL and 21% of the NL sources are classified as variable, and from their Structure Function analysis, they found that BL and NL AGN have different distributions on the SF parameter space. Simm et al. (2016) used a sample of variable X-ray selected AGN from the catalog of Brusa et al. (2010), to study optical variability. They reported that 96% of the sources were classified as type I, and they mention that 7 type II AGN in their sample were variable, which were not included in the analysis. Simm et al. (2016) also found that the amplitude of the variability anti-correlates with the bolometric luminosity.

In section 2.6.1 we showed that the variable NL sources have different distributions of the Structure Function A parameter compared to the variable BL sources. This result is in agreement with the results of Cartier et al. (2015). For the case of the XR I and XR II sources, the differences in the same parameter are not statistically significant. We think that the most likely explanation for the existence of variable NL sources in the optical rest-frame range is that most of them, which are also characterized by low AGN luminosities, correspond to BL AGN whose host galaxy is damping the variability signal, since most of them have a photometric classifications of Galaxy. We also proposed in section 2.6.3 that the four sources with variability properties similar to BL, with photometric classification of Type I, unobscured in X-rays and with high luminosities, correspond to “True type II” AGN or to BL AGN whose spectrum does not cover the region where the broad lines are present. For the case of the 4 variable sources classified as NL–TypeII sources in the Y band, we propose that they are examples of variable optically obscured sources.

The regression analysis of the Structure Function parameters with redshift and L_{BOL} (section 2.6.2) showed that A – related with the amplitude of the variability in a time scale of 1 year – has a positive correlation with redshift, and a weak anti-correlation with L_{BOL} . These results are in agreement with previous analysis (e.g., Simm et al. 2016; Caplar et al. 2017). On the other hand, the correlation between A and HL_{int} is not statistically significant. The correlation of the amplitude of the variability with redshift reveals an anti-correlation of the amplitude with the wavelength of emission. As we move to redder bands, the amplitude of the variability decrease (see Figure 2.11).

For the case of the parameter related with the structure of the variability, γ , the results of our analysis in section 2.6.2 did not show any correlation with redshift, L_{BOL} or HL_{int} . We also showed in section 2.6.1 that several sources have values of γ consistent with DRW processes, however there are a non-negligible number of variable sources with $\gamma > 1.0$ (36 for the Y band and 43 for the J band), which reveals deviations from a simple DRW process.

We also showed in section 2.6.3 that the fraction of variable sources unobscured in the X-ray is lower than the fraction of variable sources unobscured in the optical (in the spectroscopic and photometric classifications). We demonstrated that when we split the sources by their spectroscopic and X-ray classifications and by their photometric and X-ray classification (Figure 2.16), the differences in the fraction of variable sources are given by optical obscuration (i.e. spectroscopic and photometric classifications) and not by X-ray obscuration (X-ray classification). In other words, optical rest-frame variability is indifferent to X-ray obscuration. We think that an explanation is that optical and X-ray obscuration are caused by different mechanisms, and that X-ray obscuration might be due to dust-free material surrounding the inner part of the nuclei, as it was proposed by Merloni et al. (2014) and Marchesi et al. (2016b).

Chapter 3

The QUEST–La Silla AGN variability survey: connection between AGN variability and black hole physical properties

The material presented in this chapter is based on the work published in Sánchez-Sáez et al. (2018)

3.1 Introduction

Even though variability is one of the defining characteristics of AGN we do not completely understand the mechanisms that drive such variations. In particular it is not clear yet how physical properties of the central engine (e.g., luminosity, black hole mass, Eddington ratio, etc) are related to variability properties of the system (e.g., characteristic time-scale, variability amplitude, etc). If we can establish a firm statistical correlation between certain AGN variability features and some SMBH physical properties, we will be able to use the variability as a tool in the future to derive physical properties for huge samples of objects from dedicated synoptic surveys such as the LSST. Several efforts have been made in the past to assess this issue, some of them restricting the analysis to small numbers of well sampled light curves (e.g., Kelly et al. 2009, 2013; Simm et al. 2016; Smith et al. 2018), or studying large samples of sources through ensemble light curve analysis, assuming that sources with similar physical properties would have similar variability features (e.g., Wilhite et al. 2008; Bauer et al. 2009; MacLeod et al. 2010; Caplar et al. 2017). In order to test whether this assumption is correct, we need to perform an analysis of well sampled individual AGN light curves, with known physical properties. Hence, long and intensive campaigns are crucial.

An anti-correlation between the amplitude of the UV–optical variability and luminosity has been consistently observed by previous studies (e.g Angione & Smith 1972; Hook et al.

1994; Cristiani et al. 1997; Vanden Berk et al. 2004; Wilhite et al. 2008; Bauer et al. 2009; Kelly et al. 2009; MacLeod et al. 2010; Kelly et al. 2013; Simm et al. 2016; Caplar et al. 2017). However, the existence of correlation between the amplitude of the variability and the black hole mass or the Eddington ratio is not clear yet. Wold et al. (2007) used a sample of ~ 100 quasars from the Quasar Equatorial Survey Team, Phase 1 (QUEST1) variability survey (Rengstorf et al., 2004). They found a positive correlation between the black hole mass and the amplitude of the variability. Wilhite et al. (2008) found a positive correlation between the amplitude of the variability with black hole mass, and proposed that this could be explained by an anti-correlation with the Eddington ratio. MacLeod et al. (2010) also found a positive correlation with black hole mass, and propose that the anti-correlation between the amplitude of the variability and the Eddington ratio exists, but an additional dependence on luminosity or black hole mass is required. Kelly et al. (2009) found no evidence of correlation between the amplitude of the variability and the black hole mass or the Eddington ratio, and Kelly et al. (2013) found a scattered correlation between the amplitude and the black hole mass, and a weak anti-correlation with the Eddington ratio. Simm et al. (2016) found no correlation with the black hole mass, and an anti-correlation with Eddington ratio. More recently, Li et al. (2018) used a large sample of quasars ($\sim 10^5$) to perform an ensemble variability analysis. They found that the amplitude of the variability correlates positively with redshift, and negatively with bolometric luminosity, rest-frame wavelength and Eddington ratio. They also found that the correlation with black hole mass was uncertain. This uncertainty can be produced by the use of ensemble light curves and also by the large uncertainties that might be present in the black hole mass estimations used in their analysis (taken from Kozłowski 2017b), since they are calculated by using luminosities derived from broadband extinction-corrected magnitudes obtained from the Sloan Digital Sky Survey (SDSS; York et al. 2000), and by using the full width at half maximum (FWHM) of the lines obtained by Pâris et al. (2017b). It is clear that all these results on the correlation with black hole mass and Eddington ratio are inconsistent, most likely due to the shortcomings on the samples used, as highlighted before.

Rakshit & Stalin (2017) used a large sample of narrow-line Seyfert 1 (NLSy1) and broad-line Seyfert 1 (BLSy1) from the Catalina Real Time Transient Survey (CRTS; Drake et al. 2009). The light curves used in their analysis have a minimum of 50 epochs of data spanning 5 to nine years, thus they could perform a variability analysis for individual light curves. They found a strong anti-correlation between the amplitude of variability and the Eddington ratio, and they proposed that the accretion disk is the main driver of the variability observed in both broad and narrow line Seyfert 1 galaxies. However, since Rakshit & Stalin (2017) used Damp Random Walk (DRW) modelling to measure the variability amplitude, which has several limitations for the analysis of ground-based light curves, since they tend to have gaps and time coverages of a few months or years (see section 2.5.2), their results must be confirmed using a different method (e.g., the structure function).

In this Chapter we present our statistical analysis of the connection between AGN variability and physical properties of SMBH. For the variability analysis we used light curves from the QUEST-La Silla survey (see Section 1.3.1), and derived physical properties from spectra taken from SDSS. We perform the spectral fitting using the procedure of Mejía-Restrepo et al. (2016) (MR16 hereafter), from which we could derive physical parameters and also line fitting properties such as the FWHM of the emission lines and continuum luminosities. For

the variability analysis, we used the same approach as in Chapter 2. In this work, we used single object light curves, in order to test the claim that sources with similar physical properties have similar variability behaviors (like proposed by Vanden Berk et al. 2004; Wilhite et al. 2008; MacLeod et al. 2010; Caplar et al. 2017, among others).

The Chapter is organized as follows. In section 3.2 we describe the optical imaging and spectroscopic data used for the analysis. In section 3.3 we describe the different variability features used, and we report the results of the variability analysis for our sample. In section 3.4 we explain the procedure followed to obtain the physical properties from the SDSS spectra, and show the distribution of these parameters for our sample. In section 3.5 we define the different sub-samples used in our analysis. In section 3.6 we show the results of our statistical analysis done to connect the variability and physical properties. In section 3.7 we analyse the differences in the variability parameters of sources classified as Broad Line QSO and normal sources, and sources classified as radio-loud and radio-quiet. Finally, in section 3.8 we discuss the physical implications of our findings and summarize the main results. The photometry reported here is in the AB system. We adopt the cosmological parameters $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

3.2 Data

3.2.1 Optical light curves

We constructed light curves for all the sources from the photometric SDSS (Gunn et al., 1998; Doi et al., 2010) and DES (Abbott et al., 2018) catalogs with detections in the QUEST-La Silla data, using the same methodology as in Cartier et al. (2015). We decided to bin our light curves every three days, in order to reduce the noise in our light curves, produced by changes in atmospheric conditions, the relatively low quality of the QUEST camera, among other factors. We generated a total of ~ 300000 binned light curves.

3.2.2 SDSS spectra

Three of our fields (COSMOS, Stripe 82 and XMM-LSS) have spectroscopic information from the SDSS survey. We used the SDSS Data Release 14 Quasar catalog (DR14Q) (Pâris et al., 2017a), in order to identify sources with a detection in QUEST-La Silla already classified as quasars. We found 2345 sources with both QUEST-La Silla light curves and SDSS spectra, classified as quasars in DR14Q, this sample inherits the selection criteria of SDSS spectroscopic survey. We downloaded the calibrated SDSS spectra from the SDSS Catalog Archive Server, and then corrected the spectra by Galactic extinction using the maps of Schlegel et al. (1998) and the model of Cardelli et al. (1989). The wavelength coverage of the SDSS spectra ranges from 3800 to 9200 Å for the SDSS survey and from 3650 to 10400 Å for the BOSS survey (Dawson et al., 2013), with a spectral resolution of 1500 at 3800 Å, and 2500 at 9000 Å.

From the SDSS quasar catalogs we obtained the spectroscopic redshift (z) for every source. We then used these redshifts to transform every light curve to the AGN rest-frame: $t_{\text{rest}} = t_{\text{obs}}/(1+z)$, where t_{rest} is the light curve time at the rest-frame in days and t_{obs} is the observed time. The following analysis has been done considering the light curves in rest-frame time.

3.3 Variability analysis

3.3.1 Variability features

To characterize the variability of our sources, we used the same approach as in Chapter 2. We used the P_{var} and the excess variance (σ_{rms}) parameters, and the Structure Function and the DRW statistics. For further details see Sections 1.2.2 and 2.5.

Following the same approach of Chapter 2, we classify a source as variable if its light curve satisfies $P_{\text{var}} \geq 0.95$ and $(\sigma_{\text{rms}}^2 - \text{err}(\sigma_{\text{rms}}^2)) > 0$.

Kelly et al. (2009) proposed that a DRW process can be a good descriptor for AGN light curves. However, in Section 2.5.2, we demonstrated the limitations of the use of DRW processes for short light curves (see also Kozłowski 2017a). For light curves with $t_{\text{rest}} < 10 \times \tau$, the correct value of τ cannot be determined. Since our light curves have an observed time coverage of $t_{\text{obs}} \leq 5$ years, while the characteristic time-scale is expected to be of the order of hundred of days, we decided to excluded τ from our analysis.

For the case of the SF, during this Chapter we adopt the Bayesian definition of Schmidt et al. (2010) (see Section 1.2.2), since, as we demonstrated in Section 2.5.1, it provides the best results for irregular and noisy light curves.

3.3.2 Sample filtering by light curves properties

Since we want to study individual light curves, we have to consider only those sources with sampling dense enough to get statistically significant variability features. Figure 3.1 shows the distribution of the number of epochs ($\#\text{epochs}$) and rest-frame time length of our 2345 light curves with SDSS spectra. Following a similar approach than in Chapter 2, we selected for our analysis those light curves with $t_{\text{rest}} \geq 200$ days, and in order to ensure a high number of epochs, we also selected those light curves with $\#\text{epochs} \geq 40$. In the figure we can see that most of our sources satisfy these conditions. After we filter our sample by the number of epochs and the length of the light curve, we ended with 1751 sources. Hereafter, we refer to it as the ‘well-sampled’ sub-sample.

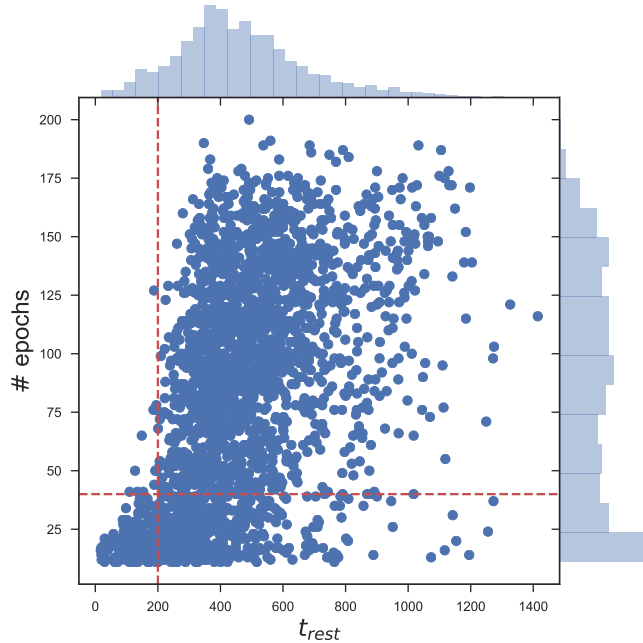


Figure 3.1: Number of epochs vs. rest-frame time length of the 2345 light curves with SDSS spectra. The red vertical dashed line shows the position where $t_{rest} \equiv 200$ days, and the red horizontal dashed line shows the position where $\#epochs \equiv 40$.

3.3.3 Biases of the variability features

Figure 3.2 shows the different measured variability features versus the light curve properties (i.e. the number of epochs or the length in days). We can see that the SF parameters γ and A are practically unaffected by the light curve length and the number of epochs. The excess variance σ_{rms} is also unaffected by the light curve sampling. On the other hand, σ_{DRW} is affected by the length of the light curve and strongly affected by the number of epochs. This is quantified by the Spearman’s rank coefficient which gives values of -0.59 ($p_{val} < 1e-8$), -0.3 ($p_{val} < 1e-8$) and -0.41 ($p_{val} < 1e-8$) for the correlations of σ_{DRW} and $\#epochs$, t_{rest} , and t_{obs} , respectively. We decided to use A and γ as the main features for our analysis, and use σ_{DRW} and σ_{rms} as references.

3.3.4 Variability of simulated light curves

In this work, we used the parameters of the structure function as the main variability features, therefore it is important to understand how these parameters respond to different factors.

In order to understand how the sampling of a given light curve affects the measurement of its SF parameters, we simulated artificial light curves, following a similar approach than in Chapter 2 (see Sections 2.5.1 and 2.5.2). We simulated light curves using the same sampling of the light curves shown in Figure 3.3. The light curve at the top of the figure (short light curve), has a length of 492 days and 39 observing epochs. The light curve at the bottom

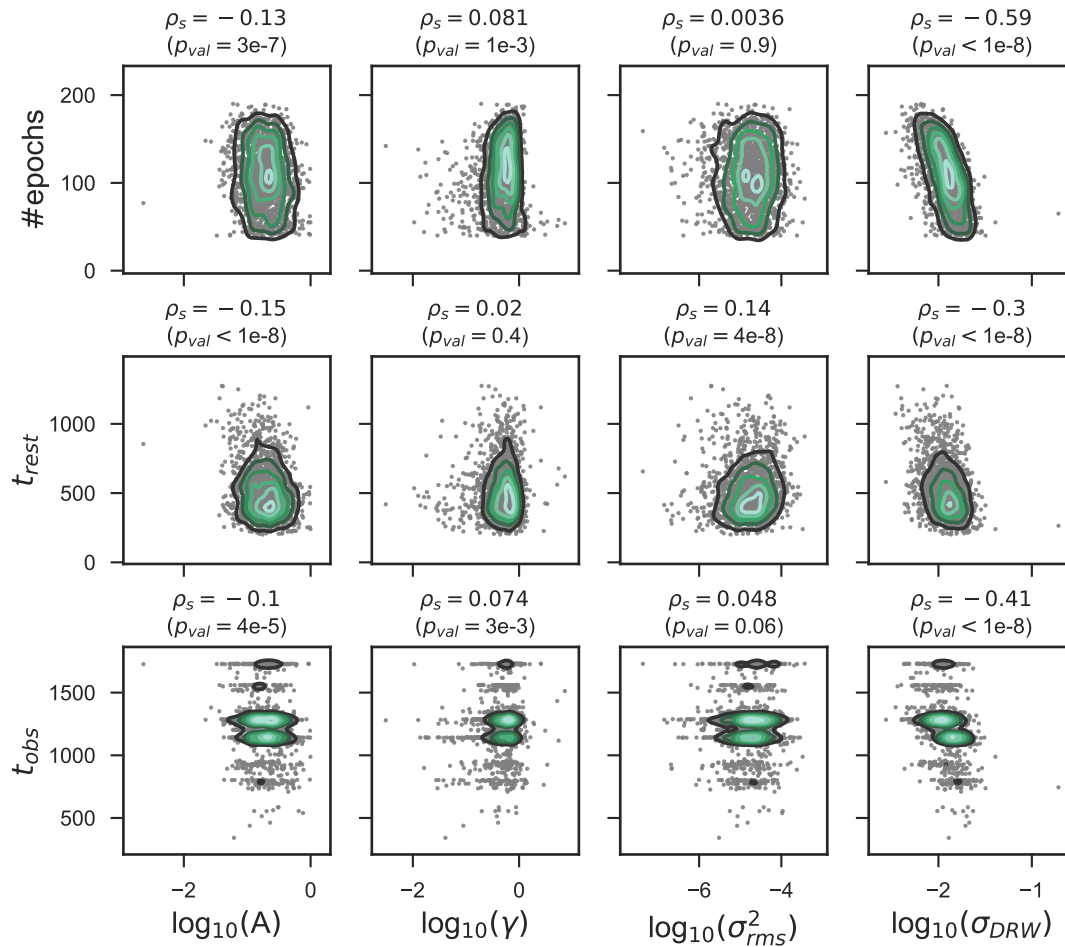


Figure 3.2: Variability features vs. light curve properties. The Spearman's rank correlation coefficient is shown as reference for every pair of variables.

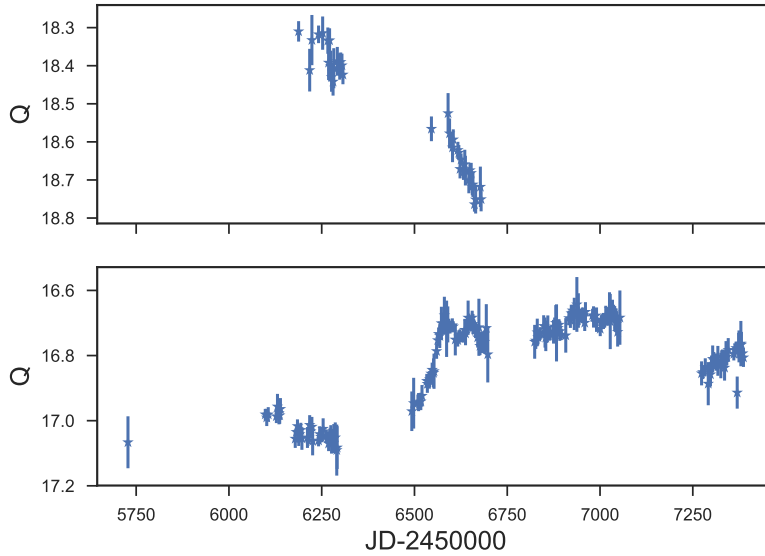


Figure 3.3: Light curves with different number of epochs and length, but similar cadence.

(long light curve) has 175 observing epochs and a length of 1659 days.

We simulated light curves from a DRW process with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$ mag. For each long and short light curves, we simulated 1000 light curves. Figure 3.4 shows the results of the SF parameters measured for the short and long simulated light curves. First of all, we can see that the distributions of the parameters measured for the short light curves have a larger dispersion than the values measured for the long light curves. For the case of the short light curves, the median, mean and standard deviation of A are 0.19, 0.22 and 0.12 respectively, and for γ are 0.3, 0.32 and 0.21 respectively. For the case of the long light curves, the median, mean and standard deviation of A are 0.22, 0.23 and 0.07 respectively, and for γ are 0.37, 0.37 and 0.14 respectively. From Figure 3.4, we can also see that there is a correlation between the measured parameters. For the case of the short light curves the Spearman’s rank coefficient is $\rho_s = 0.91$ ($p_{\text{val}} < 1e-8$), and for the long light curves we have $\rho_s = 0.87$ ($p_{\text{val}} < 1e-8$).

We tested whether other definitions of the SF show the same behavior, using the definition of Kozłowski (2016) and the “traditional” definition of Schmidt et al. (2010). The results were consistent with what was found for the Bayesian method of Schmidt et al. (2010). In addition, we simulated artificial light curves with a power-law PSD (assuming different values for the exponent), following the same approach of Chapter 2, and we found similar results.

We also tested whether longer light curves can solve the degeneracy between the SF parameters. We simulated 1000 light curves from a DRW process with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$ mag, with 7000 days and 700 epochs, and with a similar cadence than the long light curve of figure 3.3. The Spearman’s rank coefficient of A versus γ is $\rho_s = 0.88$ ($p_{\text{val}} < 1e-8$), therefore, the parameters are still correlated. The median, mean and standard deviation of A are 0.24, 0.24 and 0.04 respectively, and for γ are 0.39, 0.39 and 0.07 respectively. In order to see whether the previous results are produced by the gaps in the data, we simulated

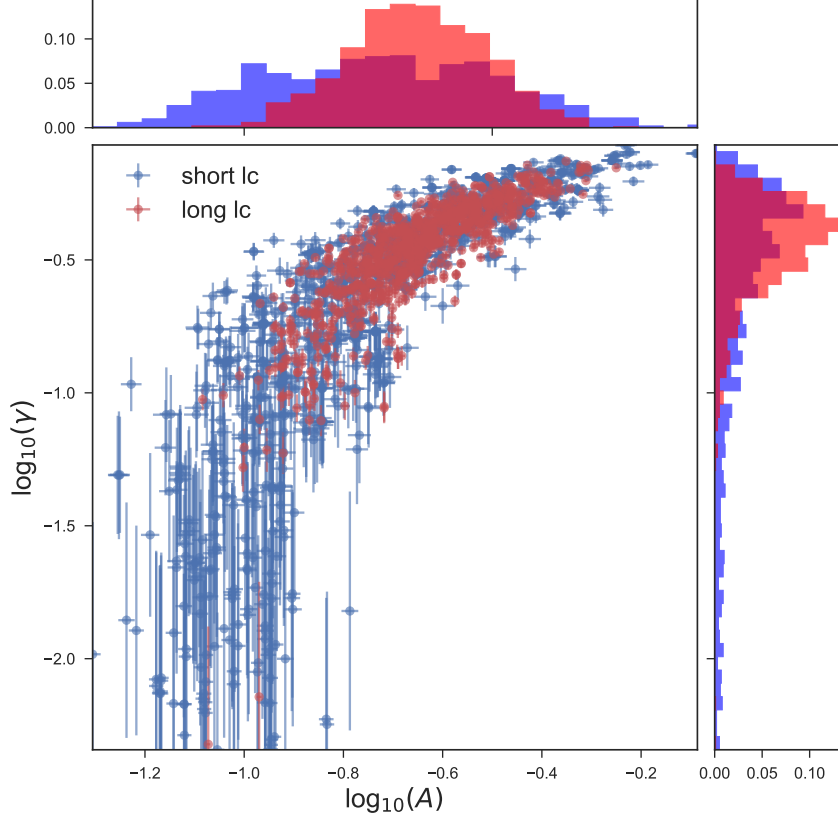


Figure 3.4: SF parameters measured for 1000 light curves simulated from a DRW process with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$ mag, with short and long samplings. Along the axes we show the histograms of every parameter.

1000 light curves from a DRW process with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$ mag, with 7000 days of length and with observations every 10 days, obtaining similar results for the correlation between A and γ .

We simulated white noise light curves, with a standard deviation of 0.2 mag, in order to test whether the correlation between the SF parameters is also present for light curves with a constant PSD. The Spearman's rank coefficient for the short light curves is $\rho_s = 0.23$ ($p_{\text{val}} < 1e-8$), and for the long light curves is $\rho_s = 0.23$ ($p_{\text{val}} < 1e-8$). In this case, we also see a broader distribution of the parameters for the short light curves. For the case of the short light curves, the median, mean and standard deviation of A are 0.29, 0.29 and 0.04 respectively, and for γ are 0.02, 0.02 and 0.02 respectively. For the case of the long light curves, the median, mean and standard deviation of A are 0.28, 0.28 and 0.02 respectively, and for γ are 0.006, 0.007 and 0.005 respectively. We also simulated white noise light curves, with a standard deviation of 0.02 mag and we found similar results.

In order to test whether the distribution of A and γ values measured for the real light curves is simply produced by this degeneracy and scatter, we compare the regions in the $A - \gamma$ plane covered by the real and simulated light curves. We simulated DRW light curves with $\tau = 300$ days and with different values of SF_{inf} . For every amplitude we simulated 1000

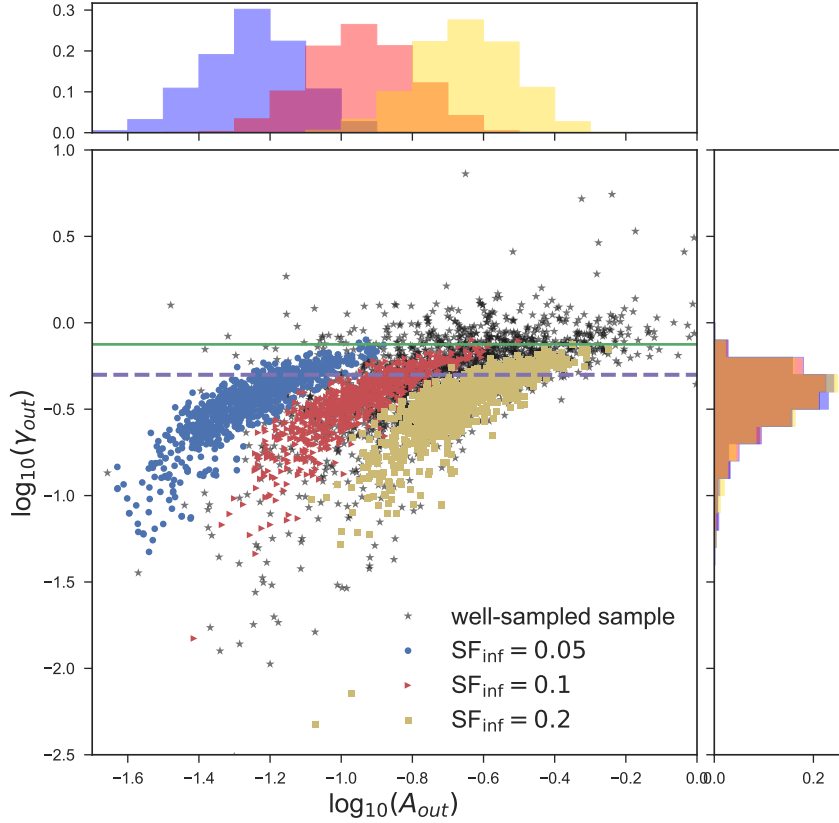


Figure 3.5: SF parameters measured for light curves simulated from a DRW process with $\tau = 300$ days and different values of SF_{inf} (with the long sampling). The blue circles correspond to light curves simulated with $SF_{\text{inf}} = 0.05$, the red triangles correspond to light curves simulated with $SF_{\text{inf}} = 0.1$, and the yellow squares correspond to light curves simulated with $SF_{\text{inf}} = 0.2$. Along the axes we show the histograms of every parameter for the different SF_{inf} . The green solid line shows the position where $\gamma = 0.75$, and the magenta dashed line shows the position where $\gamma = 0.5$, the expected value for a DRW process. We show with black stars the measurements done for the variable sources of the well-sampled sub-sample.

light curves, with the same sampling of the long light curve in Figure 3.3. Figure 3.5 shows the measured values of A and γ for three different values of SF_{inf} (0.05, 0.1, and 0.2). In the figure we can see that there is no change in the distribution of γ for different values of SF_{inf} . We can also see that independently of the value of SF_{inf} , the measured values of γ range from 0.0 to 0.75. Therefore, if we measure a value of γ between this range for a given light curve, we cannot discard a DRW process as the best model to describe the variability. However, the distribution of gamma values is significantly different for the simulated and real light curves, so as a population the AGN light curves are not well represented by a DRW model, at least because it incorporates many outliers

As can be seen in Figure 3.5, no single value of the intrinsic amplitude SF_{inf} reproduces the entire parameter space, from where we conclude that the measured values of the amplitude correlate on average with the intrinsic amplitude.

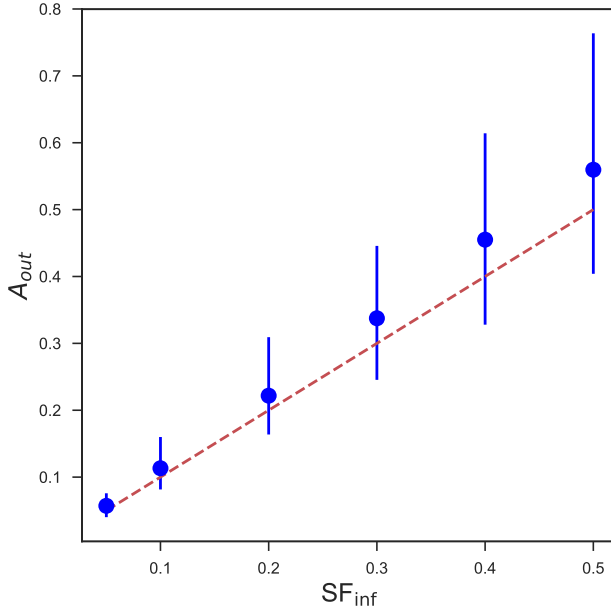


Figure 3.6: A measured for light curves simulated from a DRW process with $\tau = 300$ days and different values of SF_{inf} (with the long sampling). The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value of A (1:1 relation).

Figure 3.6 shows the distribution of measured values of A for different input SF_{inf} . We can see that we can recover the input amplitude value, but with larger dispersions for larger values of SF_{inf} . Therefore, we can say that the amplitude measured with the structure function is a good estimator of the intrinsic variability amplitude, albeit with significant scatter due to the $A - \gamma$ degeneracy.

We also tested the effects of the source redshift in the measured parameters, since sources at higher redshifts have shorter rest-frame light curves for a given observed light curve ($t_{\text{rest}} = t_{\text{obs}}/(1 + z_{\text{spec}})$). We simulated DRW light curves with $\tau = 300$ days and $SF_{\text{inf}} = 0.2$. In order to account for the redshift of the source, we use the sampling of the long light curve of Figure 3.3, dividing the time by $(1 + z_{\text{spec}})$, for different values of z_{spec} . We simulated 1000 light curves per every value of z_{spec} . Figures 3.7 and 3.8 show the results for A and γ respectively. We can see that both values do not change with z_{spec} , but the dispersion of the measured values increase a little bit with redshift. We can also see that the measured value of A is close to the input value (red dashed line in the figure), but the measured value of γ is below the expected value. This is consistent with the findings of Chapter 2 (see Figure 2.7).

From all these results, we can conclude that the $A - \gamma$ degeneracy is much lower for the case of light curves with a constant PSD. For the case of stochastic light curves, the broad distribution of the measured parameters is produced by the fact that light curves with a few months or years of coverage are not a well representation of the general behavior of variability with decorrelation time-scales of months or years, or with power-law PSD, and by the $A - \gamma$ degeneracy. The correlation between the measured SF parameters is present independently

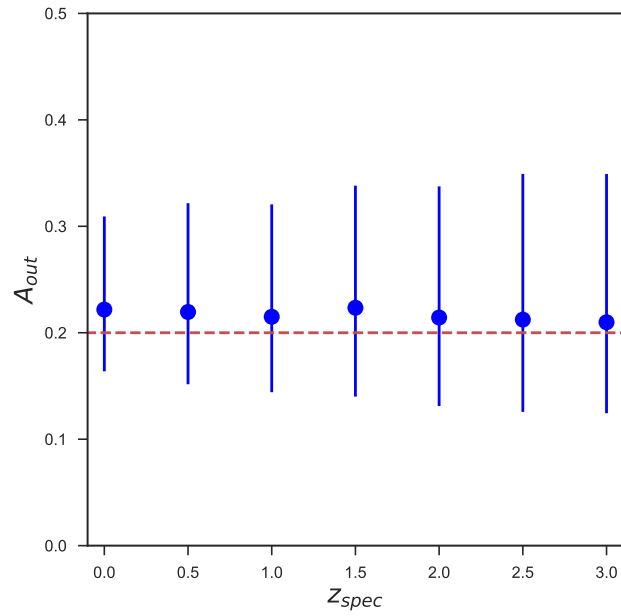


Figure 3.7: A measured for light curves simulated from a DRW process with $\tau = 300$ days and $SF_{inf} = 0.2$ mag. The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value $A = 0.2$.

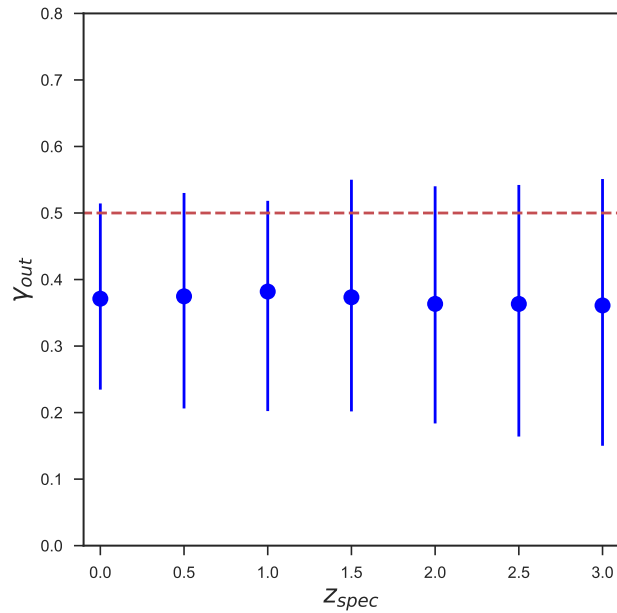


Figure 3.8: γ measured for light curves simulated from a DRW process with $\tau = 300$ days and $SF_{inf} = 0.2$ mag. The circles correspond to the median value measured, and the error bars correspond to the 15.9 and 84.1 percentiles. The red dashed line shows the expected value $\gamma = 0.5$.

on the SF definition used. We can reduce the effects of this degeneracy by using light curves with several years of length. Besides, we can conclude that the values of the amplitude of the variability obtained from the SF, are a good estimation of the real amplitude, independently of the redshift of the source. On the other hand, for the case of γ , we must consider that if we measure a value between 0.0 and 0.75 we cannot discard a DRW process as the best model to describe the variability.

3.3.5 Variability of the QUEST–La Silla light curves

Following the criteria that sources with $P_{var} \geq 0.95$ and $(\sigma_{rms}^2 - err(\sigma_{rms}^2)) > 0$ are classified as variable, 1579 of the well-sampled light curves are variable, which corresponds to the 90.2% of the sample. Figure 3.9 shows the distribution of the SF parameters, for the variable sources of the well-sampled sub-sample. The weighted average of the parameters are $\bar{A} = 0.21 \pm 0.12$ and $\bar{\gamma} = 0.64 \pm 0.22$. The median, and the percentiles 15.9 and 84.1 of the measured values of A are 0.19, 0.11, and 0.34 respectively. The median, and the percentiles 15.9 and 84.1 of γ are 0.53, 0.39, and 0.80 respectively. The measured values of A are consistent with previous findings for optical variability (e.g., Schmidt et al. 2010; MacLeod et al. 2010; Cartier et al. 2015; Suberlak et al. 2017; Rakshit & Stalin 2017). From Figure 3.5 we can see that our measurements are consistent with amplitudes ranging between 0.05 and 0.2, however there are some sources with larger amplitudes.

For a DRW process, the expected value of γ is 0.5, however, in section 3.3.4, we showed that the Bayesian method tend to underestimate the value of γ (see Figures 3.5 and 3.8). From Figure 3.5, we can see that the distribution of γ measured for the well-sampled sub-sample (black stars) differs from the distribution of γ obtained for light curves simulated from a DRW process with different amplitudes. Therefore we cannot say that a DRW process can explain the general behavior of the variability in our sample.

Figure 3.10 shows the correlations of the variable features (2D distributions) and their individual distributions (1D distributions) for all the sources classified as variable in the well-sampled sub-sample. Our results for σ_{DRW} are consistent with Kelly et al. (2009). Our measurement of σ_{rms}^2 can be characterized by a mean and standard deviation of $(2.8 \pm 3.3) \times 10^{-5}$, which is smaller than the values reported by previous analyses (e.g., Cartier et al. 2015; Simm et al. 2016). When we compare the results for σ_{rms}^2 , we must consider that it depends on the length of the light curve, since this parameter consider the total variance of the light curve, and also it depends on the photometric errors, since with low photometric errors we can detect lower variability amplitudes. It is also important to consider the differences in the definition of the excess variance used by different authors. For example, Cartier et al. (2015) used a non-normalized version of the excess variance for optical light curves measured in magnitudes and with a shorter coverage in time compared to our light curves, but with the same photometric errors, since they used QUEST–La Silla data. If we correct their measurements by the mean magnitude, we obtain consistent results. Simm et al. (2016) used optical light curves measured in flux and with different coverage in time, which introduces differences with our results.

We can also see from Figure 3.10 that the strongest correlation is shown between A and

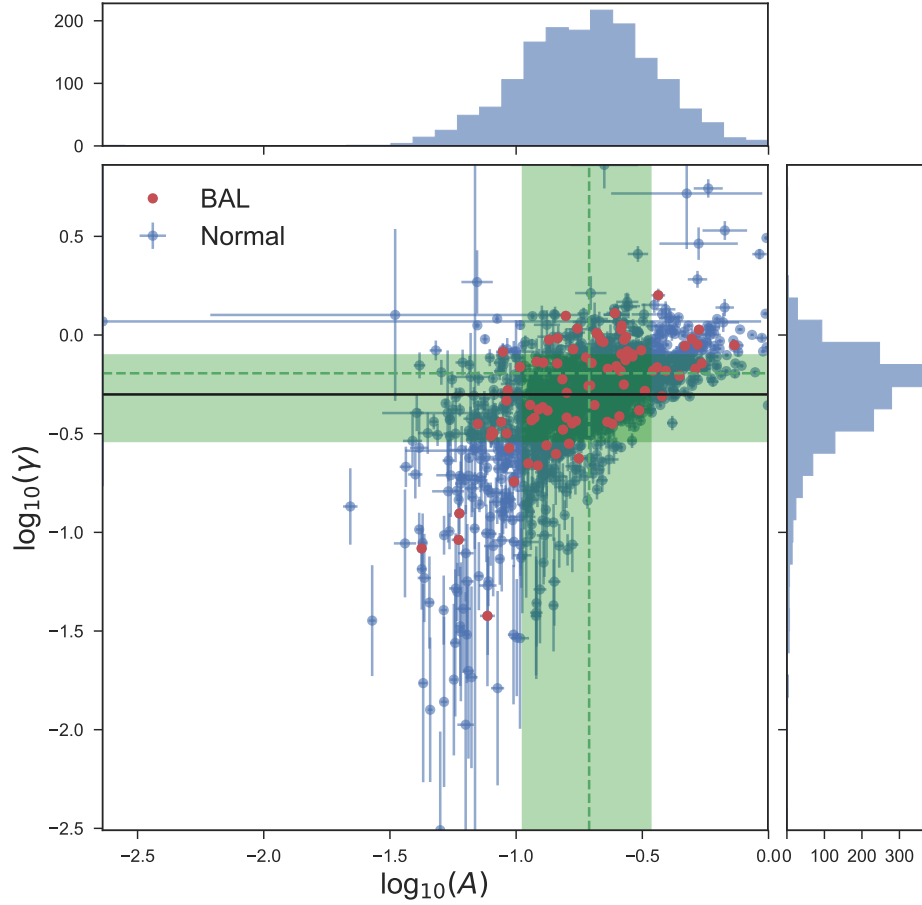


Figure 3.9: Distribution of the SF parameters A and γ for the variable and well sampled sources. Along the axes we show the histograms of every parameter. We plot in red the sources classified as BAL-QSO (sources with broad absorption-lines in their spectra), and in blue rest of the sample. The black solid line shows the expected value of γ for a DRW process. The green dashed lines show the median of the parameters. The green shaded regions show the 15.9 to 84.1 percentile range.

σ_{rms}^2 . This can be produced by the sampling of the light curve, since several light curves has lengths in rest-frame close to 1 year (see Figure 3.1).

From Figures 3.9 and 3.10, we can see that the SF parameters are correlated. The Spearman’s rank correlation coefficient for A versus γ is $\rho_s = 0.62$ ($p_{val} < 1e-8$). In section 3.3.4 we showed that this correlation is a product of the degeneracy in the SF parameters, produced by the stochastic nature of the light curves, the sampling of the light curves, and the structure function method by itself. Therefore it does not have any physical implication.

3.4 Spectral analysis

3.4.1 Host galaxy subtraction

For low redshift sources ($z_{spec} \leq 0.8$) we can have a significant degree of host galaxy contamination in the optical SDSS spectra, depending on the brightness of the nucleus. We follow the simple procedure to subtract the galactic continuum of Greene & Ho (2005) and Kim et al. (2006), where the stellar continuum is modelled using the scaled spectrum of a K giant star. In order to know how much starlight must be subtracted from the spectra, we use the equivalent width (EW) of the Ca II K absorption line ($\lambda = 3934$).

We isolated the AGN component only for those objects with $z_{spec} \leq 0.8$ (349 sources), since for sources with $z_{spec} > 0.8$ we can ensure the presence of the Mg II line in the SDSS spectra, and the H_β line would be located in the edges of the spectra. For 304 of these 349 sources, the quality of the spectra was good enough to obtain the AGN and host galaxy components (i.e. with $S/N(\text{Ca II K}) > 10$).

3.4.2 Spectral fitting and measurement of physical properties

We used the procedure proposed by MR16 to estimate the black hole masses (M_{BH}), the luminosity at 5100Å (L_{5100}), the accretion rate (\dot{M}), and the Eddington ratio (L/L_{Edd}) for our AGN. MR16 proposed new calibrations for the measurement of these physical properties from single-epoch spectrum, by fitting the H_α , H_β , Mg II $\lambda 2798$ and C IV $\lambda 1549$ lines. Their method relies on the assumption of virialized BLR kinematics and consider the FWHM of the line as a proxy to the virial velocity of the gas in the BLR (V_{BLR}). Additionally, the continuum luminosity in the proximity of the emission line is used to estimate the BLR radius by means of the empirical luminosity–radius relationship derived from several reverberation mapping experiments (e.g., Bentz et al. 2013). Their method model the broad emission lines with two broad Gaussian components, and for the case of doublet lines (Mg II and C IV) they use two additional Gaussians, which are separated by the theoretical wavelength doublet separation, and are forced to have the same profiles and intensity of the other two Gaussians, which is valid for optically thick BLR clouds. For the case of H_α and H_β their method also includes a third narrow line component, modelled with a single Gaussian, which account for the narrow line emitting region. To model the continuum emission of the AGN, they followed the local

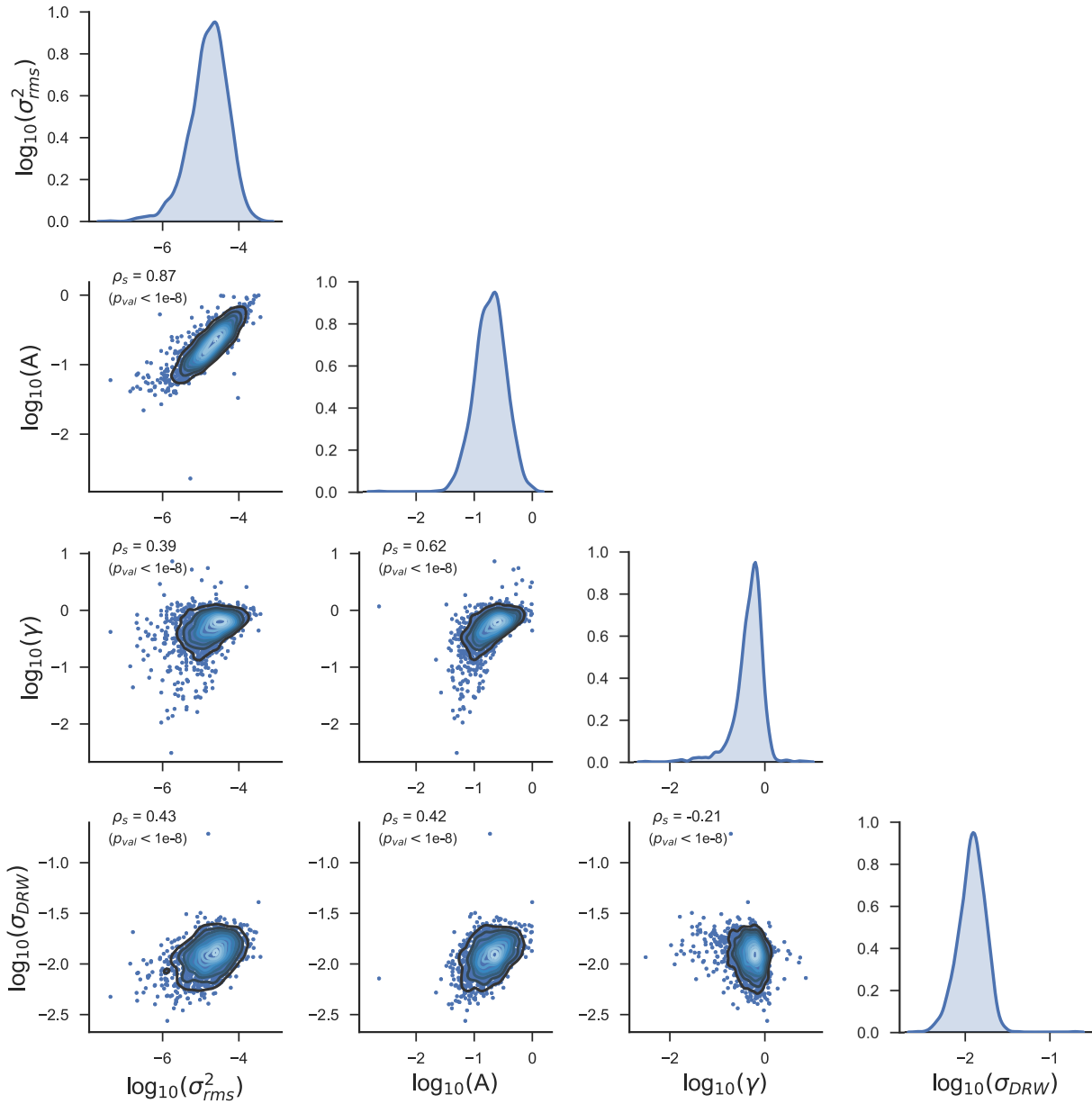


Figure 3.10: Correlations of the variable features for all the variable and well sampled sources. The diagonal shows the individual distributions. As a reference we provide the Spearman's rank correlation coefficient for every pair of variables.

approach described in MR16 which consists of fitting a single power-law limited within two narrow pseudo-continuum windows around the emission line. For the case of the H_β and Mg II lines, their method also includes the modelling of an iron pseudo-continuum, using an iron template, which originates from a large number of blended features of Fe II and Fe III (for further details see MR16).

The black hole mass is calculated as $M_{\text{BH}} = fG^{-1}R_{\text{BLR}}V_{\text{BLR}}^2 = K(\lambda L_\lambda)^\alpha \text{FWHM}^2$. The values of K and α for every line can be found in Table 7 of MR16. The accretion rate is estimated from equation 1 of Netzer & Trakhtenbrot (2014), as follows

$$4\pi D_L^2 F_\nu = f(\theta)[M_8 \dot{M}]^{2/3} \left[\frac{\lambda}{5100} \right]^{-1/3} \text{erg s}^{-1} \text{Hz}^{-1},$$

where M_8 corresponds to M_{BH} measured in units of $10^8 M_\odot$, \dot{M} is the accretion rate in units of M_\odot/year , D_L is the luminosity distance, and $f(\theta)$ is the inclination-dependent term, that describes the orientation of the accretion disk to the line of sight.

We used the calibrations derived by MR16 to estimate L_{5100} from L_{6200} , L_{3000} and L_{1450} (see their Table 5). From this, we estimated Eddington ratio as

$$L/L_{\text{Edd}} = \frac{C_{\text{BOL}} L_{5100}}{1.5 \times 10^{38} (M_{\text{BH}}/M_\odot)},$$

where C_{BOL} is the bolometric correction. In this analysis we adopted $C_{\text{BOL}} = 9.26$ (see Shen et al. 2008; MacLeod et al. 2010 and references therein).

Mejía-Restrepo et al. (2018a) proposed new corrections for the estimation of M_{BH} , which intend to account for the effect of the unknown distribution of the gas clouds in the BLR. They suggest that the virial correcting factor is inversely proportional to the width of the broad emission line used to compute M_{BH} . This can be explained either by line of sight inclination effect on a planar BLR or by radiation pressure effects on the BLR gas distribution (Kollatschny & Zetzl, 2013). The corrected black hole mass is calculated as $M_{\text{BH}}^C = f M_{\text{BH}}$, with $f = (\text{FWHM}_{\text{obs}}(\text{line})/\text{FWHM}_{\text{obs}}^0)^\beta$, where $\text{FWHM}_{\text{obs}}(\text{line})$ is the FWHM of the emission line. $\text{FWHM}_{\text{obs}}^0$ and β are parameters calculated per every line (see Table 1 in Mejía-Restrepo et al. 2018a). We computed M_{BH}^C for our sample, and used it to calculate the corrected Eddington ratio $(L/L_{\text{Edd}})^C$. In the next sections we will compare our results when using both the original and the corrected black hole masses.

3.4.3 Spectral properties of the selected SDSS spectra

We have a total of 2345 sources with SDSS spectra, however not all of them have a signal to noise (S/N) high enough to allow the fitting of the emission lines. In our analysis we only consider those spectra with a mean S/N per pixel, in the continuum region around the emission line of interest, larger or equal to 10. We also exclude from our analysis those sources classified as Broad Line Absorption QSO (BAL-QSO) in the catalogs of Shen et al. (2011), Pâris et al. (2017b) and Pâris et al. (2017a), with strong absorption lines in the region of the emission line under analysis. After we eliminated sources from our sample with low S/N

and classified as BAL–QSO, we end with 102 sources having H_α inside the SDSS wavelength coverage, 304 sources with H_β , 1561 sources with Mg II, and 801 sources with C IV. Most of our sources have more than one line available.

We fit the H_α and H_β lines for sources with $z_{\text{spec}} \leq 0.8$). For 81 sources we could obtain a fit of H_α and for sources 224 we obtained a fit of H_β . Besides, we obtained a fit of Mg II for 1487 objects, and a fit of C IV for 718.

After the line fitting, we calculated M_{BH} , L_{5100} , \dot{M} , and L/L_{Edd} using the equations given in section 3.4.2. We considered only those line fits where the height to noise (H/N) of the line is $H/N \geq 5$, with the height defined as the distance between the peak of the line fitted and the continuum. This is done to avoid the fit of fake lines when the broad lines are weak or are not present. Since for some objects we have more than one line available, we decided to estimate the final M_{BH} as the weighted average of the measured M_{BH} for the different lines, with the exception of the C IV line. It is well known that the C IV line width is not a good estimator of V_{BLR} , and therefore the measurements done using this line must be taken carefully (see MR16 and references therein). Therefore, whenever a source has C IV and other lines available, we excluded C IV from the estimation of M_{BH} , and we only consider the results of C IV when there is no other line available. In the next sections, the analyses are done with and without the results of the C IV fitting.

Following the previous procedure, we computed M_{BH} for 1899 sources, and L_{5100} for 1951 sources. Figure 3.11 shows the correlations and individual distributions of M_{BH} , L_{5100} , \dot{M} , and L/L_{Edd} , for all the sources with both M_{BH} and L_{5100} available (1899 sources). We measure the M_{BH} in units of solar masses [M_\odot], L_{5100} in units of [erg s^{-1}], \dot{M} in units of [M_\odot/year], while L/L_{Edd} is dimensionless. Our range covered for M_{BH} , L_{5100} , and L/L_{Edd} is similar than in previous variability analysis (e.g Wilhite et al. 2008; Kelly et al. 2009; MacLeod et al. 2010; Simm et al. 2016; Caplar et al. 2017).

The strongest correlations in Figure 3.11 are: a) \dot{M} and L/L_{Edd} , which is explained by the not particularly broad distribution of M_{BH} ; b) L_{5100} and both M_{BH} and \dot{M} , which is related with the use of L_{5100} in the determination of both quantities; and c) z_{spec} with L_{5100} , M_{BH} , and M_\odot , which are mostly caused by a selection effect coming from the flux limited nature of the observations.

Figure 3.12 shows the comparison of the M_{BH} and L/L_{Edd} measured using the standard single–epoch method, versus the measurements obtained using the new method proposed by Mejía-Restrepo et al. (2018a).

3.5 Final sample definition

Our final sample is composed by all those sources for which we could measure M_{BH} , L_{5100} , \dot{M} , and L/L_{Edd} , and have light curves with $t_{\text{rest}} \geq 200$ days and $\#\text{epochs} \geq 40$. We have 1473 sources in our original sample that satisfy all these conditions. We call this sub–sample the “QUEST–SDSS sample”. 1348 of these sources are variable (91.5%).

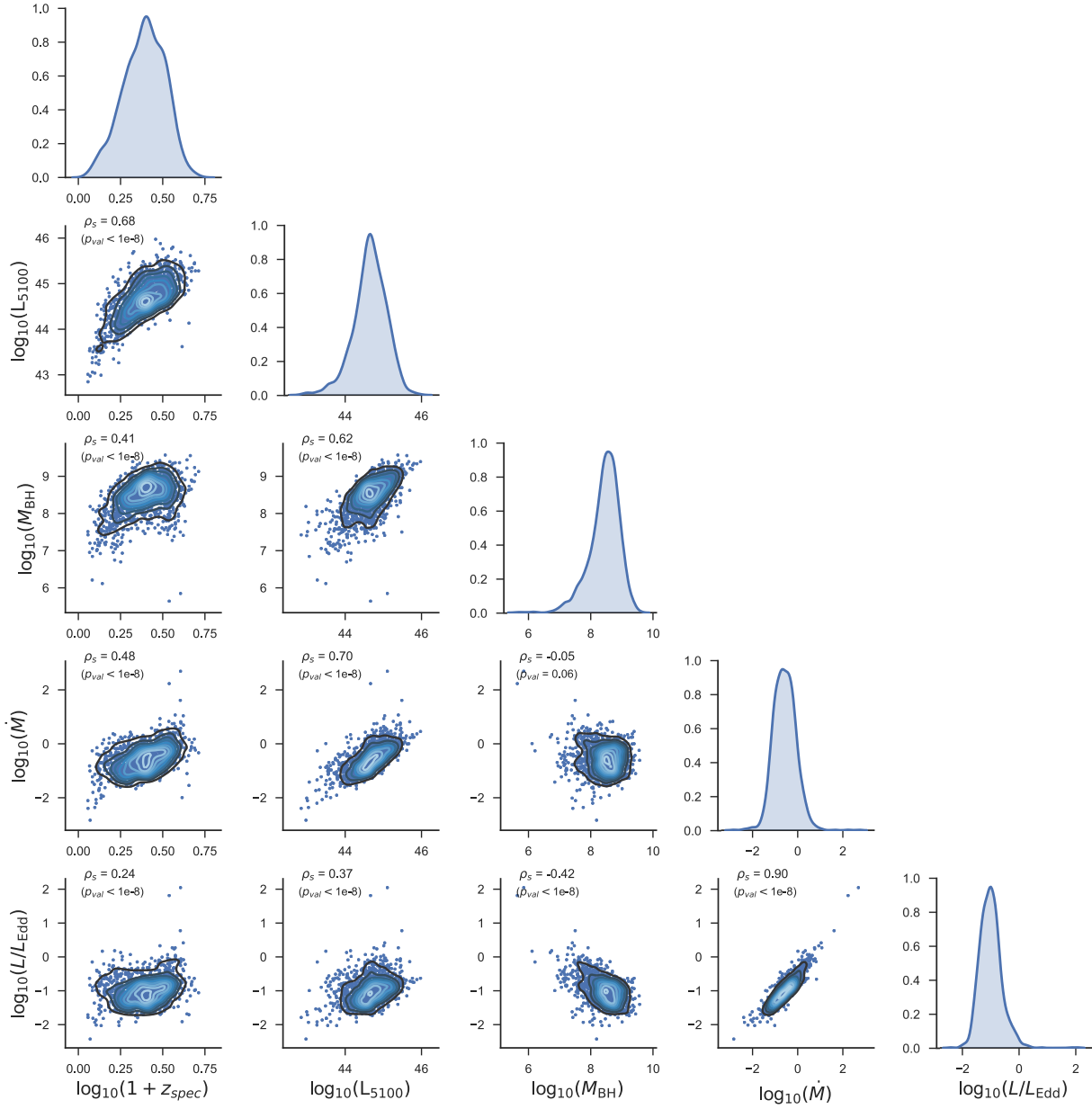


Figure 3.11: Correlations of the spectroscopic parameters for all the sources with M_{BH} and L_{5100} available. The diagonal shows the individual distributions. As a reference we provide the Spearman's rank correlation coefficient for every pair of variables.

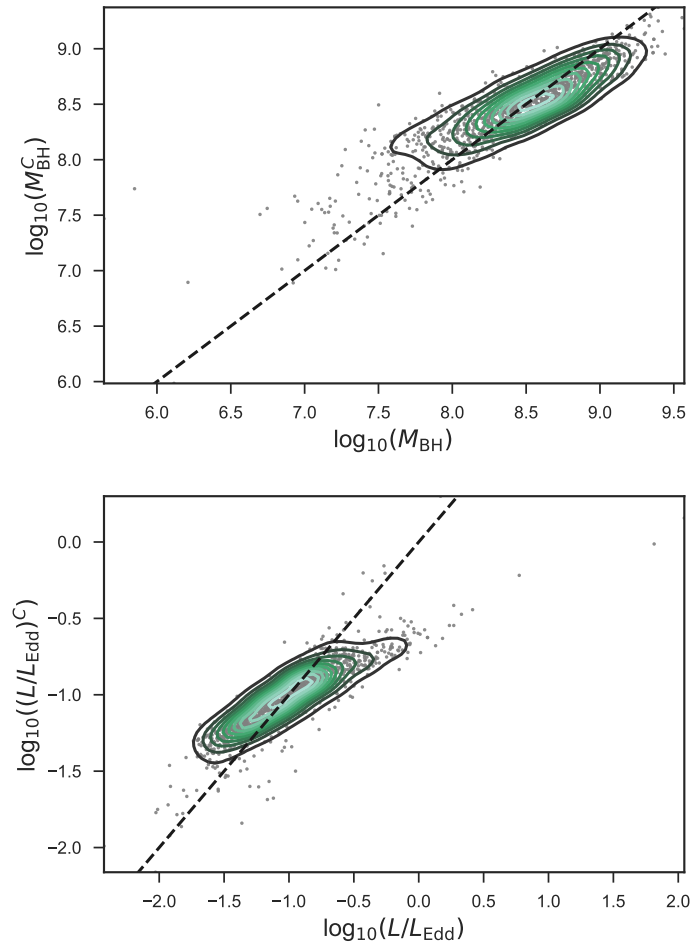


Figure 3.12: Comparison of M_{BH} vs. M_{BH}^C , and L/L_{Edd} vs. $(L/L_{\text{Edd}})^C$. The black dashed lines shows the 1:1 relations.

We also define another sub-sample composed by all the sources with M_{BH} determined from H_α , H_β , or Mg II line fitting. We call these sub-sample as the “not – C IV sample”. There are 1204 sources in this sub-sample, and 1112 are variable (92.4%).

The sub-sample composed by all the sources with M_{BH} determined from Mg II line fitting is called the “Mg II sample”. There are 1108 sources in this sub-sample, and 1029 are variable (92.9%). For 107 of the sources of this sample, the H_β line was available, and used along with the Mg II line to estimate M_{BH} .

Finally we define the sub-sample composed by all those sources whose only available emission line is C IV. We call this sample the “C IV sample”. 236 of the 269 sources in this sub-sample are variable (87.7%).

3.6 Variability parameters versus physical properties

In this section we discuss the different correlations between the physical parameters measured from the SDSS spectra and the variability features measured from the QUEST–La Silla light curves. For this analysis, we considered the different sub-samples defined in section 3.5, but in general, we worked with the most statistically significant results, which are found for the “not – C IV sample”.

3.6.1 Bivariate correlations

We first analysed the bivariate correlations between the variability features and the spectral properties, using the Spearman’s rank correlation coefficient (ρ_s), which does not consider errors measurements in the variables. It is important to remember that variability features can depend on more than one spectral property, and they would define an hyperplane which is not seen “on-edge” but instead through a projection onto specific axis. Therefore, some correlations can present large dispersions, even when there is a dependency of the variability feature on the spectral property.

Figure 3.13 shows the bivariate correlations between the variability features and the spectral properties. From the figure we can see that γ shows no correlation with z_{spec} , \dot{M} and L/L_{Edd} , and a very weak correlation with M_{BH} and L_{5100} . We also see that σ_{DRW} correlates weakly with z_{spec} and anti-correlates weakly with M_{BH} , L_{5100} , \dot{M} and L/L_{Edd} . This is consistent with the findings of Kelly et al. (2009), however we have to consider the strong effect of the light curve properties in this parameter when we interpret these results. Finally, from the figure we see that σ_{rms}^2 anti-correlates weakly with \dot{M} and L/L_{Edd} .

Crucially, we see that A correlates weakly with z_{spec} and M_{BH} , and anti-correlates weakly with \dot{M} and L/L_{Edd} . These weak correlations can be driven by the large dispersion produced by correlations with other variables. Moreover, we see a lack of correlation with L_{5100} , which is contrary to previous findings. It must be considered that the A parameter is measured for sources located at different redshifts and therefore the wavelength of rest-frame emission

(λ_{rest}) is different for every source. It is well known that the amplitude of the variability anti-correlates with rest-frame wavelength (as we point out in Chapter 2), which implies a positive correlation with redshift. Since L_{5100} correlates with z_{spec} (Figure 3.11), the anti-correlation between A and L_{5100} can be hidden by the positive correlation of A with redshift. Therefore, in order to detect correlations between A and any physical property, instead of looking for bivariate correlations, we must perform a multivariate analysis.

3.6.2 Principal Component Analysis

Principal Component Analysis (PCA, F.R.S. 1901; Francis & Wills 1999) is a mathematical tool used to reduce the dimensionality of a data set, and it is useful to understand the correlations present in multivariate data. PCA decompose the sample into a set of linearly independent Eigenvectors that are linear combinations of the original variables. We performed a PCA on our data set, for the case of the “not – C IV sample”, in order to see the dependencies between the different SMBH physical properties and the AGN variability features. We homogenized the data set by subtracting the mean values and normalizing by the variance. In the analysis we did not include the accretion rate (\dot{M}) and the excess variance (σ_{rms}), since these variables are highly correlated with the Eddington ratio (L/L_{Edd}) and the amplitude of the SF (A) respectively, and including them in the analysis produces principal components dominated by these correlations.

Table 3.1 shows the results of performing a PCA on the normalized variables. We show the first five principal components (PCs). The first row gives the variances (eigenvalues) associated with every PC. The second row gives the percentage of contribution of every PC to the total variance. The third row shows the cumulative percentage of variance carried by each eigenvector. It can be seen that the first four PC together contribute $\sim 95\%$ of the variance. In Table 3.1 we also show for each PC the weights associated to every input variable (eigenvectors). It can be seen that the first PC is dominated by the positive correlations between redshift, luminosity and BH mass. These correlations with redshift are produced by a selection effect, since at higher redshifts our sample will naturally contain more luminous and massive sources. The second PC is dominated by the anti-correlation between L/L_{Edd} and the amplitude of the variability, either measured from the SF or the DRW process. The third PC is dominated by γ , and the fourth by L/L_{Edd} . The fifth component is not very informative, since it contributes a small fraction of the total variance.

In order to have a better idea of the degree of correlation between the input variables, we computed the Spearman’s rank coefficients between the input variables and the first four PCs. The results are shown in Table 3.2. We can see again that the first PC is dominated by the positive correlation between z_{spec} , L_{5100} and M_{BH} . The second PC is dominated by the anti-correlation of the amplitude of the variability (A and σ_{DRW}) with L/L_{Edd} and L_{5100} . The third PC is dominated by γ , and demonstrates the positive correlation between γ and A . Besides, the third PC shows a possible correlation between γ and L/L_{Edd} . Finally, the fourth PC is dominated by L/L_{Edd} , and demonstrates an anti-correlation between L/L_{Edd} and M_{BH} , which is expected from the definition of L/L_{Edd} . From these results we can conclude that the most important correlation between variability features and physical properties is for the case of the amplitude of the variability with L/L_{Edd} and L_{5100} .

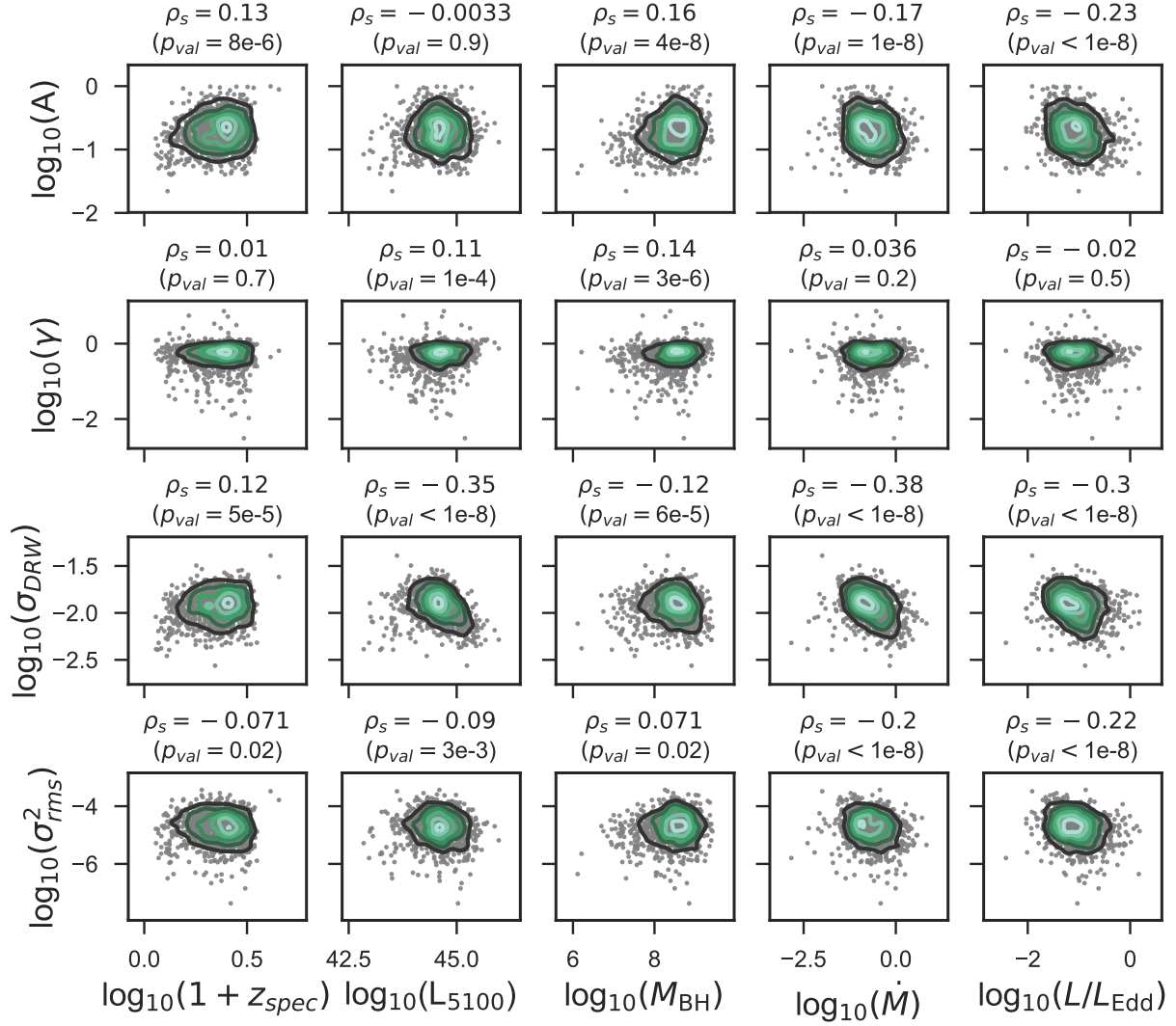


Figure 3.13: Bivariate correlations between the variability features and the spectral properties, for the “not – C IV sample”. The Spearman’s rank correlation coefficient is shown as reference for every pair of variables.

Table 3.1: PCA results: Eigenvectors and Eigenvalues.

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	2.457	1.808	1.346	1.029	0.257
Percentage	35.1%	25.8%	19.2%	14.7%	3.7%
Cumulative	35.1%	60.9%	80.1%	94.8%	98.5 %
Variable	PC1	PC2	PC3	PC4	PC5
$\log_{10}(1+z_{spec})$	-0.520	-0.118	0.231	-0.318	0.706
$\log_{10}(L_{5100}/10^{44})$	-0.520	-0.387	-0.046	-0.118	-0.414
$\log_{10}(M_{BH}/10^8)$	-0.572	-0.001	0.171	0.366	-0.262
$\log_{10}(L/L_{Edd})$	0.131	-0.478	-0.289	-0.644	-0.158
$\log_{10}(A)$	-0.271	0.540	-0.331	-0.280	-0.332
$\log_{10}(\gamma)$	-0.202	0.213	-0.749	0.097	0.343
$\log_{10}(\sigma_{DRW})$	-0.022	0.521	0.403	-0.499	-0.089

Table 3.2: PCA results: Spearman correlation coefficients between the input variables and the four first principal components (PC1, PC2, PC3, and PC4). the p-values of the coefficients are given in parentheses.

Variable	PC1	PC2	PC3	PC4
$\log_{10}(1+z_{spec})$	-0.670 (<1e-8)	-0.251 (<1e-8)	0.161 (1e-7)	0.006 (0.836)
$\log_{10}(L_{5100}/10^{44})$	-0.808 (<1e-8)	-0.615 (<1e-8)	-0.085 (0.005)	0.034 (0.252)
$\log_{10}(M_{BH}/10^8)$	-0.932 (<1e-8)	-0.066 (0.027)	0.243 (<1e-8)	0.643 (<1e-8)
$\log_{10}(L/L_{Edd})$	0.192 (<1e-8)	-0.712 (<1e-8)	-0.465 (<1e-8)	-0.831 (<1e-8)
$\log_{10}(A)$	-0.325 (<1e-8)	0.626 (<1e-8)	-0.426 (<1e-8)	-0.045 (0.131)
$\log_{10}(\gamma)$	-0.315 (<1e-8)	0.275 (<1e-8)	-0.802 (<1e-8)	0.057 (0.058)
$\log_{10}(\sigma_{DRW})$	0.130 (1e-5)	0.591 (<1e-8)	0.317 (<1e-8)	-0.173 (<1e-8)

Table 3.3: Linear Regression α , β and ε coefficients for A as the dependent variable (for the not – C IV sample). The columns headed by physical quantities refer to their slope in the regression model (β).

#	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{5100}/10^{44})$	$\log_{10}(M_{BH}/10^8)$	$\log_{10}(L/L_{Edd})$	$\log_{10}(\dot{M})$	ε
1	-0.88 ± 0.03	0.44 ± 0.08	X	X	X	X	0.26 ± 0.01
2	-0.74 ± 0.01	X	0.03 ± 0.02	X	X	X	0.26 ± 0.01
3	-0.77 ± 0.01	X	X	0.12 ± 0.02	X	X	0.25 ± 0.01
4	-0.93 ± 0.03	X	X	X	-0.19 ± 0.02	X	0.25 ± 0.01
5	-0.79 ± 0.01	X	X	X	X	-0.09 ± 0.02	0.26 ± 0.01
6	-0.87 ± 0.03	0.52 ± 0.11	-0.21 ± 0.03	0.18 ± 0.03	X	X	0.25 ± 0.01
7	-1.08 ± 0.04	0.50 ± 0.11	-0.02 ± 0.03	X	-0.18 ± 0.03	X	0.25 ± 0.01
8	-1.14 ± 0.06	0.57 ± 0.11	X	-0.04 ± 0.03	-0.22 ± 0.03	X	0.25 ± 0.01
9	-0.92 ± 0.03	0.72 ± 0.10	-0.10 ± 0.03	X	X	X	0.25 ± 0.01
10	-0.83 ± 0.03	0.18 ± 0.10	X	0.09 ± 0.02	X	X	0.25 ± 0.01
11	-1.09 ± 0.04	0.45 ± 0.07	X	X	-0.19 ± 0.02	X	0.25 ± 0.01

3.6.3 Multiple Linear Regression

In the previous section we showed that the amplitude of the variability anti-correlates with L/L_{Edd} and L_{5100} , however from our PCA we cannot say whether the amplitude of the variability is mainly driven by L/L_{Edd} or L_{5100} . Besides, is still not clear whether γ correlates with any physical property, but from the PCA there is a possible positive correlation between γ and L/L_{Edd} .

In order to have a better idea of the correlations between variability parameters and physical properties we computed Bayesian multiple linear regression. We used the Bayesian linear regression procedure of Kelly (2007), which takes into account the measurement uncertainties of every variable and includes the intrinsic scatter inherent to the relation. The following sections give the results of this analysis.

Trends of the amplitude of the SF with physical properties

In the previous section we showed that A presents an anti-correlation with L/L_{Edd} and L_{5100} , a multiple linear regression analysis can help us to differentiate which physical property drives these anti-correlations.

In Chapter 2 showed that there is a positive correlation between A with z_{spec} , which is produced by an anti-correlation between A with λ_{rest} (see Figure 2.11). Therefore, given the wide range in redshift of our sample, we must always consider the correlation with redshift when we analyse correlations with any other physical parameter.

Table 3.3 shows the results of the Bayesian multiple linear regression for A as the dependent variable, and different combinations of the spectral properties as independent variables. In the table, every column gives the value of the intercept (α), the slope (β) associated with a given physical property, and the intrinsic scatter associated to the regression model (ε). When the value of the slope is replaced by X, it means that the parameter was not included in the regression model.

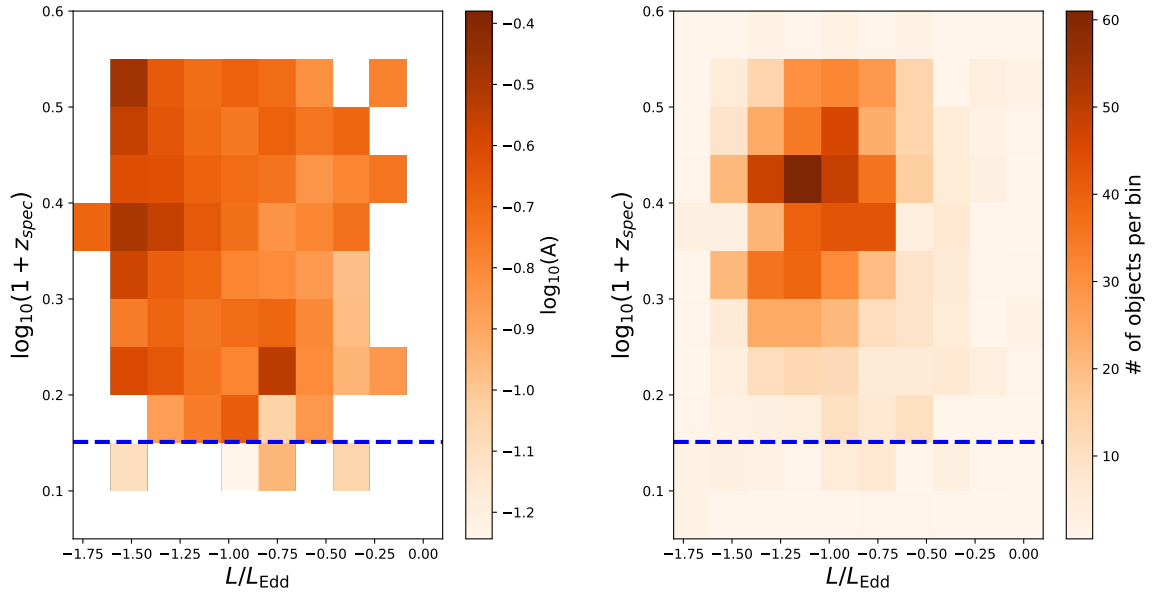


Figure 3.14: Left: Mean value of A in a grid of z_{spec} and L/L_{Edd} , for bins with 3 or more sources. Right: number of sources per bin of z_{spec} and L/L_{Edd} . The red dashed line shows the redshift from which we have available the Mg II line in the SDSS spectra (0.42).

Regressions #1 to #5 in Table 3.3 correspond to models with one single independent variable. We see that the most significant correlations are for z_{spec} , L/L_{Edd} and M_{BH} . Since \dot{M} and L/L_{Edd} are highly correlated (see Figure 3.11), and including these two variables together can produce multicollinearity in the regression model, we decided to exclude \dot{M} from the regression models, and keep L/L_{Edd} .

For the regressions #6 to #10, we decided to include always z_{spec} as one of the independent variables, because we are analysing light curves observed in a fixed photometric band (Q), which implies that the rest-frame wavelength of every light curve will depend on the redshift of the source. Besides, we do not include in Table 3.3 a regression model with z_{spec} , L_{5100} , M_{BH} , and L/L_{Edd} as independent variables, because the multicollinearity of the variables does not allow the Bayesian method to converge and return confident regression coefficients.

Regression #6 shows that when the model includes z_{spec} , M_{BH} and L_{5100} , the slopes for M_{BH} and L_{5100} satisfy (within 1σ) the relation: $\beta_{L_{5100}} \sim -\beta_{M_{\text{BH}}}$. This would be expected if L/L_{Edd} is the driver of the amplitude variability.

Regression #11 corresponds to a model which includes both the Eddington ratio and redshift. We can see that the coefficients are statistically significant, and thus we propose this model as the best regression model for the amplitude of the variability. This can be confirmed when we see regressions #7 and #8, where adding L_{5100} or M_{BH} in the model, besides L/L_{Edd} and redshift, gives statistically insignificant slopes for M_{BH} or L_{5100} .

Figure 3.14 shows the dependency of A with z_{spec} and L/L_{Edd} . We can see the anti-

Table 3.4: Linear Regression α , β and ε coefficients for A as the dependent variable (for standard black hole masses). The columns headed by physical quantities refer to their slope in the regression model (β).

sample	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L/L_{Edd})$	ε	
QUEST-SDSS	-1.01 ± 0.03	0.41 ± 0.06	-0.13 ± 0.02	0.25 ± 0.01	
not - C IV	-1.09 ± 0.04	0.45 ± 0.07	-0.19 ± 0.02	0.21 ± 0.01	
Mg II	-1.02 ± 0.04	0.23 ± 0.09	-0.21 ± 0.03	0.24 ± 0.01	
C IV	-0.47 ± 0.20	-0.37 ± 0.36	0.02 ± 0.04	0.25 ± 0.01	
sample	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{5100}/10^{44})$	$\log_{10}(M_{BH}/10^8)$	ε
QUEST-SDSS	-0.86 ± 0.02	0.52 ± 0.08	-0.17 ± 0.03	0.13 ± 0.02	0.25 ± 0.01
not - C IV	-0.87 ± 0.03	0.52 ± 0.11	-0.21 ± 0.03	0.18 ± 0.03	0.25 ± 0.01
Mg II	-0.79 ± 0.04	0.43 ± 0.11	-0.26 ± 0.03	0.18 ± 0.03	0.24 ± 0.01
C IV	-0.51 ± 0.20	-0.16 ± 0.34	-0.08 ± 0.09	-0.02 ± 0.04	0.25 ± 0.01

correlation of A with L/L_{Edd} , and the positive correlation of A with redshift. The trend with L/L_{Edd} is more significant.

In order to break the effects of the $A - \gamma$ degeneracy, we re-computed the regression model #11 of Table 3.3, but considering only variable sources whose measured γ range between $\gamma_{med} \pm 0.1$, where γ_{med} corresponds to the median value of γ measured for the well-sampled sub-sample (0.53), i.e. we consider those sources with $0.43 \leq \gamma \leq 0.63$. Selecting a narrow range in measured γ allows us to better discriminate between different *intrinsic* values of the amplitude A , as can be seen in Figure 3.5. There are 322 variable sources from the not - C IV sample in this range of γ . In this case, the results of the regression are: $\alpha = -1.06 \pm 0.05$, $\beta_{\log_{10}(1+z_{spec})} = 0.45 \pm 0.1$, $\beta_{\log_{10}(L/L_{Edd})} = -0.22 \pm 0.03$, and $\varepsilon = 0.14 \pm 0.01$. The results for the slopes of z_{spec} and L/L_{Edd} are consistent with what we found for the whole not - C IV sample, at 1σ level. However, there is a considerable reduction in the intrinsic scatter measured for the reduced sample. This implies that a large part of the scatter measured in the different regression models of Table 3.3 comes from the $A - \gamma$ degeneracy.

Table 3.4 gives the linear regression coefficients for z_{spec} and L/L_{Edd} considering the different samples defined in section 3.5. From the table we can see that the results for the C IV sample are different from the results obtained for the other three samples. For the QUEST-SDSS sample, we can see a decrement in the slope of L/L_{Edd} , produced by the presence of sources with C IV measurements. The slope of L/L_{Edd} is consistent for the not - C IV and Mg II samples, but the slope for z_{spec} changes. This can be produced by the reduced dynamic range in redshift for the Mg II sample, in comparison with the not - C IV sample. Moreover, from Figure 3.14 we can see that when we only consider sources from the Mg II sample (sources above the red dashed line), we lose most of the sources with low variability.

Table 3.4 also shows the linear regression coefficient for a model with redshift, L_{5100} and M_{BH} . We can see that, as we showed in Table 3.3, for the not - C IV sample the slopes of L_{5100} and M_{BH} satisfy $\beta_{L_{5100}} \sim -\beta_{M_{BH}}$. However, for the QUEST-SDSS and Mg II samples, the relation between the slopes is not so evident. Again, for the case of the QUEST-SDSS sample we have contamination from C IV. The difference between the results for the Mg II and not - C IV samples can be driven by the change in the dynamic range of z_{spec} (see Figure 3.11), and the strong correlation between z_{spec} and L_{5100} (produced by a selection effect),

Table 3.5: Linear Regression α , β and ε coefficients for A as the dependent variable (for corrected black hole masses). The columns headed by physical quantities refer to their slope in the regression model (β).

sample	α	$\log_{10}(1+z_{spec})$	$\log_{10}((L/L_{Edd})^C)$	ε	
QUEST-SDSS	-1.19 ± 0.08	0.48 ± 0.07	-0.28 ± 0.06	0.25 ± 0.01	
not – C IV	-1.22 ± 0.08	0.56 ± 0.08	-0.28 ± 0.06	0.25 ± 0.01	
Mg II	-1.16 ± 0.12	0.36 ± 0.10	-0.30 ± 0.08	0.25 ± 0.01	
C IV	4.33 ± 20.35	-2.08 ± 11.80	4.11 ± 21.21	0.23 ± 0.04	
sample	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{5100}/10^{44})$	$\log_{10}(M_{BH}^C/10^8)$	ε
QUEST-SDSS	-0.84 ± 0.03	0.42 ± 0.09	-0.29 ± 0.05	0.31 ± 0.06	0.25 ± 0.01
not – C IV	-0.88 ± 0.03	0.55 ± 0.11	-0.28 ± 0.05	0.29 ± 0.06	0.25 ± 0.01
Mg II	-0.81 ± 0.04	0.48 ± 0.12	-0.27 ± 0.06	0.20 ± 0.08	0.25 ± 0.01
C IV	-0.59 ± 2.2	-0.33 ± 4.23	2.00 ± 6.28	-2.73 ± 8.26	0.23 ± 0.04

since the slope for z_{spec} also increases when we include L_{5100} in the model.

In order to see whether smaller ranges of redshift can reduce the effects in the regression analysis of the z_{spec} versus L_{5100} correlation, we computed the regression models again, but considering sources from the not – C IV sample with $1.5 \leq z_{spec} \leq 1.8$, since in this range of redshift the correlation between z_{spec} and L_{5100} is smaller ($\rho_s = 0.12$, $p_{val} = 0.09$). We found similar results, but the results are less statistically significant due to the low number of sources considered (there are 213 variable sources in this range of redshift). For the case of the regression model with z_{spec} , L_{5100} and M_{BH} , the results of the regression are: $\alpha = 0.03 \pm 0.55$, $\beta_{\log_{10}(1+z_{spec})} = 1.51 \pm 1.30$, $\beta_{\log_{10}(L_{5100})} = -0.20 \pm 0.08$, $\beta_{\log_{10}(M_{BH})} = -0.12 \pm 0.07$, and $\varepsilon = 0.21 \pm 0.01$. And for the case of the regression with z_{spec} and L/L_{Edd} , the results of the regression are: $\alpha = -0.22 \pm 0.56$, $\beta_{\log_{10}(1+z_{spec})} = 1.50 \pm 1.29$, $\beta_{\log_{10}(L/L_{Edd})} = -0.15 \pm 0.06$, and $\varepsilon = 0.21 \pm 0.01$. From these results, we can see that $\beta_{\log_{10}(L_{5100})} \sim -\beta_{\log_{10}(M_{BH})}$ at 1σ level, and that the anti-correlation between A and (L/L_{Edd}) is still present.

As mentioned in section 3.4.2, Mejía-Restrepo et al. (2018a) proposed new corrections for the estimation of M_{BH} . We show in Table 3.5 the linear regression coefficients for a model with z_{spec} and $(L/L_{Edd})^C$, and for a model with z_{spec} , L_{5100} , and M_{BH}^C , for the different samples of section 3.5. We can see that the slopes for z_{spec} and $(L/L_{Edd})^C$ increase for the QUEST-SDSS, not – C IV, and C IV samples, but the errors in the slopes also increase. The slopes for $(L/L_{Edd})^C$ are more similar for the different samples than the slopes for z_{spec} . This can be related with the difference in the dynamic ranges of z_{spec} for different samples. For the regression model with z_{spec} , L_{5100} , and M_{BH}^C , the slopes of L_{5100} and M_{BH}^C satisfy the relation $\beta_{L_{5100}} \sim -\beta_{M_{BH}^C}$ for the QUEST-SDSS, not – C IV and Mg II samples (at 1σ). This supports our idea that L/L_{Edd} is the driver of the variability amplitude. The difference of these results with what we showed in Table 3.4 can be given by the reduction of the scatter in the determined black hole mass when we use the corrections proposed by Mejía-Restrepo et al. (2018a).

Table 3.6 shows three linear regression models that consider different spectral properties of the Mg II line as the independent variable. The first regression model corresponds to a model with $L_{Mg II}/L_{3000}$, which is a proxy of the EW of the line. Previous analysis have found that there is a strong anti-correlation between the equivalent width of Mg II and

Table 3.6: Linear Regression α , β and ε coefficients for A as the dependent variable, for spectral properties derived from Mg II. The columns headed by physical quantities refer to their slope in the regression model (β).

α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{Mg II}/L_{3000})$	ε
0.02 ± 0.09	0.29 ± 0.09	0.44 ± 0.05	0.24 ± 0.01
α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{5100}/L_{3000})$	ε
-0.67 ± 0.21	-0.55 ± 0.93	-0.19 ± 0.13	0.22 ± 0.02
α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{3000}/L_{1450})$	ε
-0.64 ± 0.26	-0.06 ± 0.55	0.21 ± 0.15	0.27 ± 0.01

Table 3.7: Linear Regression α , β and ε coefficients for A as the dependent variable (spectral properties per emission line). The columns headed by physical quantities refer to their slope in the regression model (β).

line	α	$\log_{10}(1+z_{spec})$	$\log_{10}(\text{FWHM})$	$\log_{10}(\lambda L_{\lambda}/10^{44})$	ε
H $_{\alpha}$ (57)	-8.76 ± 5.20	1.29 ± 1.65	0.39 ± 0.16	0.14 ± 0.12	0.26 ± 0.03
H $_{\beta}$ (172)	-2.76 ± 2.31	1.62 ± 0.46	0.23 ± 0.08	0.02 ± 0.05	0.24 ± 0.01
Mg II (1063)	3.50 ± 1.11	0.44 ± 0.11	0.34 ± 0.05	-0.12 ± 0.02	0.24 ± 0.01
C IV (460)	3.60 ± 1.93	0.02 ± 0.23	-0.08 ± 0.08	-0.09 ± 0.04	0.26 ± 0.01

Note. In parentheses we show the number of variable sources considered per line.

L/L_{Edd} (see Netzer 2013 and references therein). Our results shows a positive correlation between $L_{\text{Mg II}}/L_{3000}$ and A , which supports our interpretation that L/L_{Edd} is the driver of the amplitude.

The second and third regression models of Table 3.6 include the spectral slopes L_{5100}/L_{3000} and L_{3000}/L_{1450} . We can see that there is not statistically significant correlation between the amplitude of the variability and these spectral slopes.

We looked for correlations between A and parameters derived from the line fitting. Table 3.7 shows the regression coefficients for models that consider the FWHM and continuum luminosity (λL_{λ}) for the H $_{\alpha}$, H $_{\beta}$, Mg II and C IV lines. For the case of H $_{\alpha}$, the statistics is poor given the low number of variable sources with this line available in the SDSS spectra. Despite that, we can see a positive correlation between A and $\text{FWHM}(H_{\alpha})$. For the case of H $_{\beta}$, the results are similar, with a positive correlation between A and $\text{FWHM}(H_{\beta})$. Mg II has the best statistics, with 1063 variable sources available. In this case we also see a correlation between A and $\text{FWHM}(\text{Mg II})$, but also an anti-correlation between A and L_{3000} . The results for C IV are completely different, with no significant correlation between A and $\text{FWHM}(\text{C IV})$ or L_{1450} . This can be related with the known problems of using the C IV line to measure black hole masses, since the line profile deviates considerably from Keplerian-type motion, and can be influenced by winds emanating from the accretion disk (see Netzer 2013; Mejía-Restrepo et al. 2018b, and references therein).

The positive correlations between A and the FWHM of H $_{\alpha}$, H $_{\beta}$, and Mg II are expected for a variability process whose amplitude is driven by L/L_{Edd} , since $L/L_{\text{Edd}} \propto \text{FWHM}^{-2}(\lambda L_{\lambda})^{1-\alpha}$ (following the equations of section 3.4.2). Under this assumption, the anti-correlation between A and L_{3000} is also expected. The lack of correlation between A and L_{6200} , and L_{5100}

Table 3.8: Linear Regression α , β and ε coefficients for other amplitude features as dependent variables (not – C IV sample). The columns headed by physical quantities refer to their slope in the regression model (β).

feature	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L/L_{Edd})$	ε	
σ_{DRW}	-2.20 ± 0.02	0.26 ± 0.04	-0.16 ± 0.01	0.12 ± 0.004	
σ_{rms}^2	-5.07 ± 0.07	-0.13 ± 0.14	-0.34 ± 0.05	0.48 ± 0.01	
feature	α	$\log_{10}(1+z_{spec})$	$\log_{10}(L_{5100}/10^{44})$	$\log_{10}(M_{BH}/10^8)$	ε
σ_{DRW}	-2.13 ± 0.01	1.04 ± 0.05	-0.33 ± 0.01	0.06 ± 0.01	0.09 ± 0.0034
σ_{rms}^2	-4.64 ± 0.06	-0.31 ± 0.21	-0.29 ± 0.06	0.35 ± 0.05	0.48 ± 0.01

can be given by the differences in the continuum luminosity range covered by these lines compared to Mg II.

Trends of other amplitude features with physical properties

We tested whether σ_{rms}^2 and σ_{DRW} also show correlations with z_{spec} and L/L_{Edd} or L_{5100} and M_{BH} . The results are shown in Table 3.8. For the case of σ_{rms}^2 , we see a significant anti-correlation with L/L_{Edd} , which is consistent with what we found for A . We also found a lack of significant correlation with z_{spec} . This can be given by the positive correlation between the amplitude of the variability and z_{spec} and the negative correlation between the length of the light curve and z_{spec} . Since σ_{rms}^2 considers the variance of the whole light curve, for sources at high redshift we observe shorter light curves than at low redshift, and therefore the correlation with z_{spec} is considerably diminished.

For the case of σ_{DRW} , we see an anti-correlation with L/L_{Edd} and a positive correlation with z_{spec} . This is in contrast with the results reported by Kelly et al. (2009), who found no correlation between L/L_{Edd} and σ_{DRW} . This can be given by the strong dependency of σ_{DRW} on the sampling of the light curve, and the considerably small number of sources, with respect to our sample, used by Kelly et al. (2009). Our results also show that σ_{DRW} correlates negatively with L_{5100} and has no correlation with M_{BH} . Kelly et al. (2009) found a similar slope for L_{5100} for their model with z_{spec} included (see their Eq. 25). Since we found no correlation with M_{BH} , we propose that the anti-correlation between σ_{DRW} and L/L_{Edd} is given by the anti-correlation between σ_{DRW} and L_{5100} . We must consider the implication of these results with caution, since σ_{DRW} is strongly affected by the light curve sampling. Particularly, Figure 3.2 shows that σ_{DRW} anti-correlates with the number of epochs and the length of the light curve. More luminous sources have higher probabilities to be detected in more epochs than fainter sources. In fact, the Spearman’s rank correlation coefficient for L_{5100} and the number of epochs is 0.45 ($p_{val} = 1e-4$). Therefore, the anti-correlation between σ_{DRW} and L_{5100} can be just a reflection of the anti-correlation between σ_{DRW} and the number of epochs.

Table 3.9: Linear Regression α , β and ε coefficients for γ as the dependent variable (not – C IV). The columns headed by physical quantities refer to their slope in the regression model (β).

α	$\log_{10}(1+z_{spec})$	ε
-0.32 ± 0.03	0.01 ± 0.08	0.28 ± 0.01
α	$\log_{10}(L_{5100}/10^{44})$	ε
-0.37 ± 0.01	0.09 ± 0.02	0.28 ± 0.01
α	$\log_{10}(M_{BH}/10^8)$	ε
-0.35 ± 0.01	0.08 ± 0.02	0.28 ± 0.01
α	$\log_{10}(L/L_{Edd})$	ε
-0.34 ± 0.03	-0.01 ± 0.03	0.28 ± 0.01
α	$\log_{10}(M)$	ε
-0.30 ± 0.02	0.03 ± 0.02	0.28 ± 0.01

Trends of the logarithmic gradient of the variability (γ) with physical properties

In section 3.6.1 we showed that γ correlates very weakly with L_{5100} and M_{BH} (see Figure 3.13), and from the PCA there is evidence of a positive correlation between γ and L/L_{Edd} . In order to test whether any of these correlations exists, we performed a linear regression analysis. Table 3.9 shows the linear regression coefficients for γ , when we consider spectral features as single independent variables in the regression model, for the case of the not – C IV sample. We can see that γ does not have statistically significant correlation with any physical parameter, since the absolute values of the slopes for every the regression model are small (lower than 0.1) and/or have high errors compared to the measured values.

Some sources have values of γ that are inconsistent with a DRW process. In section 3.3.4 we showed that if a measured value of γ range between 0.0 and 0.75, we cannot discard a DRW process as the best model to describe the variability. In the well-sampled sub-sample, 325 of the 1579 variable sources have values of γ higher than 0.75 (20.6% of the sample). For these sources, the value of γ differs considerably from 0.5, and therefore, a DRW model is not sufficient to model the variability. When we compare the distributions of the SMBH physical properties of a) the 325 sources with $\gamma > 0.75$, and b) the rest of the sample; we do not observe any difference between the populations. We also do not observe differences in the light curve sampling of these two populations, we can therefore discard an observational bias in the distribution of γ .

3.6.4 Differences between variable and non-variable sources

From the QUEST-SDSS sample, 1348 sources are variable and 125 are non-variable. Figure 3.15 shows the normalized distribution of the different physical properties considered in this work, for the variable and non-variable sources. In the figure we can see that the distributions of L_{5100} , M_{BH} and L/L_{Edd} are similar, but for the case of non-variable sources, the distribution of z_{spec} is in general shifted towards higher values of redshift, with the exception

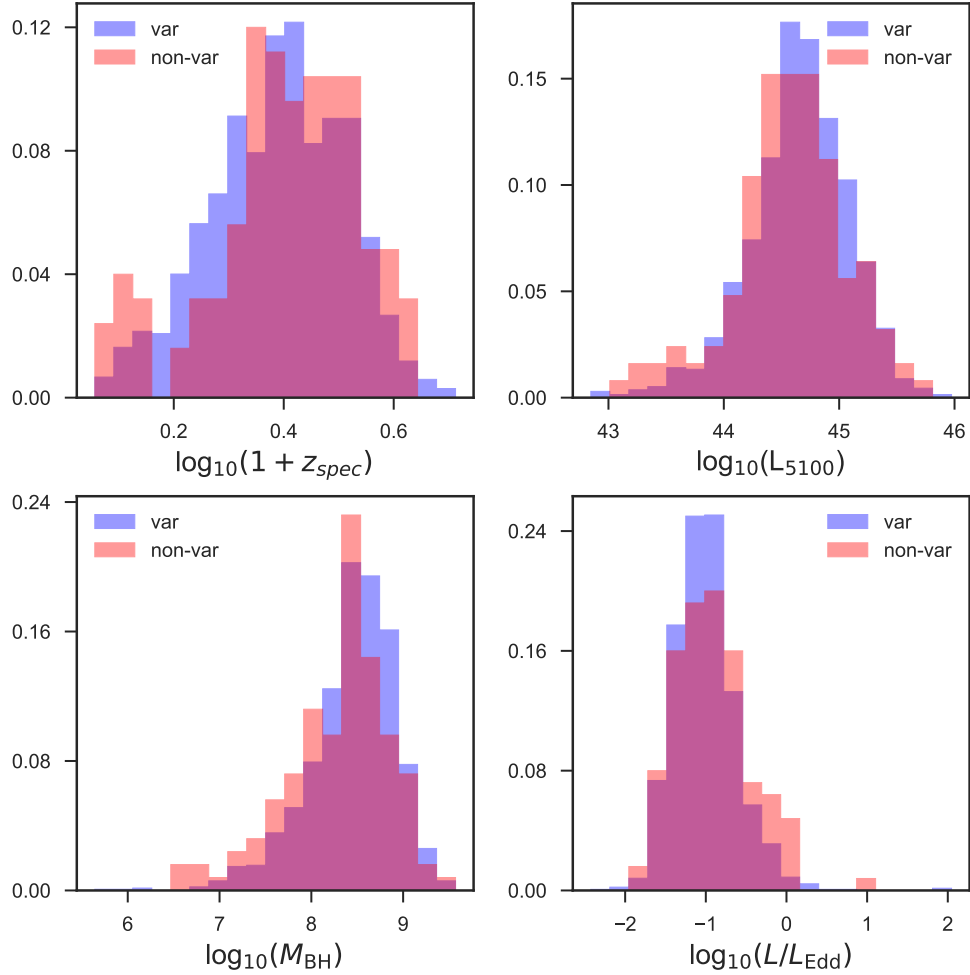


Figure 3.15: Normalized histogram of the physical properties of variable (blue) and non-variable (red) sources from the QUEST-SDSS sample.

of a few sources located at low z_{spec} . This difference in redshift can be related with the fact that high redshift sources have shorter rest-frame light curves, because of the time dilation. In Figure 3.16 we show the normalized distribution of the light curve properties of variable and non-variable sources from the QUEST-SDSS sample. In the figure we can see that non-variable sources tend to have lower number of epochs, shorter light curves in rest-frame, and fainter mean magnitudes. Therefore, in our sample, the light curve properties are more relevant for the classification of variable and non-variable sources than the physical properties of the SMBH.

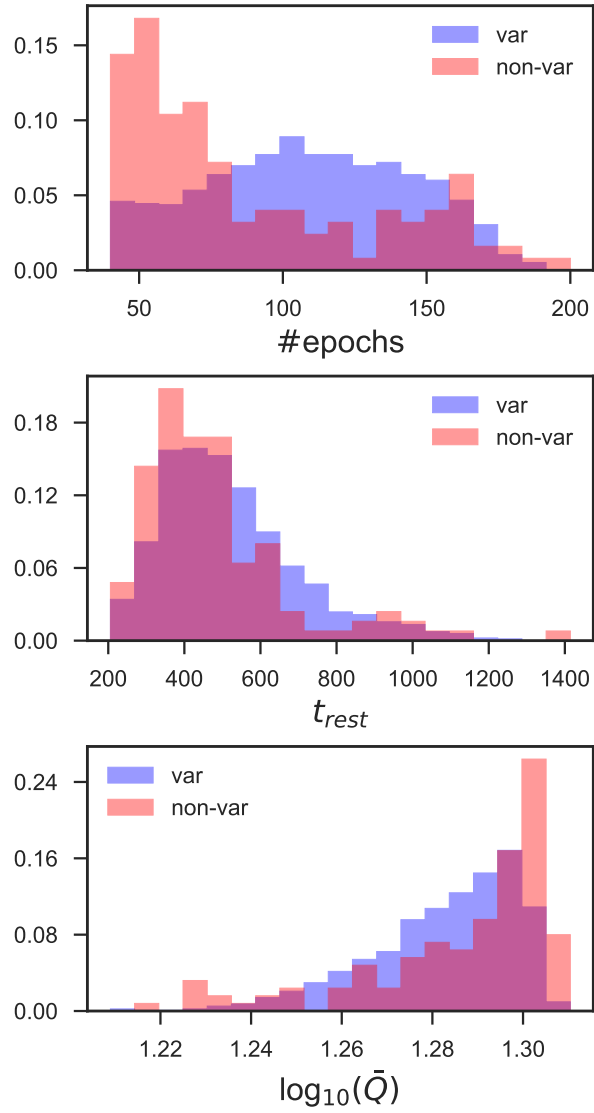


Figure 3.16: Normalized histogram of the light curve properties (number of epochs, t_{rest} , and mean magnitude) of variable (blue) and non-variable (red) sources from the QUEST-SDSS sample.

3.7 Variability behavior of different classes of AGN

3.7.1 BAL–QSO

We used the catalogs of Shen et al. (2011), Pâris et al. (2017b) and Pâris et al. (2017a) to classify 133 sources of our sample as BAL–QSO. 99 of these sources have light curves with good sampling, and 86 are variable (86,9%). In Figure 3.9 we show in red the distribution of A and γ for the well sampled and variable BAL–QSO, we can see that there is no evident difference in the SF parameters.

In order to have a more quantitative comparison of the distribution of the SF parameters of BAL–QSO and the rest of the sample, we performed a two–sample Anderson–Darling test (Pettitt, 1976) for the A and γ parameters. Since the Anderson–Darling test does not take into account the errors of the parameters, we only considered in the test those variable sources with a measured parameter having a signal to noise ratio higher than 3. According to the test, the distributions of the A parameter are the same at a 99.5% significance level, with a p_{val} of 0.9, and the distributions of the γ parameter are different, with a p_{val} of 0.02. The difference in the distributions of γ are related with the fact that for the rest of the sample, most of the sources are concentrated around $\gamma = 0.5$, but for the case of BAL–QSO, the sources are more homogeneously distributed in the SF parameter space. Nevertheless, we do not see that the values of γ for BAL–QSO are systematically different than the rest of the sample.

3.7.2 Radio Classification

We divided our sources as radio–loud (RL) or radio–quiet (RQ) according to their radio and optical emissions. We used data from the Faint Images of the Radio Sky at Twenty cm survey (FIRST, Becker et al. 1994) to obtain fluxes at 20 cm of our sources. FIRST used the Very Large Array (VLA) to produce a map of the 20 cm (1.4 GHz) sky with a beam size of $5''.4$ and an rms sensitivity of about $0.15 \text{ mJy beam}^{-1}$. The last version of the FIRST survey catalog (14Dec17 Version¹), provides all the sources detected with a threshold of 1 mJy. We cross–matched our QUEST–SDSS sample with the FIRST catalog, using a radius of $1''$.

We classified our sources as RL and RQ using the ratio:

$$R = \frac{F_{\nu}(5 \text{ GHz})}{F_{\nu}(4400 \text{ \AA})} \quad (3.1)$$

where $F_{\nu}(5 \text{ GHz})$ is the radio flux density of the source measured at 5 GHz and $F_{\nu}(4400 \text{ \AA})$ is the flux density at 4400 \AA (Kellermann et al., 1989). We applied K–corrections to the photometry provided by the SDSS and FIRST catalogs, considering that the radio and optical emissions follow a power–law like $F \propto \nu^{-0.8}$ and $F \propto \nu^{-0.44}$ respectively. Then, we estimate

¹<http://sundog.stsci.edu/first/catalogs/readme.html#coverage>

$F_\nu(5 \text{ GHz})$ from the measurements at 1.4 GHz provided by FIRST, and $F_\nu(4400 \text{ \AA})$ from the g SDSS band (4770 \AA) measurements. Therefore, the final flux values used to determine R were:

$$\begin{aligned} F_\nu(5 \text{ GHz})_{\text{rest}} &= F_\nu(1.4 \text{ GHz})_{\text{obs}} \left(\frac{1.4}{5}\right)^{0.8} (1+z)^{-0.2} \\ F_\nu(4400 \text{ \AA})_{\text{rest}} &= F_\nu(4770 \text{ \AA})_{\text{obs}} \left(\frac{6.29}{6.81}\right)^{0.44} (1+z)^{-0.56} \end{aligned} \quad (3.2)$$

All the sources of the QUEST–SDSS sample are located in regions mapped by FIRST, however not all of them have a radio detection associated. 55 objects from the QUEST–SDSS sample have a FIRST counterpart. For those sources without a detection reported, we assumed that the measured flux corresponds to the detection threshold of 1 mJy. Then, we classify sources as RL if they have $R \geq 10$ and are detected by FIRST, and we classify sources as RQ if they have $R < 10$.

From the QUEST–SDSS sample, 373 sources are classified as RQ and 354 are variable (94,9 %). 48 are classified as RL and 38 are variable (79.2%). Figure 3.17 shows the distribution of the SF parameters for RL and RQ sources, and also for sources without radio classification. In the Figure we can see that there is no evident difference in the distributions of RL and RQ sources. For a more quantitative comparison, we performed an Anderson–Darling test comparing the SF parameters distributions of the RL and RQ sources. As before, we only considered those variable sources with a measured parameter having a signal to noise ratio higher than 3. According to the test, the distributions of A and γ are the same for RQ and RL sources, with p_{val} of 0.64 for A and 0.22 for γ . This could imply that the radio loudness may not be relevant for the optical variability of type I AGN.

3.8 Discussion and Conclusions

In section 3.3.5 we showed that there is a correlation between A and γ , however in section 3.3.4 we demonstrated that such a correlation is produced by the stochastic nature of the light curves together with the light curve sampling. We might need longer light curves to reduce this degeneracy, however, having access to long (decades of coverage) and well sampled light curves for large samples of sources is not possible currently. In section 3.6 we demonstrated that A anti-correlates with both λ_{rest} and L/L_{Edd} , but γ does not correlate with any of the physical parameters studied. This confirms our assumption that the correlation between A and γ is produced by the light curve properties, and by the stochastic nature of the variability. Nonetheless, the structure function is the best option that we have today to analyse typical ground-based light curves (i.e. with a few years of coverage, a few epochs, and with gaps), since other techniques, like Fourier analyses, requires well sampled light curves (i.e. with several epochs, and without gaps).

In sections 3.4.3 and 3.6.2 we reported that our data set presents correlations between

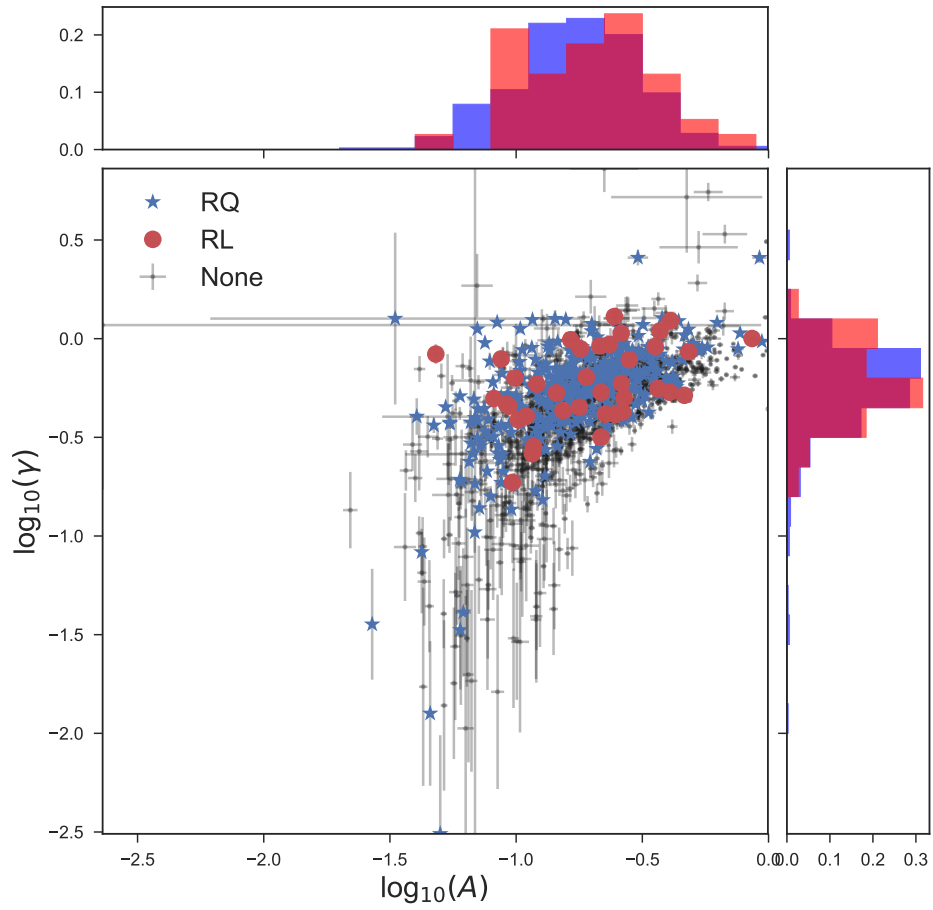


Figure 3.17: Distribution of the SF parameters A and γ , for sources classified as RL (red circles), RQ (blue stars), and sources without radio classification (None, black dots). Along the axes we show the histograms of every parameter, for the case of RL (blue) and RQ sources (red).

the SMBH physical properties. Some of these correlations are produced by selection effects, since our sample is flux limited. For example, the correlations of redshift with luminosity, BH mass, and accretion rate, are produced by the fact that at higher redshifts our sample will naturally contain sources with higher luminosities, higher SMBH masses and higher accretion rates.

The results shown in section 3.6 tell us that the observed amplitude of the variability depends on two variables, z_{spec} and L/L_{Edd} , which means that, at a fixed z_{spec} , sources with similar L/L_{Edd} will have similar variability amplitudes. The positive correlation with redshift can be interpreted as an anti-correlation with the wavelength of rest-frame emission. MacLeod et al. (2010) analysed *ugriz* light curves of ~ 9000 spectroscopically confirmed SDSS S82 quasars. Since they had multiple bands for each quasar, they could separate the dependency of the amplitude of the variability with redshift and λ_{rest} , finding an anti-correlation with λ_{rest} and no correlation with z_{spec} . In Chapter 2 we analysed the near infrared variability of X-ray selected AGN. We also found a correlation between the amplitude of the variability and redshift. By comparing the trends between A and z_{spec} for two different bands (Y and J), we showed that the correlation with z_{spec} is explained by an anti-correlation with the wavelength of emission. From this, and other previous results (e.g., Kozłowski et al. 2010), we conclude that the positive correlation of A with z_{spec} is produced by a dependency on λ_{rest} , and is not given by evolution over cosmic time. This anti-correlation between λ_{rest} and A can be explained considering that the innermost regions of the disk can be the most variable, either intrinsically or by reprocessing. Since at shorter wavelengths a larger fraction of the disc emission is produced by the innermost region, it follows that shorter wavelengths display larger amplitudes of variability (Arévalo et al., 2008; Lira et al., 2011, 2015; Edelson et al., 2015; Fausnaugh et al., 2016).

We found an anti-correlation between A and L/L_{Edd} . When we used the standard method to determine black hole masses from single epoch spectra we found a slope of $\beta_{L/L_{Edd}} = -0.19 \pm 0.02$ for the regression model with A as the dependent variable, and L/L_{Edd} and z_{spec} as the independent variables, for the not - C IV sample. When we apply the corrections proposed by Mejía-Restrepo et al. (2018a), which intend to account for the effect of the unknown distribution of the gas clouds in the BLR, we found a slope of $\beta_{(L/L_{Edd})^C} = -0.28 \pm 0.06$. An anti-correlation between L/L_{Edd} and the amplitude of the variability has also been reported by previous works (e.g., Wilhite et al. 2008; MacLeod et al. 2010; Simm et al. 2016; Rakshit & Stalin 2017). MacLeod et al. (2010) reported a power-law slope of -0.23 ± 0.03 . This value was calculated by binning the parameter space of M_{BH} and M_i (absolute magnitude), and using ensemble light curves, which can explain the small difference with the value found by us.

MacLeod et al. (2010) also proposed that an additional dependency with luminosity or black hole mass is needed in order to explain their findings. Here we conclude that such a dependency is not necessary when intrinsic scatter is included in the model. In fact, in Table 3.3 we can see that the value of the intrinsic scatter found with our method is pretty stable, and that including L_{5100} or M_{BH} in the regression model with z_{spec} and L/L_{Edd} produces statistically insignificant slopes for L_{5100} or M_{BH} . From the results of sections 3.3.4 and 3.6.3 we can say that the main contributors to this scatter are the $A - \gamma$ degeneracy (produced by the stochastic nature of the AGN variability), the definition of the SF by itself, the light

curve sampling, and the fact that light curves with coverage of a few years are not a good representation of the whole variability behavior. We could notice in section 3.6.3 that when we performed the linear regression model with z_{spec} and L/L_{Edd} as independent sources, selecting only those sources from the not-C IV sample whose measured values of γ were in the range between 0.43 and 0.63, the measured scatter in the regression was reduced considerably. This confirms our assumption that the $A - \gamma$ degeneracy is one of the main contributors to the measured scatter in the regression models.

Possible interpretations of the inverse dependency of the amplitude of the variability with L/L_{Edd} are discussed by Wilhite et al. (2008), MacLeod et al. (2010), Simm et al. (2016), and Rakshit & Stalin (2017). One explanation can be that L/L_{Edd} is a proxy of the age of the AGN (e.g., Martini & Schneider 2003; Haas 2004; Hopkins et al. 2005). Sources with lower L/L_{Edd} can suffer from a dwindling of the fuel supply, as they become old, thus, the accretion flow can be more variable, producing larger amplitude in the variability. But, the time-scales of the amplitudes measured in this work are ~ 1 year, and therefore, it is unlikely that the variability amplitudes observed are given by variations in the external fuel supply, which requires much longer time-scales to be effective (10^5 to 10^7 days).

Other possible interpretation is that sources with higher L/L_{Edd} have hotter accretion disks, as predicted by classical accretion physics (Shakura & Sunyaev, 1973). For typical values of black hole mass and accretion rate, it is expected that the innermost part of the disk emits in the far UV. Because of its smaller size, this region is also the one showing the largest variability amplitude. For lower accretion rates however, the disk becomes cooler, and the innermost, most variable region will shift its emission from the UV to optical wavebands ($r_\lambda \propto M_{\text{BH}}^{2/3} (L/L_{\text{Edd}})^{1/3} \lambda^{4/3}$). This would be true regardless of whether the variation of the disk emission is produced by intrinsic processes or by reprocessing of highly variable X-ray emission by the disk surface. MacLeod et al. (2010) discarded this assumption because the time-scales (τ) that they measured were not in agreement with this scenario. However, they used DRW modelling to find τ , while it is now clear that DRW models cannot be used to properly describe the time-scales of typical ground-based light curves (see section 3.3.1).

A third possible explanation for the anti-correlation between A and L/L_{Edd} can be related with the positive correlation between L/L_{Edd} and the ratio of the UV/optical-to-X-ray flux (α_{ox}) reported by several studies (e.g., Shemmer et al. 2008; Grupe et al. 2010; Lusso et al. 2010; Jin et al. 2012). If the UV/optical variability is produced by reflection of the variable X-ray emission, then disks located in systems with higher α_{ox} values will receive fractionally less X-ray radiation, and therefore the amplitude of the variability detected in the UV/optical range will be small. On the other hand, for sources with lower α_{ox} , the disk will be irradiated with more X-ray light, and therefore we will detect higher UV/optical variability amplitudes. Kubota & Done (2018) developed a new spectral model for the SED of AGN that includes a hot corona, an inner warm optically thick Comptonising region and an outer disk. Considering this model, they studied the UV/optical variability resulting from the reprocessing of the rapidly variable X-ray flux. Their model predicts an anti-correlation between the amplitude of the variability and L/L_{Edd} . However their model also predicts a much lower amount of UV/optical variability than what is observed by our analysis and previous studies (e.g., MacLeod et al. 2010) at time-scales of 1 year or longer. This means that the model needs an extra source of UV/optical variability in order to explain the amplitudes

observed at long time-scales, as has been found by previous analyses (Krolik et al., 1991; Arévalo et al., 2008; Lira et al., 2015; Edelson et al., 2015). Therefore, the anti-correlation between A and L/L_{Edd} , cannot be solely explain by the correlation between L/L_{Edd} and (α_{ox}) .

In this work, we also found that the logarithmic gradient of the variability (γ) does not correlate significantly with any of the physical parameter studied, and that the general distribution of γ measured for our sample differs from the distribution of γ obtained for light curves simulated from a DRW process. We showed in sections 3.3.5 and 3.6.3 that 20,6% of the light curves have values of γ higher than 0.75, for which a DRW model is not appropriate to explain the variability. Kasliwal et al. (2015) and Smith et al. (2018) used *Kepler* light curves to study whether DRW modelling is sufficient to explain the variability of light curves with high cadence. They concluded that most of the *Kepler* AGN light curves analysed cannot be described by a simple DRW model. Smith et al. (2018) also proposed that it is possible that DRW modelling can be correct for ground-based quasar light curves, which in general study different time regimes than *Kepler*. We need larger samples of high cadence light curves, to see whether the results of Kasliwal et al. (2015) and Smith et al. (2018) are representative for the whole AGN population.

Chapter 4

The QUEST–La Silla AGN Variability Survey: selection of AGN candidates through optical variability

The material presented in this chapter is extracted from Sánchez-Sáez et al. 2019, submitted to *Astrophysical Journal Supplement* (ApJS).

4.1 Introduction

Traditionally, AGN selection follows the philosophy of finding regions in UV/optical/mid-IR color-color space in which AGN can be cleanly separated from stars and galaxies (e.g., Schmidt & Green 1983; Fan 1999; Richards et al. 2002, 2004; Lacy et al. 2004; Stern et al. 2005; Smith et al. 2005; Richards et al. 2009; Ross et al. 2012). However, some AGN populations have observed colors that fall outside the region occupied typically by bright, “blue” AGN, mimicking those of stars, such as type 2 or obscured AGN, broad absorption-line quasars (BAL-QSO), high-redshift quasars (high- z QSO) (Butler & Bloom, 2011; Palanque-Delabrouille et al., 2011, 2016), and low-luminosity AGN (LLAGN), whose colors can be highly contaminated by the emission from the host galaxy. Therefore, alternative methods to identify AGN candidates missed by traditional selection techniques are required, in order to obtain complete AGN samples. One promising selection method involves the use of variability techniques.

Butler & Bloom (2011) implemented a variability-based selection algorithm to classify high-redshift quasars in the Sloan Digital Sky Survey (SDSS; York et al. 2000) Stripe 82 field. They used damp random walk modelling (Kelly et al., 2009) to separate sources showing quasar-like variability from those with temporally uncorrelated variability. Particularly, they targeted unresolved sources with redshifts in the range $2.5 \leq z \leq 3$, where color-based selection of AGN is less efficient due to stellar contamination. Palanque-Delabrouille et al. (2011) used the variability structure function (e.g., Schmidt et al. 2010) to separate quasars,

variable stars, and non-variable stars, in the SDSS Stripe 82 data. They implemented a neural network algorithm that separates point-like objects by their structure function parameters. A similar technique has been used by the SDSS IV *the extended Baryon Oscillation Spectroscopic Survey* (eBOSS) team to select quasar candidates with $z > 2.1$ by variability (Myers et al., 2015; Palanque-Delabrouille et al., 2016). Peters et al. (2015) used color, variability, and astrometric data from SDSS to select point-like AGN candidates. They implemented a non-parametric Bayesian Classification Kernel Density Estimation (NBC KDE), to classify 35,820 type 1 quasar candidates in the Stripe 82 field. They tested various combinations of color and variability parameters, finding that using a combination of optical colors and variability parameters improves quasar classification efficiency and completeness over the use of colors alone. More recently, Tie et al. (2017) used data from the supernova fields of the Dark Energy Survey (DES; Abbott et al. 2018) to select quasars by combining color and variability selection methods. All these previous studies have shown the potential of selecting AGN candidates through variability analyses, demonstrating that variability-based techniques can increase considerably the number of AGN candidates in the redshift range where the colors of stars are similar to those of AGN. Less clear is how deep into the low-luminosity and obscured AGN populations they can probe. This is of substantial importance given the eventual mismatch that X-ray and MIR surveys will have compared with LSST.

In this Chapter we present our variability-based technique to select AGN candidates using data from the QUEST-La Silla AGN variability survey (Cartier et al., 2015). In this work we aim to detect wider sets of AGN populations. Particularly, we expect to detect sources that show clear signatures of a nonstellar continuum emitting process in their centers, with emission lines broader than ~ 1800 km/s, regardless of their luminosity or shape in the QUEST images. We do not expect to detect many type 2 or obscured AGN candidates, since our technique requires the detection of a variable continuum component. Variability features, like the structure function, have been used to characterize the variable sources (e.g., Cartier et al. 2015; Sánchez et al. 2017). We then used a Random Forest algorithm to classify our objects as either AGN or non-AGN. We tested three classifiers, one that includes only variability features, and two that include optical colors and variability features. The main difference of our selection technique with previous variability-based AGN selection methods is the use of light curves with higher cadence, and the exclusion of any morphology indicator. Hereby, we expect to detect more low-redshift AGN and LLAGN candidates than previous analyses. For some of our candidates we have obtained optical spectra to confirm their nature. Four of the fields observed by the QUEST-La Silla AGN variability survey correspond to the LSST Deep Drilling Fields (DDFs), whose expected cadence will be similar to the nightly cadence used by the QUEST-La Silla AGN variability survey (but extending the time baseline to 10 years)¹. The QUEST-La Silla AGN variability survey is an important test bed to study AGN selection in time-domain surveys, like LSST, or the Zwicky Transient Facility (ZTF, Bellm 2014), which has a depth similar to the QUEST-La Silla AGN variability survey.

The Chapter is organized as follows. In Section 4.2 we describe the light curve construction procedure. In Section 4.3 we describe the Random Forest algorithm, the variability features, and the labeled set used for the selection. We also discuss the performance of our Random Forest classifiers, and we comment about the selected candidates. In Section 4.4 we provide the results on confirming the nature of some of our candidates by using public data and

¹<https://www.lsst.org/scientists/survey-design/ddf>

spectroscopic follow-up. In Section 4.5, we provide a comparison of our results with previous works. Finally, in Section 4.6 we summarize the main results.

4.2 Light curve construction

We performed aperture photometry of the QUEST–La Silla images using SExtractor (Bertin & Arnouts, 1996), with the same optimal aperture found by Cartier et al. (2015) for the QUEST camera ($\sim 6''.18$). We then constructed light curves for all the sources from the SDSS (Gunn et al., 1998; Doi et al., 2010) and DES (Abbott et al., 2018) photometric catalogs with detections in the QUEST–La Silla data, using the same methodology as in Cartier et al. (2015). In summary, we constructed light curves by cross-matching the SDSS and DES catalogs with every QUEST–La Silla catalog, that we generated for each observation, for which we knew their associated Julian dates, using a radius of $1''$. We then constructed light curves for each source, keeping only those epochs where the SExtractor FLAG parameter was equal to zero, to prevent false detection of variability due to bad photometry. Finally, we only saved those light curves with more than three epochs. From the SDSS catalog, we could obtain single-epoch photometry of every source in the COSMOS, XMM–LSS, and Stripe 82 fields in the u , g , r , i , and z bands, and from the DES catalog we obtained single-epoch photometry in the g , r , i , and z bands for the ELAIS–S1 and ECDF–S fields.

We decided to bin our light curves using three-day bins, in order to reduce the noise in our light curves produced by several factors, including changes in atmospheric conditions and the relatively low cosmetic quality of the QUEST CCD camera chips. This might affect the detection of variability of sources with short time-scale variations, like some variable stars (e.g., RR Lyrae or Cepheid stars), however, we do not expect to detect many variable stars in the QUEST–La Silla fields (e.g., Medina et al. 2018). Moreover, in this work we are focused in the detection of sources with long time-scale variations (with time-scales of months or years), thus the three-day binning does not affect our detection of AGN.

In this work, we excluded the Stripe 82 field, since it is a crowded field, and requires point spread function (PSF) photometry. We generated a total of 277629 light curves for sources located in the COSMOS, ECDF–S, ELAIS–S1, and XMM–LSS fields. In order to have statistically significant variability features of the sources, we decided to include in our analysis only those light curves with at least 40 epochs and a length greater than or equal to 200 days, after the three days binning was applied (hereafter “well-sampled” light curves). There are 208,583 well-sampled light curves in the four fields. The median, mean and standard deviation of the number of epochs of each light curve are 118, 119.3, and 47.2, respectively; and the median, mean and standard deviation of the total length of each light curve are 1283.7, 1306.4, and 254.3, respectively. In Table 4.1 we summarize the total number of light curves and the number of well-sampled light curves in each field.

Table 4.1: Number of light curves per field.

Field	total light curves	well-sampled light curves
COSMOS	68514	45323
XMM-LSS	104962	82697
Elais-S1	49504	38106
ECDF-S	54649	42457
Total	277629	208583

4.3 Selection of AGN candidates

We implemented a supervised automatic classification using a Random Forest algorithm (RF; Breiman 2001) to classify our 208,583 objects with well-sampled light curves as either AGN or non-AGN according to their variability features and optical colors. We did not include a morphological parameter during the classification (e.g., SExtractor CLASS_STAR parameter), in order to be able to detect sources with AGN-like variability with extended shapes. We tested three classifiers: one that includes only variability features, and two that include optical colors and variability features. In the following sub-sections we describe the selection methodology, the features used in our analysis, and the results of the classification for sources from the QUEST-La Silla survey.

4.3.1 Random Forests

A decision tree is a hierarchical structure that performs successive partitions on the data, each of them according to a certain criteria, such as a cut-off value in one of the descriptors or features. In this way, the data are divided into smaller and smaller subsets as the tree goes deeper, until it reaches the leaves of the tree. Each of the leaves is associated with a single class. A given class, however, may be associated with several leaves. Thus the elements that fall on any of the leaves corresponding to a particular class will be classified as belonging to that class.

A RF algorithm consist of a collection of single decision trees, where each tree is trained using a random sub-set of sources, sampled with repetition, from a training set (a set of objects with known classification, selected from a labeled set), and a random selection of features. The final classification function of the algorithm weighs each of these results according to the size of the sub-set used by each tree, and generates an average score, which can then be interpreted as the probability that the input element belongs to a certain class (predicted class probability, P_{RF}). Then, the classifier is validated using a sub-set of the labeled set that was not used to train (the test set). Finally, a prediction is made on the unlabeled data. RF has several advantages, it can handle thousands of features, it provides a ranking of feature importance during the classification, it does not need to scale the feature values to the same “units”, it handles numerous objects, and it is easily parallelizable.

For the selection of AGN candidates we used the *scikit-learn*² Python package implementation of RF. We performed a hyperparameter selection procedure in order to obtain the optimal values for the RF classifier, by means of a K-Fold Cross-Validation procedure³ (with $k = 5$ folds) and using the “accuracy” (see its definition in Section 4.3.4) as the target score to optimize. This hyperparameter selection procedure was executed as part of the model training phase (i.e. on the training set). In this procedure, the training set is divided in k folds, using $k - 1$ of them to compute the RF model, and testing it in the remaining data (the validation set). This is done k times, using every time a different fold as validation set. The parameters considered in this cross-validated search include the number of trees in the forest, and the number of features to consider when looking for the best split in a tree. In order to take into account the class imbalance in the classification process, we initialized the class weight hyperparameter as “*balanced_subsample*”.

The variability features used by the RF classifier are described in the following section (4.3.2), and are listed in Table 4.2. We trained the RF classifier using a labeled set of type 1 AGN and stars with spectroscopic classification from SDSS and with well-sampled light curves from the QUEST–La Silla survey (see Section 4.3.3). During the RF classifier training, we used 70% of the labeled set as a training set, and then we tested the performance of the classifier using the remaining 30% of the labeled set (the test set), as normally done during supervised learning procedures. We then applied the trained RF classifier to our unlabeled set, composed by our 208,583 sources with well-sampled QUEST–La Silla light curves, to classify them as either AGN or non-AGN. As a result, we obtain a predicted class and the predicted class probability (P_{RF}) associated to each source of the unlabeled set.

4.3.2 Variability features

In order to have a complete description of the variability of our sources, we used several variability features. Following the same approach of Sánchez et al. (2017) and Sánchez-Sáez et al. (2018), we used two parameters related to the amplitude of the variability, P_{var} and the excess variance (σ_{rms}), and one parameter that describes the shape of the variability between two observations separated by a given time, the structure function (SF).

In particular, P_{var} (see Sánchez et al. 2017 and references therein) corresponds to the probability that the source is intrinsically variable; it considers the χ^2 of the light curve, and calculates the probability $P_{var} = P(\chi^2)$ that a χ^2 lower or equal to the observed value could occur by chance for an intrinsically non-variable source.

σ_{rms} is a measure of the intrinsic variability amplitude (see Sánchez et al. 2017 and references therein), and it is calculated as $\sigma_{\text{rms}}^2 = (\sigma_{LC}^2 - \bar{\sigma}_m^2) / \bar{m}^2$, where σ_{LC} is the standard deviation of the light curve, $\bar{\sigma}_m$ is the mean photometric error, and \bar{m} is the mean magnitude.

The SF (e.g., Schmidt et al. 2010) is the average variability amplitude between two observations separated by a given time (τ), and it can be modelled as a power-law: $\text{SF}(\tau) =$

²<http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

³https://scikit-learn.org/stable/modules/cross_validation.html

$A_{\text{SF}} \left(\frac{\tau}{1\text{yr}} \right)^{\gamma_{\text{SF}}}$, where A_{SF} corresponds to the amplitude of the variability at 1 year time-scale, and γ_{SF} is the logarithmic gradient of this change in magnitude.

We also used some variability features from the Feature Analysis for Time Series (FATS; Nun et al. 2015) Python package, related with the amplitude of the variability (e.g., the mean variance and the percent amplitude) and the structure of the light curve (e.g., the linear trend and the auto-correlation function length), as well as the period of the Lomb-Scargle periodogram (VanderPlas, 2018), derived by using the AstroML module for Python (VanderPlas et al., 2012). A list of all the variability features used in this work is shown in Table 4.2, together with a brief description of each feature and its reference.

4.3.3 Labeled set

To train our RF classifier we need a labeled set, which has to be representative of the populations that we want to classify. Since in this analysis we only include extragalactic fields, we do not expect to detect a high fraction of variable stars, because of their low density at high Galactic latitudes (e.g., RR Lyrae or Cepheid stars, see Medina et al. 2018 and references therein). Moreover, since we implemented a three-day binning to our light curves, the detection of variable signals with short time-scales is not possible. Therefore, any variable star with a short period will have a light curve that will not be very different from a non-variable star. Only variable stars with long periods would be detectable using our QUEST-La Silla light curves. We cross-matched the positions of the 208,583 sources with well-sampled light curves with the General catalogue of variable stars (Version GCVS 5.1, Samus' et al. 2017), which provides a detailed compilation of catalogs of variable stars in the Galaxy. We found that only three known variable stars are present in our data, one RR Lyrae and two cataclysmic variables. Therefore, we did not include variable stars in our RF classifiers.

In this analysis, galaxies are not included in the labeled set, since in general their variability and color properties will be similar to those of stars, unless they host an AGN (which might not have been previously detected). Therefore, we constructed a labeled set composed by stars and type 1 AGN (i.e. AGN with broad permitted emission lines). We decided to include only type 1 AGN since we want to characterize properly the variability of the optical continuum emission, which cannot be detected in most type 2 or obscured AGN.

Three of our fields (COSMOS, Stripe 82, and XMM-LSS) have spectroscopic information from SDSS. We constructed light curves for sources with spectral classification from the SDSS-DR14 database (Abolfathi et al., 2018). There are 3,313 type 1 AGN and 3,332 stars with at least three epochs in the QUEST-La Silla light curves, and 2,405 type 1 AGN and 2,608 stars with well-sampled light curves. We considered the sources with well-sampled light curves to define a labeled set for the RF classifier training. As mentioned in Section 4.3.1, 30% of the labeled set was used as a test set and 70% as training set for the RF modelling. Figure 4.1 provides examples of QUEST-La Silla light curves for four sources of the labeled set.

It is well-known that a fraction of AGN are misclassified in the SDSS databases, therefore

Table 4.2: List of features.

Feature	Description	Reference
P_{var}	Probability that the source is intrinsically variable	McLaughlin et al. (1996)
σ_{rms}	Measure of the intrinsic variability amplitude.	Allevato et al. (2013)
A_{SF}	Amplitude of the variability at 1 year, derived from the SF	Schmidt et al. (2010)
γ_{SF}	Logarithmic gradient of the change in magnitude, derived from the SF	Schmidt et al. (2010)
Std*	Standard deviation of the light curve (σ_{LC})	Nun et al. (2015)
Meanvariance*	Ratio of the standard deviation to the mean magnitude (σ_{LC}/\bar{m})	Nun et al. (2015)
MedianBRP*	Fraction of photometric points within amplitude/10 of the median magnitude	Richards et al. (2011)
Autocor-length*	Lag value where the auto-correlation becomes smaller than e^{-1}	Kim et al. (2011)
StetsonK*	A robust kurtosis measure	Kim et al. (2011)
η^{e*}	Ratio of the mean of the square of successive differences to the variance of data points	Kim et al. (2014)
PercentAmp*	Largest percentage difference between either the max or min magnitude and the median	Richards et al. (2011)
Coni*	number of three consecutive data points that are brighter or fainter than $2\sigma_{LC}$	Kim et al. (2011)
LinearTrend*	Slope of a linear fit to the light curve	Richards et al. (2011)
Beyond1Std*	Percentage of points beyond one σ_{LC} from the mean	Richards et al. (2011)
Q31*	Difference between the third quartile and the first quartile of a light curve	Kim et al. (2014)
PeriodLS	Period from the Lomb-Scargle periodogram	VanderPlas (2018)

Note. (*) Features from FATS

we cross-match our labeled sample with the Million Quasars Catalog (MILLIQUAS v5.7 update⁴, 7 January 2019, Flesch 2015, see Section 4.4.1 for further details), in order to estimate the fraction of AGN in the labeled set that are not, in fact, AGN. There are 57 AGN in the labeled set, with well-sampled light curves, that are not present in MILLIQUAS, which correspond to the 2.4% of the AGN in the labeled set. 21 of these 57 sources are classified as variable according to their variability features, thus we can estimate that less than 2% of the AGN in the labeled set are misclassified as AGN.

In Figure 4.2 we show three color-color diagrams of the labeled set: $u - g$ vs $g - r$, $g - r$ vs $r - i$, and $r - i$ vs $i - z$. As a reference, we mark the regions of the $u - g$ versus $g - r$ color-color diagram dominated by a particular type of source, from Sesar et al. (2007). We can see that a high fraction of the AGN in the labeled set are located in a region of the $u - g$ vs $g - r$ diagram where low-redshift, luminous AGN (II) are the dominant population, which corresponds to 78.8% of the AGN in the labeled set. This corresponds to the classical color-color selection of “blue” AGN. Moreover, we can see in the figure that several high-redshift ($z_{spec} > 2.5$) AGN in the labeled sample are located in the region where high-redshift luminous AGN are the dominant component (VI), as expected, but a non negligible fraction is located in other regions, where binary stars or cool dwarf stars (III), RR Lyrae stars, and main sequence stars or “stellar locus” (V) are the dominant population.

On the other hand, we can see in the right panel of Figure 4.2 that most of the AGN in the labeled set are located in a region where $r - i \lesssim 0.7$ and $i - z \lesssim 0.8$, cleanly isolating a sub-population of cool dwarf stars. This can be understood considering that stellar colors become monotonically redder as the effective temperature decreases (Covey et al., 2007), thus we normally observe a high concentration of cool dwarf stars in a region around $g - r \sim 1.5$, with $r - i \gtrsim 0.8$. Besides, extragalactic sources (i.e., galaxies and AGN) are normally located in regions of the color-color space with $r - i \lesssim 1.0$ (e.g., Rahman et al. 2016), since their integrated emission typically has a low contribution from cool dwarf stars. Therefore, we can use $r - i$ and $r - z$ colors to separate AGN from cool dwarf stars. The separation of AGN and stars from the general “stellar locus” is more complicated if we only use optical colors, since there is a high overlap between these two populations in the different color-color diagrams, particularly in the $u - g$ vs $g - r$ diagram, as already discussed. Thus, including variability information in the selection of AGN candidates will be extremely useful to improve AGN selection.

4.3.4 Performance of the Random Forest classifier

We tested three different RF classifiers. The first one includes only variability features (hereafter RF1). The second one includes variability features and the $r - i$ and $i - z$ colors (hereafter RF2). And the third classifier includes variability features and the $g - r$, $r - i$, and $i - z$ optical colors (hereafter RF3). We exclude $u - g$ since we do not have photometry in the u band for the Elais-S1 and ECDF-S fields, and because our labeled set does not cover properly the $u - g$ space, compared to the unlabeled set, thus including it might produce poor results. For this reason we did not test a pure color selection, as u band is highly

⁴<https://heasarc.gsfc.nasa.gov/W3Browse/galaxycatalog/milliquas.html>

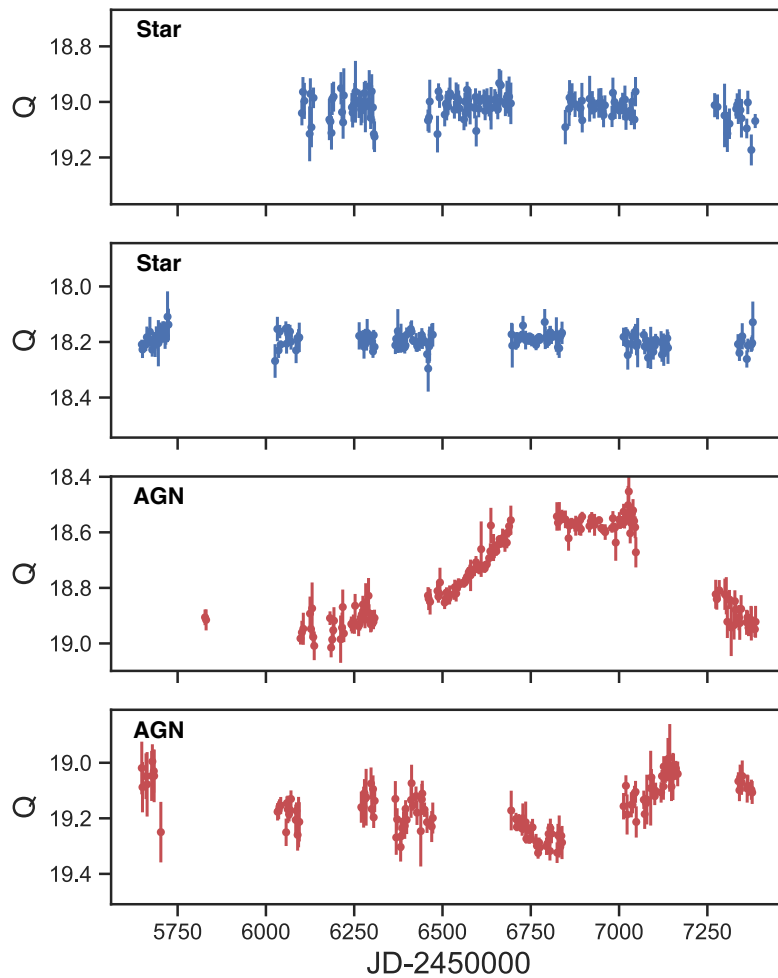


Figure 4.1: Example of four light curves from the QUEST–La Silla labeled set: two stars (blue dots, top panels) and two AGN (red dots, bottom panels).

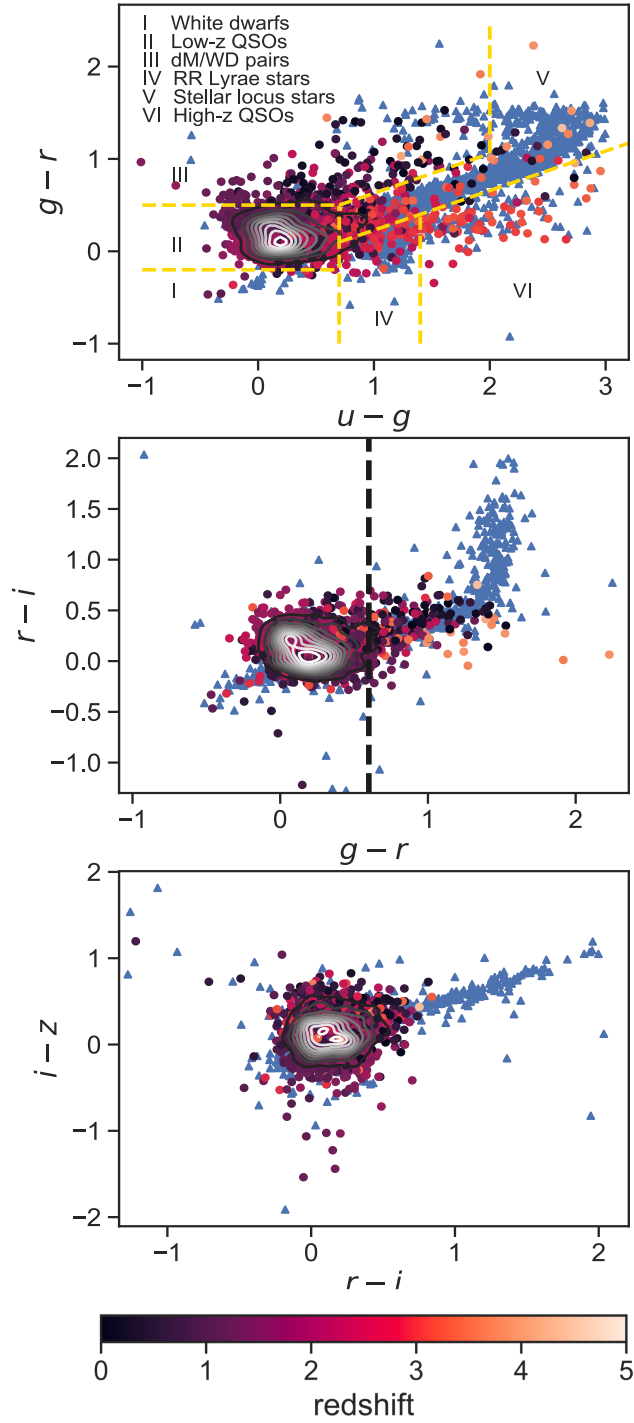


Figure 4.2: Color-color diagrams of the labeled set. In the top panel we show $u - g$ versus $g - r$, in the middle panel $g - r$ vs $r - i$, and in the bottom panel $r - i$ vs $i - z$. The stars are represented by blue triangles, and the AGN are represented by circles whose colors depend on the redshift of each source. The contour plots show the distribution of AGN. In the top panel we show with yellow dashed lines the division used in Sesar et al. (2007) to identify regions of the $u - g$ versus $g - r$ diagram dominated by a particular type of source. In the middle panel, the black dashed line shows the position where $g - r = 0.6$.

discriminating for selecting AGN (e.g., Richards et al. 2002, 2009; Ross et al. 2012).

As can be seen, the difference between RF2 and RF3 is the exclusion of the color $g - r$ in RF2. As mentioned in the previous section, the $r - i$ and $i - z$ colors can easily separate cool dwarf stars and AGN. However, the separation of AGN and stars from the stellar locus is difficult when we use optical colors, particularly for the case of $u - g$ and $g - r$. Thus, with RF2 we can test whether avoiding the use of $g - r$ can improve the detection of redder AGN populations. On the other hand, RF1 excludes optical colors, thus the amount of information used by this classifier is lower compared to RF2 and RF3. Optical colors have been exhaustively used in the literature for the selection of AGN candidates (e.g., Fan 1999; Richards et al. 2002, 2004; Smith et al. 2005; Richards et al. 2009; Kirkpatrick et al. 2011; Bovy et al. 2011; Ross et al. 2012), thus with RF1 we can test whether single-band variability-based techniques can provide results as competitive as the ones obtained using optical colors.

As mentioned in Section 4.3.1, we trained each classifier using 70% of the labeled set as a training set, and the remaining 30% of the labeled set as a test set. The selection of the training and test sets is done randomly, using the “train_test_split” procedure of *scikit-learn*. The labeled set, by definition, has the same limiting magnitude of the QUEST-La Silla images ($r \sim 21$), therefore the training and test sets have limiting magnitudes of $r \sim 21$.

Since we are interested in selecting only AGN candidates, for the rest of the analysis we will refer to stars, and any source that is not an AGN as non-AGN.

RF1: selection of AGN based solely on variability

Our first RF classifier (RF1) includes only variability features. We show the results from this classifier using a confusion matrix, which is shown in Figure 4.3 (see its RF1 results). It can be seen that AGN (true positives) are in general well classified, and also that the fraction of non-AGN classified as AGN (false positives) is very low.

We also computed the following scores to assess our classifiers: accuracy (A), precision (P), recall (R), and F1. These scores are defined by means of the True Positives (TPs: known AGN classified as AGN by the RF classifier), the False Positives (FPs: known non-AGN classified as AGN), the True Negatives (TNs: known non-AGN classified as non-AGN), and the False Negatives (FNs: known AGN classified as non-AGN):

$$\begin{aligned}
 A &= \frac{TPs + TNs}{Total\ Sample} \\
 P &= \frac{TPs}{TPs + FPs} \\
 R &= \frac{TPs}{TPs + FNs} \\
 F1 &= 2 \times \frac{P \times R}{P + R}
 \end{aligned}
 \tag{4.1}$$

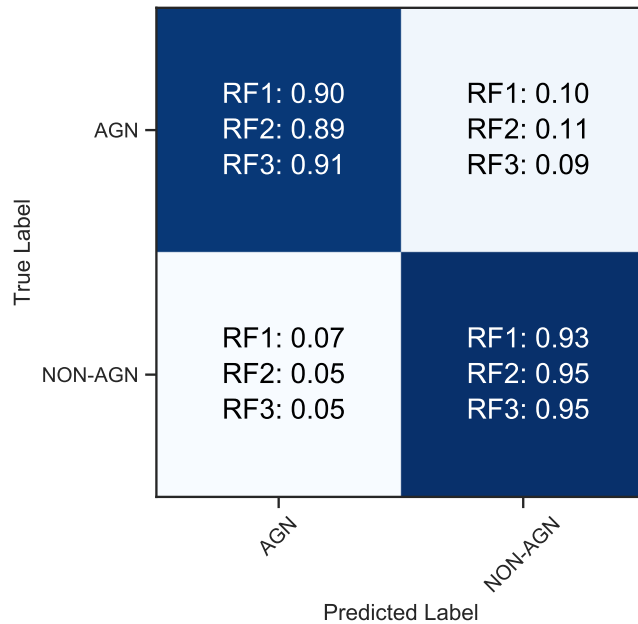


Figure 4.3: Confusion matrix from testing the RF1, RF2, and RF3 in the test set. True Label represent the classification done from SDSS spectra, and Predicted Label is the outcome of each classifier.

Table 4.3: Scores measured in the test set for each classifier

Score	RF1	RF2	RF3
Accuracy	0.916	0.923	0.931
Precision	0.909	0.909	0.921
Recall	0.933	0.950	0.951
F1	0.921	0.930	0.936

Table 4.3 shows the computed scores for the RF1 classifier. From these scores, and from the confusion matrix, we can say that RF1 presents a low fraction of False Positives, thus, the sample of predicted AGN has low contamination from non-AGN. However, we tend to miss a fraction of real AGN ($\sim 10\%$). This results from the difficulty of detecting a variable signal from AGN with low amplitude variability, and since we are only considering variability properties for the classification, they could be classified as non-AGN.

It is important to consider that we are testing the RF1 classifier in a sample of AGN selected mostly by means of their optical colors, and since we are only considering variability features in our selection, the confusion matrix and the different scores, obtained from our labeled sample, might not necessarily be an optimal prediction of the performance of our method in the unlabeled sample.

One of the advantages of the RF classification is that we can easily know the feature importance, since it provides a ranking score for each feature, or how well every feature separates the two classes. In the first columns of Table 4.4, we provide the list of features, ordered by importance (rank value), for the RF1 classifier. It can be seen that the four

Table 4.4: Feature importance for each classifier.

RF1		RF2		RF3	
Feature	Rank	Feature	Rank	Feature	Rank
A_{SF}	0.197	A_{SF}	0.209	A_{SF}	0.189
σ_{rms}	0.139	σ_{rms}	0.149	σ_{rms}	0.142
Meanvariance	0.127	Q31	0.102	Q31	0.113
Q31	0.111	P_{var}	0.093	Meanvariance	0.099
P_{var}	0.095	Std	0.088	P_{var}	0.095
Std	0.090	Meanvariance	0.086	Std	0.074
PercentAmp	0.040	PercentAmp	0.045	Autocor-length	0.042
γ_{SF}	0.036	Autocor-length	0.035	PercentAmp	0.039
Autocor-length	0.033	γ_{SF}	0.031	$g - r$	0.031
MedianBRP	0.025	$r - i$	0.028	γ_{SF}	0.029
LinearTrend	0.023	MedianBRP	0.021	$r - i$	0.023
PeriodLS	0.023	PeriodLS	0.020	MedianBRP	0.020
η^e	0.023	η^e	0.019	η^e	0.020
Beyond1Std	0.019	Beyond1Std	0.019	$i - z$	0.019
StetsonK	0.018	LinearTrend	0.019	LinearTrend	0.018
Con	0.002	$i - z$	0.019	PeriodLS	0.017
		StetsonK	0.014	Beyond1Std	0.017
		Con	0.002	StetsonK	0.012
				Con	0.002

most important features are the amplitude of the structure function, the excess variance, the Meanvariance, and Q31. In Figure 4.4, we show the distribution of the A_{SF} and Q31 features for the labeled set. We highlight using black dots those AGN classified as variable, according to the definition proposed by Sánchez et al. (2017), where a source is classified as variable when its light curve satisfies $P_{\text{var}} \geq 0.95$ and $(\sigma_{\text{rms}}^2 - \text{err}(\sigma_{\text{rms}}^2)) > 0$. From the figure, it can be seen that AGN and non-AGN are separated by these two features, with A_{SF} providing a much stronger division than Q31, as there is substantial source overlap between the two classes with the latter indicator. It can be also seen that the majority of the AGN with low variability amplitude are classified as non-variable.

RF2: selection of AGN based on variability and $r - i$, and $i - z$ optical colors

Our second RF classifier (RF2) includes variability features and the $r - i$ and $i - z$ colors. Figure 4.3 shows the confusion matrix for RF2 (see its RF2 results). In this case, the confusion matrix is similar to the confusion matrix of RF1, however, in the case of RF2 we have a slightly cleaner population of AGN candidates (i.e. the fraction of False Positives is lower). The accuracy, precision, recall, and F1 scores are given in Table 4.3. There are not significant differences between the score values of RF1 and RF2.

In Table 4.4 we list the ranking of features for the RF2 classifier. There are not significant

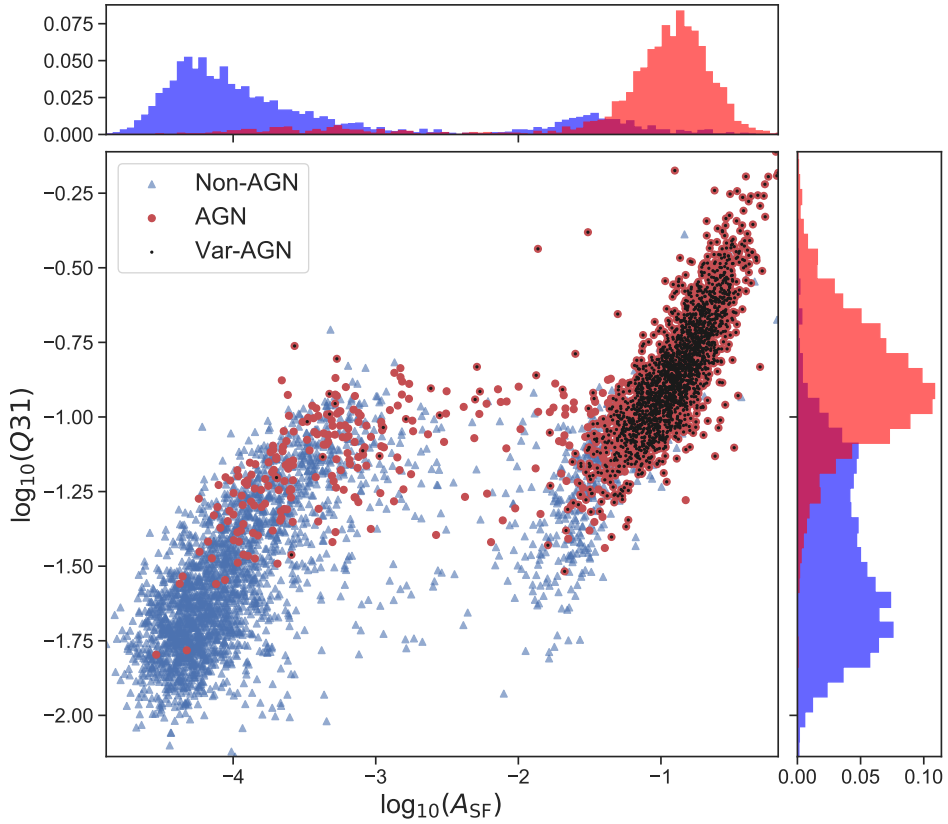


Figure 4.4: Distribution of the A_{SF} and Q31 features for the labeled set. Blue triangles correspond to non-AGN, and red circles correspond to AGN. We mark with black dots those AGN classified as variable, according to the definition used in Sánchez et al. (2017).

differences when compared to RF1, and notably we found that variability features are more relevant for AGN selection than the $r - i$ and $i - z$ colors. In this case, the most important features are A_{SF} , σ_{rms} , Q31, and P_{var} . We also found that the $r - i$ color seems to be more relevant than the $i - z$ to classify our sources.

RF3: selection of AGN based on variability and $g - r$, $r - i$, and $i - z$ optical colors

Our third RF classifier (RF3) includes variability features and the $g - r$, $r - i$, and $i - z$ colors. Figure 4.3 shows the confusion matrix for RF3 (see its RF3 results). In this case, the fraction of True Positives (True AGN classified as AGN) is slightly higher than those of RF1 and RF2. However, we must consider that most of the AGN in the labeled set have been selected by means of their optical colors, which might explain the improvement of the results over the test set compared with the RF2 classifier.

The scores for RF3 are listed in Table 4.3. In comparison to RF1 and RF2, the scores are slightly higher, particularly the precision. In Table 4.4 we list the ranking of features for the RF3 classifier. In this case the most important features are A_{SF} , σ_{rms} , Q31, and the Meanvariance. The most important color is $g - r$, which is expected due the distribution of non-AGN and AGN in Figure 4.2. It is well-known that much of the discriminating power for selecting unresolved AGN is in $u - g$ (e.g., Braccetti et al. 1970), which is in agreement with our finding that $r - i$ and $i - z$ colors are not as relevant for AGN selection.

4.3.5 AGN candidates from QUEST–La Silla

We applied the trained RF1, RF2, and RF3 classifiers to our unlabeled well-sampled set of 208,583 light curves. In order to improve the purity of our selection, we considered the predicted class probability P_{RF} (computed as the mean predicted class probabilities of the trees in the forest) to select the final set of AGN candidates. We defined two samples of AGN candidates: a) the full-AGN sample, consisting of all sources classified as AGN by the RF classifier ($P_{\text{RF}} \geq 0.5$), and b) the hp-AGN sample, consisting in sources that have a high probability ($P_{\text{RF}} \geq 0.8$) of being an AGN based on the RF classifier. In Table 4.5 we provide a summary with the number of sources classified as AGN in both samples, for each classifier. For the case of the RF1 classifier, there are 17,120 sources in the full-AGN sample, and 5,941 sources in the hp-AGN sample. For the RF2 classifier there are 15,100 sources in the full-AGN sample, and 5,252 sources in the hp-AGN sample. Finally, for RF3, there are 13,810 sources in the full-AGN sample, and 4,482 sources in the hp-AGN sample. There are 4,054 candidates in common between the RF1, RF2, and RF3 hp-AGN samples. For the rest of the analysis we will only consider the hp-AGN samples of each classifier.

Figure 4.5 shows the $g - r$ vs $r - i$ color-color diagram of the unlabeled set, and the hp-AGN samples for the RF1, RF2, and RF3. Comparing with Figure 4.2, we can see that several of our AGN candidates are located in regions of the color-color space where AGN are not normally found, particularly for the case of RF1. The main difference between the candidates of RF1 and the rest of the classifiers, is the exclusion of sources in the color-color

Table 4.5: Number of AGN candidates per field, for each classifier

Field	RF1		RF2		RF3	
	full-AGN	hp-AGN	full-AGN	hp-AGN	full-AGN	hp-AGN
COSMOS	3,968	1,503	3,562	1,201	3,424	1,018
XMM-LSS	6,441	2,374	5,774	2,106	5,516	1,879
Elais-S1	3,374	988	2,936	942	2,441	777
ECDF-S	3,337	1,076	2,828	1,003	2,429	808
Total	17,120	5,941	15,100	5,252	13,810	4,482

region where we normally find cool stars. For example, there are 4,890 candidates in common between RF1 and RF2, and 1,051 RF1 candidates that are not candidates for RF2. 54.1% of the former ones have $r - i > 0.7$, where we expect to find mostly cool stars. The main differences between the candidates of RF2 and RF3, is the exclusion of the redder candidates in the case of RF3 ($g - r \gtrsim 1.0$). There are 4,178 candidates in common between RF2 and RF3, and 1,074 candidates in RF2 that are not candidates for RF3, with 70.5% of these having $g - r > 1.0$.

In the top panel of Figure 4.5 we also show a selection of AGN candidates observed during spectroscopic follow-up (see Section 4.4.2). In the middle and bottom panels of the figure we show the position in the $g - r$ vs $r - i$ diagram of four candidates located in different positions of the stellar locus, marked with letters (ABCD), and in Figure 4.6 we show their light curves, where it can be seen that they are clearly variable. These four sources are selected as candidates by RF1, two are selected as candidates by RF2 (A and B), and none of them is selected as a candidate by RF3.

4.4 Confirmation of AGN candidates

In the following sub-sections, we aim to confirm the nature of our candidates. In Section 4.4.1, we use ancillary data to confirm the nature of our AGN candidates. In Section 4.4.2 we show the results of our spectroscopic follow-up campaign, conducted between December 2016 and September 2018, to test the efficiency of our selection method.

We are particularly interested in identifying the nature of sources located in positions of the color-color space dominated by stars (e.g., in the stellar locus). We divided our high probability candidates according to their $g - r$ colors, we avoid $u - g$ since the u band is not available for all the fields. We define the blue sub-sample as the one composed of sources with $g - r \leq 0.6$ and the red sub-sample as that composed of sources with $g - r > 0.6$. As can be seen in Figure 4.2, most of the AGN in the labeled set have $g - r \leq 0.6$.

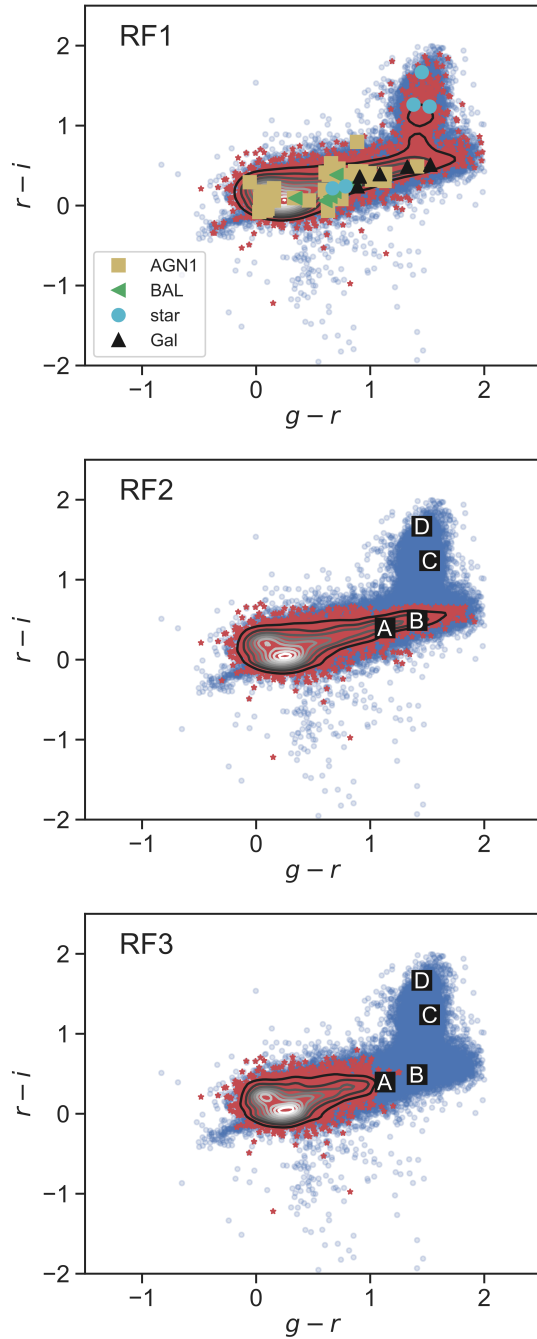


Figure 4.5: $g-r$ vs $r-i$ color-color diagrams of the unlabeled set (blue circles), and the hp-AGN sample (red stars) for the RF1 (top panel), RF2 (middle panel), and RF3 (bottom panel). The contour plots show the distribution of the hp-AGN samples for each classifier. In the top panel, we show the position of candidates observed during spectroscopic follow-up, differentiating between type 1 AGN (AGN1, yellow squares), BAL-QSO (BAL, red triangles), stars (cyan circles), and galaxies (Gal, black triangles). In the middle and bottom panels we show with letters and black squares the position of a selection of observed candidates located in the stellar locus.

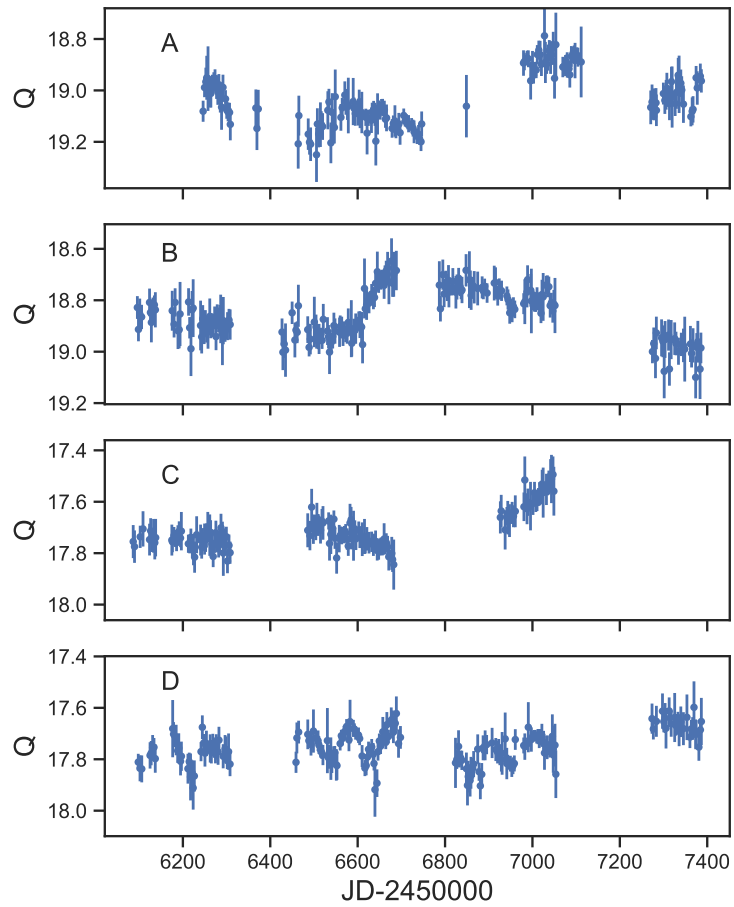


Figure 4.6: Light curves of some RF1 candidates located in the stellar locus, observed during the spectroscopic follow-up campaign, shown in the top-right and bottom-left panels of Fig. 4.5. A and B are classified as type 1 AGN, C and D are classified as M-type stars.

4.4.1 Confirmation by ancillary data

MILLIQUAS (v5.7 update, 7 January 2019, Flesch 2015) provides a very complete compendium of known AGN (both type 1 and type 2), from the literature, including the last data release of SDSS (SDSS-DR15), and several recent XMM–Newton, Swift, and Chandra catalogs (e.g., Evans et al. 2014; Marchesi et al. 2016a; Maitra et al. 2019). It also includes a list of high–confidence AGN candidates from different sources like AllWISE (Secrest et al., 2015).

We used MILLIQUAS to confirm the nature of our candidates. We cross–matched MILLIQUAS with the coordinates of our well–sampled light curves, using a radius of $1''$. There are 3,524 sources in the well–sampled sample with identifications in MILLIQUAS. For the case of the RF1 classifier, there are 2,358 (66.9%) of these sources in the hp–AGN sample, 2,757 (78.2%) in the full–AGN sample, and 767 (21.8%) sources classified as non–AGN. For the RF2 classifier, there are 2,366 (67.1%) sources in the hp–AGN sample, 2775 (78.7%) in the full–AGN sample, and 749 (21.3%) sources classified as non–AGN. Finally, for the RF3 classifier, there are 2,348 (66.7%) sources in the hp–AGN sample, 2,769 sources in the full–AGN sample (78.6%), and 755 (21.4%) sources classified as non–AGN. From these results we can say that $\sim 21\%$ of the sources are misclassified as non–AGN when we include variability features in the selection. Thus, we can conclude that when we include variability features in our selection, we obtain a completeness of $\sim 79\%$.

We plot in the top panel of Figure 4.7 the distribution of the A_{SF} variability feature, for sources in MILLIQUAS and QUEST–La Silla belonging to the RF1 hp–AGN sample, the RF1 full–AGN sample, and sources classified as non–AGN by RF1 (but classified as AGN by MILLIQUAS). It can be seen that the main difference between the sources classified as AGN and non–AGN is the value of the variability amplitude at one year, .i.e. we are not detecting variability for the sources classified as non–AGN.

In the middle panel of Figure 4.7 we compare the distribution of the mean Q magnitude (determined from their light curves) of sources from the RF1 hp–AGN sample that have and do not have detection in MILLIQUAS, and in the bottom panel we compare their A_{SF} distributions. It can be seen that the sources with detection in MILLIQUAS are in general brighter than the sources without detection in MILLIQUAS, however the amplitude of the variability is lower for the sources without detection in MILLIQUAS. It is important to note that only a small fraction ($\lesssim 3\%$) of the AGN in MILLIQUAS are classified as host–dominated (i.e., they appear extended in imaging). This demonstrates that variability can be used to augment AGN selection to include objects that are extended as well as point sources.

In Table 4.6 we show the number of hp–AGN candidates confirmed using MILLIQUAS, dividing the hp–AGN samples into red and blue sub–samples, as already described. It can be seen that most of the confirmed sources are in the blue sub–sample, with 52% of the candidates in this sample confirmed for RF1, 52.4% for RF2, and 52.1% for RF3. For the case of the red sub–sample, 5.3% of the candidates from RF1 are confirmed using MILLIQUAS, 7.1% for RF2, and 12% for RF3. Besides, there are 354 AGN listed as candidates in MILLIQUAS for RF1, 356 for RF2, and 345 for RF3. Most of them are candidates from WISE (Secrest et al., 2015). This lack of confirmed red candidates can be understood if we

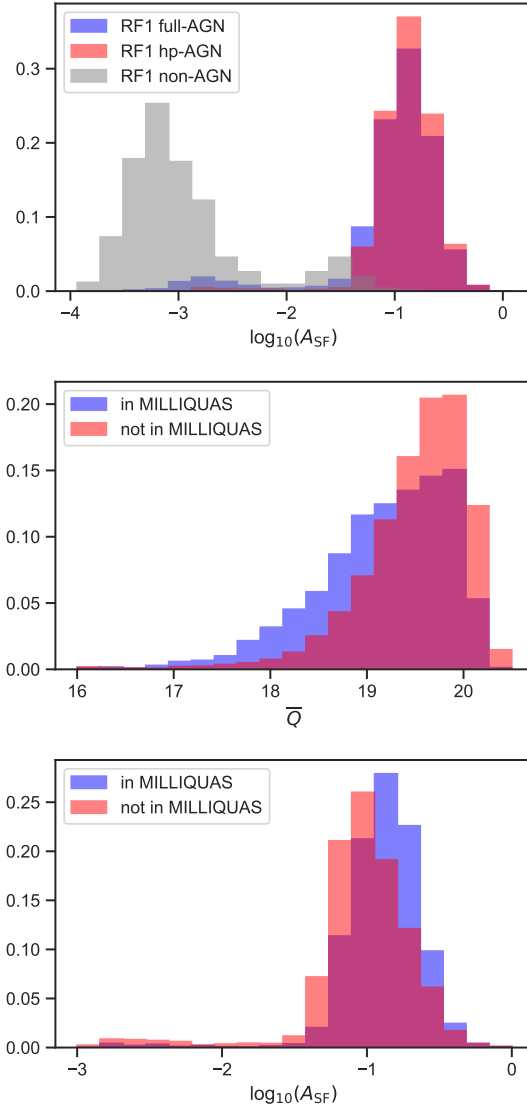


Figure 4.7: Top panel: normalized distribution of A_{SF} for sources with detection in MILLIQUAS with well-sampled light curves in QUEST–La Silla. We show sources from the RF1 full-AGN sample (blue), the RF1 hp-AGN sample (red), and sources classified as non-AGN by RF1. Middle panel: normalized distribution of the mean Q magnitude for sources from the hp-AGN sample that are present (blue) and not present (red) in MILLIQUAS. Bottom panel: normalized distribution of A_{SF} for sources from the hp-AGN sample that are present (blue) and not present (red) in MILLIQUAS.

Table 4.6: Number of hp-AGN candidates confirmed using MILLIQUAS, for each classifier

Sample	RF1		RF2		RF3	
	blue	red	blue	red	blue	red
MILLIQUAS AGN	1,882	122	1,894	116	1,901	100
MILLIQUAS candidate	324	30	322	34	320	25
X-ray detections	640	59	644	57	641	47
hp-AGN	3,618	2,323	3,613	1,639	3,646	836

consider that most of the AGN presented in MILLIQUAS come from samples that applied morphological cuts to target point sources, thus, they tend to exclude sources whose emission is dominated by their host galaxies.

Candidates with X-ray detections

MILLIQUAS provides X-ray detections associated to every source, however some recent catalogs like Luo et al. (2017) Chen et al. (2018) are not completely included. Thus, we used different X-ray catalogs to complement the information provided by MILLIQUAS and see which candidates have X-ray detections associated. For the COSMOS field we used the optical and infrared counterparts catalog of the Chandra COSMOS-Legacy Survey (Marchesi et al., 2016a), for the XMM-LSS field we used the recent XMM-SERVS survey catalog (Chen et al., 2018), for the ECDF-S field we used the Chandra Deep Field-South 7 Ms source catalog (Luo et al., 2017), and for the Elais-S1 field we used Elais-S1 field X-ray source optical/IR Identifications catalog (Feruglio et al., 2008).

In table 4.6 we provide the number of candidates with X-ray detections from the previously mentioned X-ray catalogs or MILLIQUAS (see the row “X-ray detections”). It can be seen that most of the candidates with X-ray detections are from the blue sub-sample.

4.4.2 Spectroscopic follow-up of AGN candidates

Since most of the AGN confirmed with ancillary data have blue colors ($g - r \leq 0.6$), we performed spectroscopic follow-up to confirm the nature of sources located in different regions of the color-color space. We used Goodman at SOAR (Clemens et al., 2004) and EFOSC2 at NTT (Buzzoni et al., 1984) instruments to observe 54 candidates (for details see Section 2.3.1 of the appendix).

To select the candidates for the follow-up campaign we divided the hp-AGN sample of the RF1 classifier into the blue and red sub-samples. Then, we randomly selected 100 candidates from each sub-sample, excluding sources with $r > 20.5$ for which it would be hard to obtain a good-quality spectrum with 4-meter class telescopes. We visually inspected the light curves of the selected candidates in order to exclude sources with evidence of bad photometry (produced by the relatively low cosmetic quality of the QUEST CCD camera

Table 4.7: Summary of the spectroscopic follow-up campaign

Class	blue sub-sample	red sub-sample	Total
AGN1	15	25	40
BAL-QSO	1	3	4
Galaxy	0	5	5
Star	0	5	5

chips). During the follow-up campaign we observed as much of the selected candidates as we could, observing in total 54 targets. We gave priority to sources from the red sub-sample, of the 54 candidates observed, 38 have $g - r > 0.6$, which represents 70% of the sample.

In the top panel of Figure 4.8 we show the r band magnitude distribution of hp-AGN candidates of the RF1 classifier and the observed candidates. We can see that the distributions are different. This is produced by the limitations of observing faint targets with 4-meter class telescopes. During the follow-up campaign we gave priority to sources with $r < 20$, for which we expected to obtain spectra with signal to noise higher than 10. In the bottom panel of Figure 4.8 we show the distribution of the predicted class probability (P_{RF}) for the hp-AGN candidates of the RF1 classifier and the observed candidates. It can be seen that in general we observed sources with higher probabilities, compared with the total sample of candidates. This is produced by the visual inspection of the light curves and the selection of brighter sources for the follow-up campaign.

We used the spectra to classify our targets and to estimate their redshifts. For details about the spectroscopic analysis see section 2.3.1 of the appendix. The full list of observed candidates can be found in Section 4.7.1 of the appendix. We provide the position of the observed sources, their redshift, their $g - r$ and $r - i$ colors, their r magnitude, and their spectroscopic classification. In Table 4.7 we provide a summary of the follow-up campaign. We divided the classified sources into blue and red sub-samples, and we also separate them according to their spectroscopic classes: AGN1 (type 1 AGN), BAL-QSO, galaxy, and star.

In the top-left panel of Figure 4.5, we show the color-color diagram of the observed candidates, with colors and shapes depending on their spectral classification. In the top-right and bottom-left panels of Figure 4.5 we mark with letters some candidates located in the stellar locus. The sources A and B are classified as type 1 AGN, and the sources C and D are classified as type M stars. From the light curves of sources C and D (see Figure 4.6) and from their spectra, we propose that these candidates are irregular variable stars.

In the bottom panel of Figure 4.8 we show the normalized redshift distribution of AGN from the labeled sample, and observed candidates classified as AGN or BAL-QSO. We can see that we have a much larger fraction of low-redshift sources observed during the follow-up campaign. Besides, the fraction of observed AGN with $z > 3.0$ is slightly higher compared with the AGN from the labeled sample.

There are seven targets with redshift higher than 2.5. These types of AGN are harder to detect than lower-redshift AGN in magnitude-limited, optical color-color selections (because

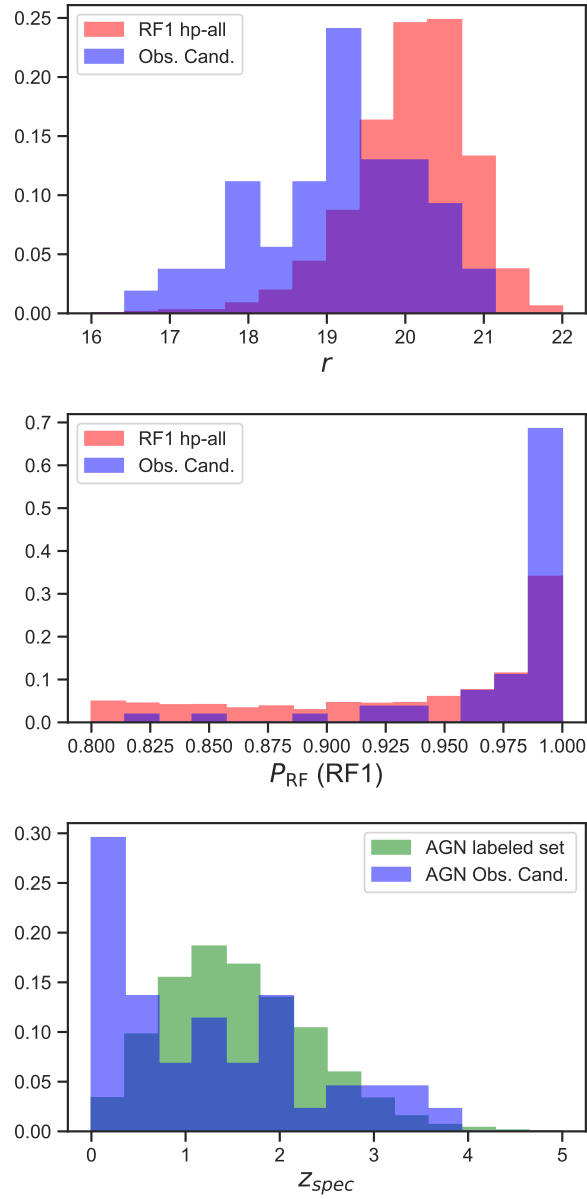


Figure 4.8: Top panel: normalized histogram of the r band magnitude of the RF1 hp-AGN sample (red) and the observed candidates (blue). Middle panel: normalized histogram of the predicted class probability (P_{RF}) of the RF1 hp-AGN sample (red) and the observed candidates (blue). Bottom panel: normalized histogram of spectroscopic redshift of AGN from the labeled sample (green) and observed candidates classified as type 1 AGN or BAL-QSO (red).

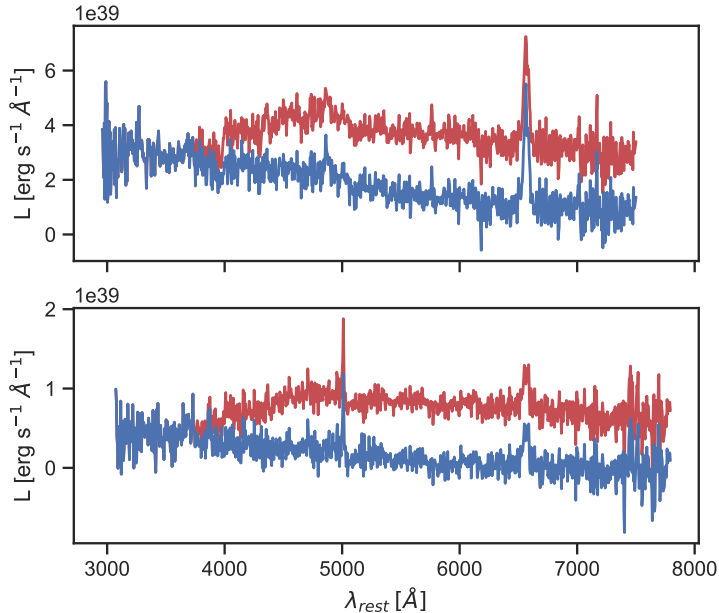


Figure 4.9: Rest-frame optical spectra of two type 1 AGN with evidence of continuum dominated by the host galaxy (LLAGN), observed with EFOSC2/NTT. In red we show the original spectra, and in blue we show the AGN component. The most prominent emission lines correspond to $H\alpha$.

their colors resemble those of stars, particularly near the magnitude limit of surveys where the stellar locus is wider), and clearly benefit from the variability criteria (Palanque-Delabrouille et al., 2011; Butler & Bloom, 2011; Palanque-Delabrouille et al., 2016). In addition, we found four BAL-QSO, with three having $g - r > 0.6$. There are 22 AGN with $z_{spec} < 0.7$, of which 20 have $g - r > 0.6$, and eight of these are LLAGN, whose continua are significantly dominated by the host galaxy, but with clearly distinguishable broad emission lines.

The case of the eight LLAGN is particularly interesting, since the continuum of their spectra is dominated by the host galaxy, but we still detect its optical variable component, which is associated with the accretion disk. This component is revealed by subtracting the galactic continuum of each spectrum following the simple procedure of Greene & Ho (2005) and Kim et al. (2006). As an example, in Figure 4.9 we show the spectra of two LLAGN sources, in red we show the original observed spectra and in blue we show the AGN component. It can be seen that in both cases the continuum is dominated by the host galaxy, but after its subtraction, the power-law AGN continuum appears, which is the one that produces the optical variations. It is important to remark that without including variability features in the selection of our candidates, these type of sources would be classified as non-AGN according to their optical colors. This reflects the importance of including variability in the selection of LLAGN.

We define the efficiency of a classifier as the number of confirmed AGN divided by the total number of observed candidates. Considering all the observed candidates, the efficiency is 100% for the blue sub-sample, and 73.7% for the red sub-sample. In Table 4.8 we show

from which classifier comes every observed candidate. As we mentioned previously, the 54 observed candidates belong to the RF1 hp-AGN sample, therefore the efficiency of the follow-up for the RF1 classifier is 100% for the blue sub-sample, and 73.7% for the red sub-sample. For the case of RF2, there are 50 observed candidates, with an efficiency of 100% for the blue sub-sample and 79.4% for the red sub-sample. There are three stars and one type 1 AGN excluded by RF2. For the case of RF3, there are 43 observed candidates. The efficiency of the RF3 blue sub-sample is 100% and for the red sub-sample is 85.2%. There are four type 1 AGN, one BAL-QSO, three galaxies, and three stars excluded by RF3.

From these results, we can conclude that RF2 has a higher efficiency compared to RF1 and RF3, since it has a high efficiency for both the blue and red sub-samples, and excludes most of the observed stars and only one type 1 AGN. RF3 also provides good results, however it excluded one BAL-QSO and four type 1 AGN, of which one has $z_{spec} = 3.5852$, and two are LLAGN.

Non-AGN observed sources

We observed 10 sources which were spectroscopically classified as stars or galaxies, but as AGN by our classifiers; all of which have clear evidence of variability. These 10 sources are selected as candidates by RF1, seven are selected as candidates by RF2 (two star and five galaxies), and four are selected as candidates by RF3 (two stars and two galaxies).

For the galaxy-classified cases, the light curves show clear signs of AGN-like variability. The obtained spectra of these sources are generally noisy (signal to noise less than 10), so we could be missing some weak emission lines. We tried to subtract a galactic component from these sources, and in some cases we found evidence of a power-law continuum component, but without evidence of emission lines. We decided to classify these sources as galaxies, however, in order to confirm their true nature, we likely need deeper observations, using 8-meter class telescopes. We note that Cartier et al. (2015) found that about 20% of the objects classified spectroscopically as galaxies showed variability, they found a similar percentage of narrow-line AGN showing variability.

For the star-classified cases, four of them are M-type stars, and one seems to be a K-type star. From their light curves, we can conclude that they are semi-periodic or irregular variable stars.

4.5 Comparison with previous works

Butler & Bloom (2011) used SDSS photometry to select AGN candidates through variability in the Stripe 82 field. They used damp random walk modelling (Kelly et al., 2009) to detect quasar-like variable sources. The light curves used in their analysis have on average 10 epochs, with a maximum of 28, obtained over ~ 6 years (Sesar et al., 2007). As can be seen in Figure 8 of Butler & Bloom (2011), most of their candidates lie in the region where typical AGN are found, and only $\sim 1\%$ of their candidates lie in the color-color space dominated by

stars (stellar locus).

Palanque-Delabrouille et al. (2011) and Palanque-Delabrouille et al. (2016) used SDSS photometry to select AGN candidates through variability in the Stripe 82 field, to be observed as part of the BOSS and eBOSS surveys. On average, the light curves used by them had 53 ± 20 epochs, and a total coverage spanning 4 and 10 years. They characterized the variability of each source using the structure function, and they classified the sources using a neural network algorithm. They demonstrated that their method is very efficient at selecting sources with $z_{spec} > 2.2$ or a BAL-QSO classification. However, the fraction of candidates with stellar-like colors is low (as can be seen in Figure 18 of Palanque-Delabrouille et al. 2011).

Peters et al. (2015) used SDSS data to perform a NBC KDE algorithm that classifies type 1 quasar, using color, variability and astrometric parameters. They used data in five broad optical bands (u , g , r , i , and z), and constructed light curves in each band with 10 to ~ 100 observations over time-scales from ~ 1 day to ~ 8 years. They used structure functions to characterize the variability of each source. They tested different combinations of these parameters, finding that by combining variability and colors, they can achieve 97% efficiency, improving particularly the efficiency in the selection of quasars at $2.7 < z < 3.5$. They selected 35,820 type 1 quasar candidates, with only the 14% of them having $g - r > 0.6$.

More recently, Tie et al. (2017) combined colors and variability properties to select AGN candidates from DES. They obtained light curves from DES, which span less than a year and typically have ~ 15 epochs. They used the chi-squared integrated probability to select AGN candidates. Since they have light curves with only a year of coverage, they did not implement more sophisticated variability selection methods. They demonstrated that combining variability with optical and infrared photometry improves the efficiency of AGN selection. Tie et al. (2017) provide a catalog of 1,263 spectroscopically confirmed quasars in the DES supernova fields brighter than $i = 22$ mag. Only 6% of their confirmed candidates have $g - r > 0.6$.

The light curves used in our analysis have a considerably higher cadence than the ones used in previous variability analyses, with an average of 119 ± 47 epochs and a total length of 1306 ± 254 days. In our case, we find that 39.1%, 31.2%, and 18.7% of the hp-AGN candidates from RF1, RF2, and RF3, respectively, have $g - r > 0.6$, where stars outnumber AGN. As we show in Section 4.4.2, most of the atypical AGN observed during our spectroscopic follow-up campaign lies in this region of color-color space (70% of the observed candidates). In general, our selection technique does not differ considerably to other techniques (e.g., Palanque-Delabrouille et al. 2011; Peters et al. 2015). The key difference is the larger number of epochs and the larger coverage of the QUEST-La Silla light curves, in addition to the exclusion of morphological parameters during the selection of candidates. This helps us to be more sensitive to atypical AGN, like BAL-QSO and LLAGN (observing four and eight, respectively, during the follow-up) compared with previous analyses, which might be related with the higher probability to detect a variable signal from our better quality light curves.

Our selection technique has the advantage of being easily applicable to LSST data, since the expected cadence for the DDF will be similar to the one used here (LSST Science Collaboration et al., 2009). However, it is important to note that LSST will provide *ugrizy*

photometry, with with single-epoch depths reaching to ~ 24 th magnitude, and ~ 27 th magnitude for the stacked images. Therefore, LSST will allow to perform variability analyses of much fainter sources than the QUEST–La Silla survey, and with multi–band light curves, which can help to weed out false positives and allow for a more complete characterization of the variability properties of the candidates.

Moreover, LSST will have the advantage of including the u band, which has been repeatedly used in the past to select AGN candidates (in combination with other photometric bands), after a morphological cut is applied (e.g., Fan 1999; Richards et al. 2002, 2004; Smith et al. 2005; Richards et al. 2009; Kirkpatrick et al. 2011; Bovy et al. 2011; Ross et al. 2012). These color–color selection techniques will be easier to apply to the full stacked depths of the LSST data, where the variability–based methods (like those presented in here) will eventually become infeasible, due to large flux errors on the light curves. However, for the case of AGN with strong host contamination variability–based methods will be in advantage over color–color selection techniques.

4.6 Conclusions

We have presented a methodology to classify AGN through variability analyses, particularly useful to find AGN populations missed by other optical selection techniques. We used data from the QUEST–La Silla AGN variability survey to construct a total of 208,583 well–sampled light curves in the COSMOS, XMM–LSS, Elais–S1, and ECDF–S fields. We characterize the variability of these sources by using different variability features (see Section 4.3.2). We used a Random Forest algorithm to classify our objects as either AGN or non–AGN using variability features and optical colors. We tested three classification schemes, one that includes only variability features (RF1), one that includes variability features and the $r - i$ and $i - z$ colors (RF2), and one that includes variability features and the $g - r$, $r - i$, and $i - z$ colors (RF3). We have a total of 5,941 AGN candidates for the RF1 classifier, 5,252 candidates for the RF2 classifier, and 4,482 candidates for the RF3 classifier.

We confirmed the nature of our candidates by using ancillary data, and we found that a high fraction of the candidates from each classifier with $g - r \leq 0.6$ are known AGN from the literature (52% for RF1, 52.4% for RF2, and 52.1% for RF3; see Section 4.4.1), but the fraction of candidates with $g - r > 0$ confirmed by ancillary data is low (5.3% for RF1, 7.1% for RF2, and 12% for RF3). This is produced because most of the AGN known from the literature are biased against bluer optical colors by their selection criteria. This motivated us to perform spectroscopic follow–up, to confirm the nature of sources located in different regions of the color–color space.

We observed 54 candidates with EFOSC2/NTT and Goodman/SOAR, with 70% of the observed targets having $g - r > 0$. We confirm the nature of several interesting sources, including four BAL–QSOs, seven sources with $z_{spec} > 2.5$, and eight LLAGN. Our method was very efficient in classifying AGN with $g - r \leq 0.6$, for which we achieved 100% of efficiency for all the classifiers. For the case of sources with $g - r > 0$, our method also demonstrated good performance, achieving 73.7% efficiency for RF1, 79.4% for RF2, and 85.2% for RF3.

From the spectroscopic follow-up campaign, we conclude that the optimal classifier is the one that includes variability features and the $r - i$ and $i - z$ colors (RF2), as it avoids the region of the color-color space where we normally find cool stars, and also shows high efficiency, excluding only one observed type 1 AGN. The RF3 classifier also provides good results, however it excluded four AGN and one BAL-QSO. For the case of RF1, we propose that most of the candidates with $g - r \sim 1.5$ and $r - i \gtrsim 0.8$ are LPV or binary stars.

Our work can be considered as a pilot study in preparation for LSST, since the selection techniques tested here can be easily implemented for LSST data. The cadence of the LSST's DDF will be similar to the one of QUEST-La Silla, but covering 10 years of observations, which will improve considerably the selection efficiency. In addition, LSST will provide observations in more than one photometric band, which should prove useful to discard artifacts and false positives. Particularly, LSST will provide u band photometry, which has been exhaustively used in the literature for the selection of point-like AGN, with very high efficiencies (e.g., Richards et al. 2004, 2009). Thus, color-color selection methods will remain a critical approach in the LSST era, particularly for the selection of faint, point-like AGN, since they can be applied to the full depths of the LSST data. However, optical color-color selections alone are not efficient at classifying morphologically extended AGN. For these type of objects, a combination of optical colors and variability-based methods will be more suitable, as we have demonstrated in this work.

4.7 Appendix

4.7.1 Catalog of observed candidates

Here we present the list of candidates observed during our spectroscopic follow-up campaign. We provide the equatorial coordinates in degrees (J2000), classifier from which the candidate was selected, measured redshift, quality flag of the measured redshift (1: low-quality z_{spec} , 2: good-quality z_{spec}), the magnitude in the r band, the $g - r$ color, the $r - i$ color, and the spectroscopic classification. For details about the spectral analysis of these targets see Section 4.7.2.

Table 4.8: Targets observed during spectroscopic follow-up.

Name	RA	DEC	Classifier	z_{spec}	FLAG $_z$	r	$g - r$	$r - i$	Class
QLS_1	7.021016	-45.806145	RF1/RF2	3.5852	1	20.14	1.13	0.31	AGN1
QLS_2	7.263506	-45.629417	RF1/RF2/RF3	1.3796	2	20.72	0.63	0.09	AGN1
QLS_3	7.323368	-43.633305	RF1/RF2/RF3	0.3242	1	18.57	0.03	-0.08	AGN1
QLS_4	7.377026	-46.529945	RF1/RF2	0.2824	1	19.32	1.53	0.51	Gal
QLS_5	7.387083	-43.664276	RF1/RF2	0.2000	1	19.16	1.08	0.39	Gal
QLS_6	7.418616	-42.320072	RF1	0.0	2	17.88	1.52	1.24	STAR
QLS_7	7.419530	-43.789948	RF1/RF2	0.3912	2	18.91	1.12	0.34	AGN1
QLS_8	7.820146	-45.645706	RF1/RF2	0.3123	2	19.07	1.41	0.49	AGN1
QLS_9	7.970508	-42.275764	RF1/RF2/RF3	0.1847	2	18.02	0.99	0.41	AGN1
QLS_10	8.304768	-45.681595	RF1/RF2/RF3	0.2676	2	18.73	0.92	0.43	AGN1
QLS_11	8.348364	-46.856922	RF1/RF2/RF3	3.5297	2	19.90	0.75	0.08	AGN1
QLS_12	8.629325	-42.310108	RF1	0.0	2	20.36	1.38	1.26	STAR
QLS_13	8.787973	-45.348194	RF1/RF2/RF3	0.1469	2	17.53	0.64	0.38	AGN1
QLS_14	9.407302	-43.000004	RF1/RF2/RF3	1.986	1	19.69	0.69	0.17	AGN1
QLS_15	9.414984	-43.422619	RF1/RF2/RF3	1.8265	2	20.45	-0.06	0.29	AGN1
QLS_16	10.021671	-43.859173	RF1/RF2/RF3	0.3710	2	19.28	0.86	0.32	AGN1

Continued on next page

Table 4.8 – *Continued from previous page*

Name	RA	DEC	Classifier	z_{spec}	FLAG $_z$	r	$g - r$	$r - i$	Class
QLS_17	10.097470	-44.866116	RF1/RF2/RF3	2.9609	1	19.94	0.70	0.38	BAL-QSO
QLS_18	10.225745	-43.855934	RF1/RF2/RF3	1.2671	1	20.49	0.63	-0.07	AGN1
QLS_19	10.758265	-42.452019	RF1/RF2/RF3	0.1000	1	19.90	0.89	0.24	Gal
QLS_20	10.866718	-43.825359	RF1/RF2/RF3	3.123	1	20.58	0.65	0.11	AGN1
QLS_21	11.090293	-43.665966	RF1/RF2/RF3	0.0	2	17.12	0.78	0.24	STAR
QLS_22	30.591522	-2.020991	RF1/RF2/RF3	2.0502	2	18.66	0.04	0.11	AGN1
QLS_23	30.603312	-1.752825	RF1/RF2/RF3	0.2101	2	18.12	0.71	0.33	AGN1
QLS_24	31.057844	-2.953287	RF1/RF2/RF3	0.1500	1	19.26	0.91	0.36	Gal
QLS_25	31.167528	-3.630512	RF1/RF2	1.22	1	19.14	0.61	0.06	BAL-QSO
QLS_26	31.533081	-2.510754	RF1/RF2/RF3	1.4319	1	18.12	0.09	0.06	AGN1
QLS_27	32.158398	-3.652800	RF1	0.0	2	18.02	1.45	1.67	STAR
QLS_28	32.456287	-3.651828	RF1/RF2/RF3	1.4976	2	18.85	0.09	0.14	AGN1
QLS_29	33.409081	-3.250347	RF1/RF2/RF3	2.8491	2	19.54	0.16	0.06	AGN1
QLS_30	33.474072	-3.279924	RF1/RF3	0.5671	1	20.81	0.88	0.80	AGN1
QLS_31	36.426113	-2.971209	RF1/RF2/RF3	0.0	2	16.53	0.67	0.22	STAR
QLS_32	36.852539	-2.401858	RF1/RF2/RF3	0.2551	2	19.94	0.84	0.38	AGN1
QLS_33	37.988422	-2.585521	RF1/RF2/RF3	0.3498	2	18.83	0.09	0.01	AGN1
QLS_34	38.680885	-2.637419	RF1/RF2/RF3	0.8437	2	19.86	0.46	0.06	AGN1
QLS_35	51.529488	-29.656691	RF1/RF2/RF3	0.2307	2	18.15	0.87	0.43	AGN1
QLS_36	51.765114	-27.740358	RF1/RF2/RF3	2.0279	2	19.20	0.07	0.17	AGN1
QLS_37	52.009411	-28.600405	RF1/RF2	0.2667	2	18.01	1.33	0.49	Gal
QLS_38	52.013538	-30.619936	RF1/RF2/RF3	0.3284	2	19.47	0.90	0.37	AGN1
QLS_39	52.397503	-27.657492	RF1/RF2/RF3	1.4175	2	19.59	0.60	0.20	AGN1
QLS_40	52.536362	-29.822481	RF1/RF2/RF3	0.1804	2	17.51	0.66	0.53	AGN1
QLS_41	52.593563	-29.353525	RF1/RF2/RF3	0.2177	2	19.94	0.61	0.41	AGN1
QLS_42	53.317341	-29.488207	RF1/RF2/RF3	1.9234	2	19.48	0.16	0.22	AGN1
QLS_43	53.730831	-27.736212	RF1/RF2/RF3	3.4986	1	19.37	0.86	0.27	BAL-QSO
QLS_44	53.749317	-27.499168	RF1/RF2	0.3598	2	19.42	1.13	0.40	AGN1
QLS_45	53.847992	-28.123224	RF1/RF2/RF3	0.8682	2	17.11	0.09	-0.06	AGN1
QLS_46	53.864979	-26.950115	RF1/RF2/RF3	0.8159	2	19.50	0.04	0.08	AGN1
QLS_47	53.911106	-28.961195	RF1/RF2/RF3	0.2116	2	18.54	0.72	0.46	AGN1
QLS_48	53.943211	-27.432163	RF1/RF2/RF3	0.4332	1	19.18	1.02	0.32	AGN1
QLS_49	54.049484	-28.095388	RF1/RF2/RF3	2.5265	2	18.87	0.10	-0.04	AGN1
QLS_50	54.782047	-26.617104	RF1/RF2/RF3	0.3825	2	19.14	0.70	0.26	AGN1
QLS_51	54.785744	-28.131121	RF1/RF2/RF3	2.1388	1	19.20	0.34	0.09	BAL-QSO
QLS_52	54.878654	-27.307127	RF1/RF2/RF3	1.0899	2	19.72	0.62	0.16	AGN1
QLS_53	55.059608	-26.485237	RF1/RF2/RF3	1.6453	2	19.43	0.05	0.17	AGN1
QLS_54	149.004532	1.161204	RF1/RF2/RF3	2.3530	1	20.36	0.33	0.09	AGN1

4.7.2 Spectroscopic analysis of the observed candidates

We obtained classification spectra for 21 of our candidates using both the red and blue cameras of the Goodman spectrograph (Clemens et al., 2004), mounted on the SOAR telescope. We used the 400 lines mm^{-1} grating and the 1.0" and 0.8" slits providing a typical resolution of ~ 6 or better. We reduced Goodman data following usual steps including bias subtraction, flat fielding, cosmic ray rejection (see van Dokkum 2001), wavelength calibration, flux calibration, and telluric correction using our own custom IRAF⁵ routines.

We also obtained classification spectra for 33 candidates using EFOSC2 (Buzzoni et al., 1984) mounted on the New Technology Telescope (NTT) at La Silla Observatory. We used the 236 lines mm^{-1} grating and the 1.0" slit providing a typical resolution of 18 Å. We followed

⁵IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

the same observing procedures as the Public ESO Spectroscopic Survey for Transient Objects (PESSTO) collaboration (Smartt et al., 2015), using the PESSTO Observing Blocks (OBs) to perform our observations, We reduced our observations using the PESSTO pipeline (Smartt et al., 2015).

We corrected the reduced and calibrated one-dimensional spectra by Galactic extinction using the maps of Schlegel et al. (1998) and the model of Cardelli et al. (1989). We then computed their redshifts and spectral classes by cross-correlating every spectrum with a set of spectral classification templates from SDSS⁶, a type 1 AGN composite spectrum (Croom et al., 2002), and a type 2 AGN spectrum (Jones et al., 2009). We define a redshift FLAG that indicates the quality of the measured redshift. We say that a computed redshift has good-quality (FLAG=2) when there are several lines in the spectrum, and these lines are not affected by absorption features; and we say that a computed redshift has low-quality (FLAG=1) when the spectrum has low signal to noise, when the number of emission lines available is low (one or two), or when the emission lines are highly affected by absorption features. We measured the full width at half-maximum (FWHM) of the emission lines (when they were present) of each spectrum, following a simple Gaussian fitting procedure with the *PySpecKit* Python package (Ginsburg & Mirocha, 2011). The final classification of every source was done complementing the results of the cross-correlation analysis with visual inspection of every spectrum. To distinguish type 1 and type 2 AGN, we requested that at least one emission line has FWHM > 1800 km/s in the rest-frame. In Figure 4.10 we provide the rest-frame optical spectra of our 54 candidates.

⁶<http://classic.sdss.org/dr5/algorithms/spectemplates/>

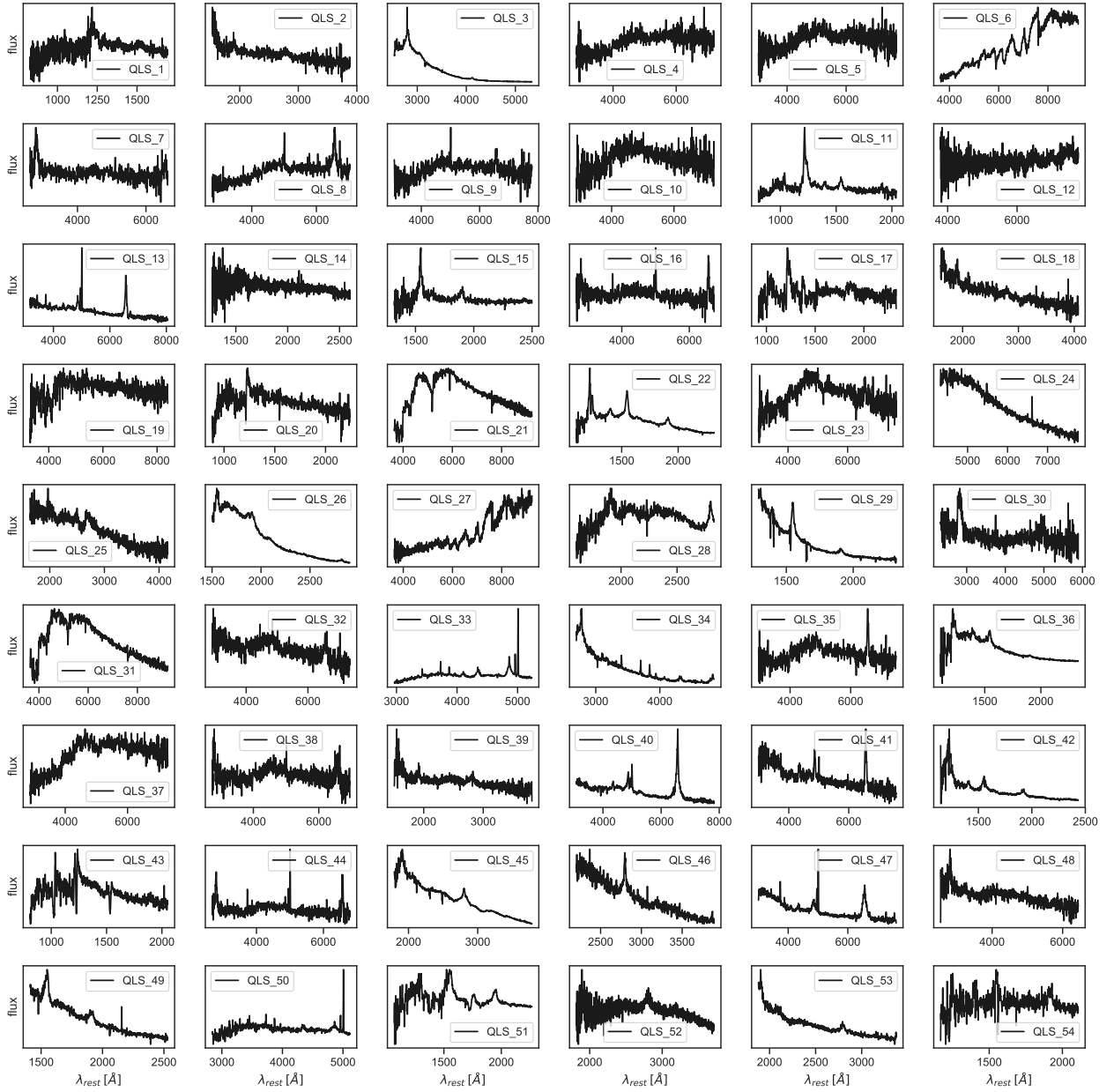


Figure 4.10: Rest-frame optical spectra of the observed candidates. The flux is in arbitrary units.

Chapter 5

Dust Reverberation Mapping of $z < 1.2$ AGN

5.1 Introduction

As we mentioned in Section 1, current AGN models suggest that two of the main structures of AGN are the accretion disk and the dusty torus. In particular, the continuum emission observed in the UV/optical range is expected to come from the accretion disk, and the infrared emission from the dusty torus. Previous studies have demonstrated that the variations in the UV/optical continuum flux are observed at a later time in the infrared continuum emission. These delays between the UV/optical and infrared emissions can be understood if we assume that the infrared emission is produced by the dusty torus reprocessing of the UV/optical emission from the accretion disk, although the corona can also illuminate the torus directly, as probed from X-ray spectroscopy (e.g., Ramos Almeida & Ricci 2017). Then, the measured time lag can be interpreted as the light-travel distance from the accretion disk to the dusty torus (e.g. Suganuma et al. 2006; Lira et al. 2011; Koshida et al. 2014; Lira et al. 2015).

Reverberation mapping (RM) is a unique technique to study the spatially-unresolved structures of AGN (e.g., Peterson 1993; Peterson et al. 2004; Suganuma et al. 2006; Arévalo et al. 2009; Breedt et al. 2009, 2010; Lira et al. 2011; McHardy et al. 2014; Koshida et al. 2014; Lira et al. 2015; Edelson et al. 2015; Fausnaugh et al. 2016; Grier et al. 2017; McHardy et al. 2018; Edelson et al. 2019). RM measures the observed time lags (Δt) between contemporaneous light curves of AGN observed in different bands, by cross-correlating the light curves, and associating these lags to the light travel time between the different structures of the AGN. In particular, dust reverberation mapping (DRM) is a useful technique to estimate the radius of the innermost hot dusty torus, by computing the time delays between optical and NIR light curves (e.g., V band versus K band) (Suganuma et al., 2006; Lira et al., 2011; Shappee et al., 2014; Koshida et al., 2014; Lira et al., 2015).

Previous analyses have used optical and NIR light curves to study the radius of the dusty torus. Suganuma et al. (2006) performed DRM of four local Seyfert 1 galaxies. They found that the measured time lags were tightly correlated with the optical luminosities,

with $\Delta t \propto L^{0.5}$, and they found no correlation between Δt and the BH masses. More recently, Koshida et al. (2014) performed DRM of 17 nearby Seyfert 1 galaxies. Again, they found a strong correlation between the optical luminosity and the measured time lags, with $\log_{10}\Delta t = -2.11 - 0.2M_V$, assuming $\Delta t \propto L^{0.5}$.

This tight correlation between luminosity and Δt is expected from torus models (e.g Barvainis 1987; Kishimoto et al. 2007), in which the innermost radius of the torus is determined by the sublimation radius (R_{sub}), which corresponds to the radius at which dust particles sublimate for a given temperature and luminosity. For the case of dust composed by graphite grains the sublimation radius can be determined by:

$$R_{\text{sub}} = 1.3 \times \left(\frac{L_{\text{UV}}}{10^{46} \text{ erg s}^{-1}} \right)^{0.5} \left(\frac{T_{\text{sub}}}{1500 \text{ K}} \right)^{-2.8} \left(\frac{a}{0.05 \mu\text{m}} \right)^{-0.5} \text{ pc}, \quad (5.1)$$

where L_{UV} , T_{sub} , and a are the UV luminosity of the accretion disk, sublimation temperature of dust, and dust grain size, respectively.

Recently, it has been proposed that we can use the time lag between optical and NIR continuum emission of AGN as standard candle for cosmology, since it shows a tight correlation with luminosity (e.g Hönig et al. 2017). However, all the DRM previous studies have been done for a small sample of local sources (less than 20 objects), as we pointed out above. It is crucial to perform DRM studies for large sample of sources located at different redshifts, in order to study whether the same correlations observed between L and Δt in the local universe are also observed at higher redshifts.

Moreover, as mentioned in Section 1, the SED of AGN around the rest-frame wavelength of $\lambda \sim 1 \mu\text{m}$ samples simultaneously two emission components, the accretion disk and the hot part of the dusty torus (Glass, 1992; Landt et al., 2011). Therefore, beyond $\lambda \sim 1 \mu\text{m}$ we expect to observe continuum emission dominated by the dusty torus. However, emission from both, the accretion disk and the dusty torus, has been detected in the J and H bands of some local sources (Lira et al., 2011, 2015). Therefore, it is crucial to study whether the NIR emission of high redshift sources is dominated by the dusty torus or the accretion disk.

In this Chapter we present our DRM analysis, done to determine the characteristic optical–NIR time lags, to study whether the emission received in the NIR of high redshift sources is consistent with emission from the dusty torus or the accretion disk, and to determine whether we can use Δt as a standard candle. We used NIR data from the UltraVISTA survey (McCracken et al. 2012; see Section 1.3.2) to construct YJHKs light curves for sample of X-ray selected AGN, with redshifts below 1.2. We avoid sources with higher redshifts since we need to detect emission beyond $\lambda \sim 1 \mu\text{m}$ in the rest-frame, in order to study the dusty torus using UltraVISTA data. The Chapter is organized as follows. In Section 5.2 we describe the data used for this analysis. In Section 5.3 we describe how we selected our AGN sample. In Section 5.4 we show the results of our SED analysis. In Section 5.5 we describe the methodology used to subtract the accretion disk component from our Ks light curves. In Section 5.6 we describe the cross-correlation (CC) methods used during this work, and provide the CC results. In Section 5.7 we study the correlation between the measured time lags with the luminosity of our sources. Finally, in Section 5.8 we summarize the main results and provide comments about the future work. The photometry reported is in the AB system. We adopt the cosmological parameters $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

5.2 Data

Our work is based on the NIR imaging data from the UltraVISTA survey (McCracken et al., 2012), which has repeatedly imaged the COSMOS field during five years in the *YJHKs* bands (see Section 1.3.2). The data considered in this work were taken between December 2009 and June 2016, which corresponds to the 4th UltraVISTA data release.

We processed the UltraVISTA individual OBs stacks following the same procedures presented in Sections 2.2.1 and 2.4.1. In summary, we further PSF homogenized the individual images and applied a photometric re-calibration to match the UltraVISTA DR3¹ catalogs (see Section 2.4.1 for further details). We then constructed light curves for all the sources in the clean-AGN catalog (described in Section 2.2.2) with a detection in the UltraVISTA single images (see Section 2.4.2).

5.3 Sample Selection

As mentioned in Section 5.2, we constructed light curves for all the sources in the clean-AGN catalog with a detection in the UltraVISTA single images. This catalog is composed by X-ray selected sources from the catalog of optical and infrared counterparts of the Chandra COSMOS-Legacy Survey (Marchesi et al., 2016a). From this catalog we obtained X-ray fluxes and rest-frame K-corrected luminosities measured in three bands (Soft: 0.5-2 keV, Hard: 2-10 keV, and Full: 0.5-10 keV), and spectroscopic and photometric classes (see Sections 2.3.1 and 2.3.2).

Unlike Chapter 2, for this analysis we only considered sources classified as type 1 by Marchesi et al. (2016a) with $z \leq 1.2$. There are 1015 sources in Marchesi et al. (2016a) catalog classified as type 1 spectroscopically (by optical spectra) or photometrically (by SED analysis). 692 of these sources have a spectroscopic redshift below 1.2, for which we can observe the rest-frame $\lambda \sim 1 \mu\text{m}$ emission using the UltraVISTA data. For 79 of these sources we could construct light curves in the four NIR bands (YJHKs). We called this sample as “UVISTA-type 1”. Figure 5.1 shows a light curve example a source from the UVISTA-type 1 sample.

We performed a variability analysis of the UVISTA-type 1 sample, following the same approach of Chapter 2. We computed the P_{var} and σ_{rms} parameters, and classify the sources as variable or non-variable, according to the $P_{var} \geq 0.95$ and $(\sigma_{rms}^2 - err(\sigma_{rms}^2)) > 0$ criteria. There are 31 sources in the UVISTA-type 1 sample classified as variable in the four NIR bands. We called this sample as “var-type 1”. Since to perform DRM analysis we need variable light curves, for the rest of this work we used the var-type 1 sample.

¹http://www.eso.org/sci/observing/phase3/data_releases/uvista_dr3.pdf

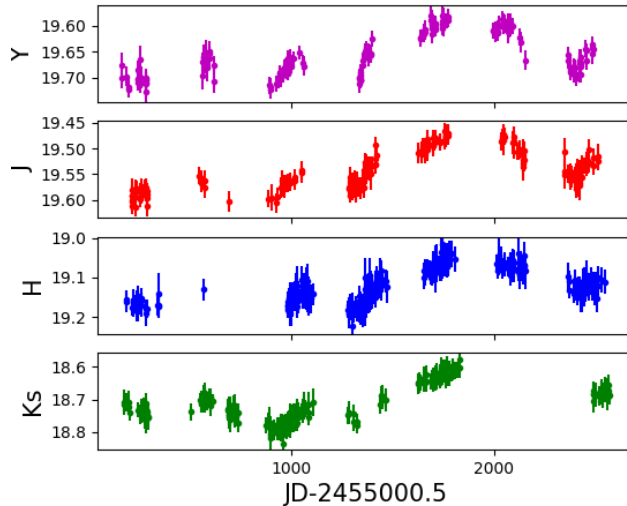


Figure 5.1: Observed light curve for a type 1 AGN located at RA= 150.194687° and DEC= 2.067890° from Marchesi et al. (2016a) catalog, located at $z=0.554$, variable in all the filter bands.

5.4 SED Analysis

Following a similar approach than Lira et al. (2011), we performed an SED analysis of the total and variable components of the YJKs light curves of every source in the var-type 1 sample. We corrected every light curve by Galactic extinction using the maps of Schlegel et al. (1998) and the model of Cardelli et al. (1989).

For the total SED component, we computed the average flux of every YJKs light curve for each object, and computed the rest-frame wavelength and rest-frame emission flux associated to a given band for each source. For the variable SED component, we computed the difference SED of each source, which corresponds to the subtraction of the observed SED around the lowest flux level from the SED observed at the highest flux level. When we do this, we subtract any constant component from the SED, such as the host galaxy component, thus, we can obtain an SED that is not contaminated by the host galaxy emission, which can be particularly strong in the NIR.

We computed the difference SED in four different ways: a) by subtracting the maximum and minimum flux of each YJKs light curve (max-min), b) by computing the difference between the percentiles 95 and 5 of each light curve ($P_{95} - P_5$), c) by computing the difference between the percentiles 90 and 10 of each light curve ($P_{90} - P_{10}$), and d) by using the non-normalized excess variance of the flux ($\sigma_f^2 = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} [(f_i - \bar{f})^2 - \sigma_{err,i}^2]$). To take into account the photometric errors of the light curves in the difference SED measurements, we constructed 5000 light curves for each source observed in a given band, by generating a random number drawn from a Gaussian distribution with mean and standard deviation equal to the measurement value and its uncertainty, respectively, at each light curve epoch. We then computed the median and the 15.86 and 84.14 percentiles of the distribution of the

max-min, $P_{95} - P_5$, $P_{90} - P_{10}$, and σ_f^2 measurements.

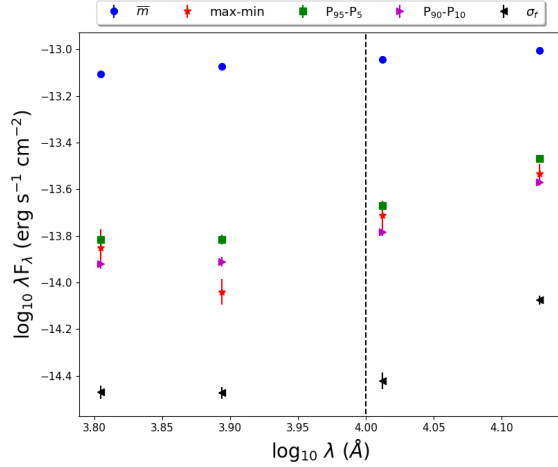
Since we have only 4 points in every SED, we could not perform an SED fitting procedure. However, we can detect the presence of a dusty torus SED component by looking at the shape of every SED around $\lambda \sim 1 \mu\text{m}$, since previous studies has shown that there is a local minimum of the SED at $\lambda \sim 1 \mu\text{m}$, produced by the end of the accretion disk dominated SED region, and the beginning of the host dust dominated SED region. If there is an SED component associated to the dusty torus emission, we might expect to see the beginning of the infrared bump at $\lambda \sim 1 \mu\text{m}$, as, for instance, in Figure 5 of Glass (1992) or in Figures 4 and 5 of Lira et al. (2011).

In Figure 5.2 we show examples of sources with, without, and unclear variable torus component in the SED. There are 11 sources in the var-type 1 sample with a clear variable torus component, nine without a torus component, but instead a disk dominated variable SED, and 11 for which we cannot clearly confirm the presence of the torus component. In Figure 5.3 we compared the redshift and the Hard band X-ray luminosity distributions of the sources with SED dominated by the torus and the accretion disk. For the case of the X-ray luminosity, the differences are not significant. However, we can see in the redshift distribution that most of the sources with SED dominated by the accretion disk have $z_{spec} > 0.9$. Eight of the nine sources without a variable torus SED component, and seven of the 11 sources with unclear variable torus SED component, are located at $0.9 < z < 1.2$. For the case of the sources with a clear torus component, 5 of the 11 sources have $0.9 < z < 1.2$. At this redshift regime, the emission observed in the Ks band can have a non-negligible contribution from the accretion disk, which might explain the lack of a clear torus component in the SED of 20 of our sources.

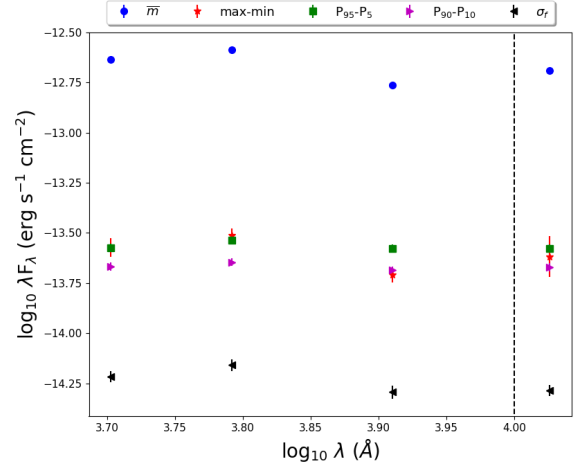
This SED analysis can help us to determine whether the rest-frame NIR emission is dominated by the dusty torus or the accretion disk emission, without doing any CC analysis. This is particularly useful, considering that for sources with monotonous light curves (i.e. light curves without peaks and valleys) the CC analysis can produce poor results, hindering the measurement of time lags and the detection of a torus from DRM analysis. From this analysis, we can conclude that 11 of the 31 sources in the var-type 1 sample have NIR variable emission that is dominated by the dusty torus, and for the 20 remaining sources, the emission is dominated by the accretion disk, or by a combination of both components.

5.5 Subtraction of the Accretion Disk Component from the rest-frame NIR emission

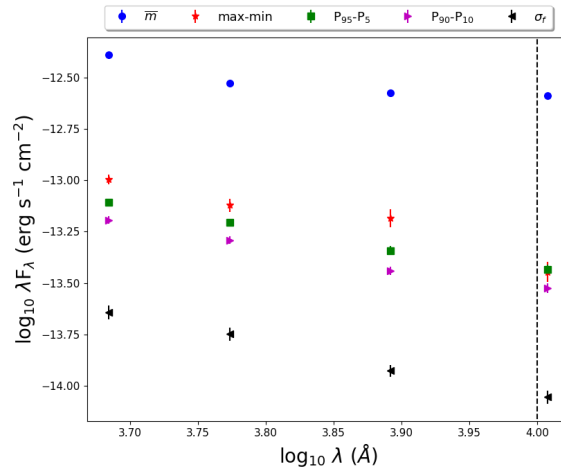
As we mentioned in the previous section, the AGN rest-frame NIR emission can be contaminated by variable flux from the accretion disk (Lira et al., 2011; Koshida et al., 2014; Lira et al., 2015). This can affect the CC analysis producing the measurement of lags that are shorter than actual light-travel distance from the accretion disk to the dusty torus. In order to obtain a proper disk-torus lag, we can subtract the accretion disk component from the rest-frame NIR emission following the same approach of Koshida et al. (2014) (see also Mandal et al. 2018), where the NIR accretion disk component at a given time t is estimated



(a)



(b)



(c)

Figure 5.2: Examples of SEDs of the total component (light curves averages), and the four difference SED estimations (max-min, $P_{95} - P_5$, $P_{90} - P_{10}$, and σ_f^2), for (a): a source with clear variable torus component, (b): a source with unclear variable torus component, and (c): a source without variable torus component. We mark with a black dashed line the position where $\lambda_{\text{rest-frame}} = 1 \mu\text{m}$.

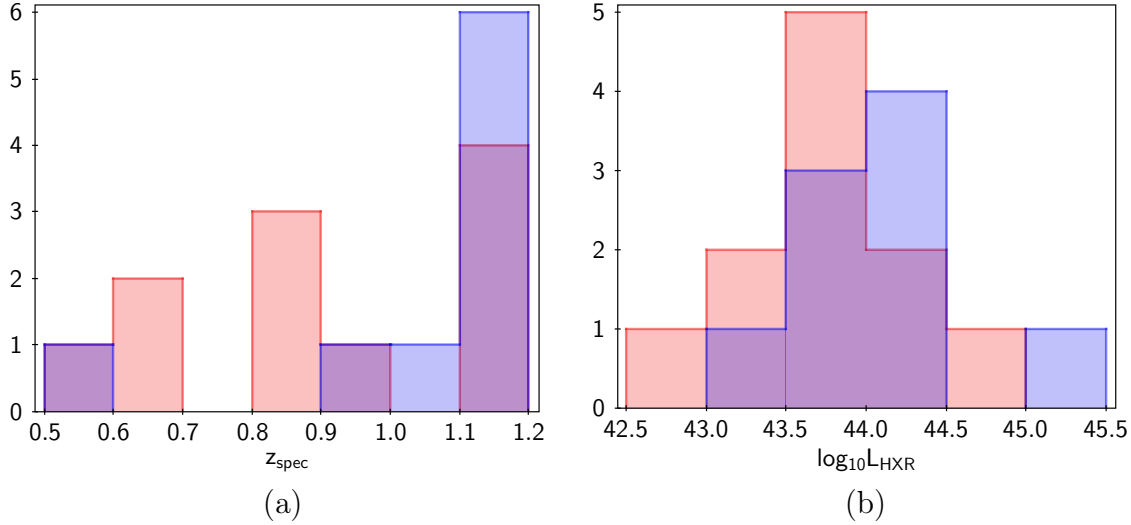


Figure 5.3: Redshift (a) and rest-frame Hard band X-ray luminosities (b) distribution of sources with SED dominated by the torus (red) and SED dominated by the accretion disk (blue).

from the rest-frame optical emission, assuming a power-law spectrum of the accretion disk emission:

$$f_{\text{NIR,disk}} = f_{\text{opt,disk}}(t) \left(\frac{\nu_{\text{NIR}}}{\nu_{\text{opt}}} \right)^{\alpha_{\nu}}, \quad (5.2)$$

where $f_{\text{NIR,disk}}(t)$ is the NIR flux of the accretion disk at time t , $f_{\text{opt,disk}}(t)$ is the optical flux of the accretion disk at time t , ν_{opt} and ν_{NIR} are the rest-frame wavelengths of the NIR and optical emission, respectively, and α_{ν} is the power-law index. Following Koshida et al. (2014) and Mandal et al. (2018), we assume that $\alpha_{\nu} = 1/3$. Then, we obtain the dusty torus NIR flux by subtracting the total flux in the NIR and the estimated $f_{\text{NIR,disk}}$.

In our case, we used the emission in the Y band to obtain the emission in the rest-frame optical range, and the Ks band to obtain the emission in the rest-frame NIR range. Since the epochs of the Y and Ks light curves of a given source are not exactly the same, we did a linear interpolation of the Y band light curves, and then we evaluated them at the Ks light curves epochs, before applying Equation 5.5. It is important to note that we did not subtract the host galaxy component from the Y and Ks light curves, since all our sources look as point sources in the UltraVISTA data, which hinders the proper modelling of the host galaxy component. However, since the host galaxies do not contribute a variable flux component to the light curves, the CC analysis will not be affected. In the panel (a) of Figure 5.4 we show, as an example, the original Y and Ks light curves, and the Ks light curve after the accretion disk component subtraction, for a source located at $z_{\text{spec}} = 0.963$.

5.6 Cross–correlation Analysis

5.6.1 Challenges of the Cross–correlation Analysis

Previous DRM studies have been done for local sources (e.g Suganuma et al. 2006; Lira et al. 2011; Koshida et al. 2014; Lira et al. 2015), for which the estimation of the AGN and host galaxy component can be done without difficulty, and for which the CC analysis can be done from densely sampled, targeted observations. However, if we want to use DRM for cosmology, we need to study sources located at higher redshift, which make the DRM analysis more complex.

One of the first problems that appears when dealing with non–local sources is the time dilation that affects the light curves, where the light curve in the rest–frame can be obtained by computing $t_{rest} = t_{obs}/(1 + z)$. For sources at $z \sim 1$ this dilation is non–negligible, since the light curve in the rest–frame will have a length that is half the time observed. Besides, the AGN/host galaxy decomposition is much harder for high redshift sources, since they might look as point sources in the images.

Other problems come from the observational bias produced by flux–limited surveys, where we tend to observe AGN with higher luminosities and higher BH masses as we go to higher redshifts (as we shown in Section 3.4.3). Since sources with higher luminosities have lower variability amplitudes (e.g Vanden Berk et al. 2004; Kelly et al. 2009; MacLeod et al. 2010; Simm et al. 2016; Caplar et al. 2017; Sánchez-Sáez et al. 2018), the detection of a variable signal for high redshift sources can be more difficult. On the other hand, as we mentioned in Section 1.2.1, there is a positive correlation between the time–scale of the variability and the BH mass. Combining this with the fact that at higher redshift we observe sources with higher BH masses, we end up observing sources with larger time–scales as we go to higher z . Thus, if we want to obtain a light curve with enough structure to perform a CC analysis, we need longer light curves as we go to higher redshifts.

Finally, in most of the cases, we use ground based light curves to perform a RM analysis. These light curves can have seasonal gaps, which produces problems in the CC analysis when the time delays that we want to measure have values of the same order as the gap lengths. For the case of the UltraVISTA light curves, the gaps have lengths of ~ 100 days, which is similar to the time delays that we can expect for the DRM analysis. Thus for some sources the CC analysis will inevitably produce poor results.

5.6.2 ICCF

To calculate the time lags between our light curves we used the interpolated cross–correlation function (ICCF; Peterson et al. 1998, 2004). The ICCF calculates the the cross–correlation function (CCF) between two time series by interpolating the light curves to a desired cadence. It assumes that a good representation of the missing information between two epochs in a light curve is the linear interpolation between these two epochs (e.g Peterson et al. 2004). In our case, we used a cadence of 2 days between $-t_S/4$ and $t_S/2$, where t_S is the time

span in common between the two light curves. The time lag between the light curves can be determined by using the peak of the CCF or its centroid. In our case, we estimated the lag using the centroid of the CCF, computed by selecting all those CCF points with values above $0.8 \times r_{\max}$, where r_{\max} correspond to the peak value of the CCF.

The uncertainties of the ICCF are determined by a model-independent Monte Carlo simulation based on a flux randomization/random subset selection (FR/RSS) method (see Peterson et al. 1998, 2004), where N_{lc} (10000 in this work) light curves are generated from an original light curve, by randomly modifying the fluxes of each epoch (FR), and randomly selecting a fraction of epochs (RSS), in our case the 70% of the epochs. Then the method computes N_{lc} CCF and provides the centroid cross-correlation distribution (CCCD). The final centroid of the CCF is then determined with the median of the CCCD, and the errors are determined with the 15.86 and 84.14 percentiles of the CCCD. During the FR/RSS process, only those simulations with $r_{\max} > 0.5$ are kept, and the remaining simulations are considered as non-significant. From these, we define the number of successful simulations N_s , as the number of simulations for which the criteria $r_{\max} > 0.5$ is accomplished.

In this work we used PyCCF², which is a Python implementation of the ICCF method (Sun et al., 2018). In order to avoid some of the problems discussed in Section 5.6.1, we used the whole light curves during the CC analysis.

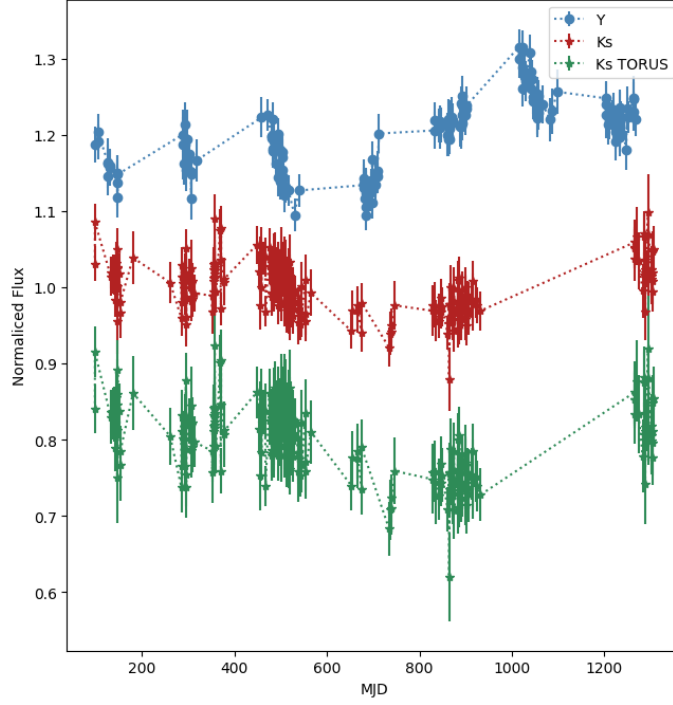
5.6.3 Lag-significance Criteria

One of the main problems of our DRM analysis is that, in some cases, the actual lag between the accretion disk and the dusty torus has a value similar to the size of the gaps present in the light curves used for the analysis. This might produce CCCD with several peaks, or CCCD with peaks located at time lags where the number of epochs in common between the two light curves is low.

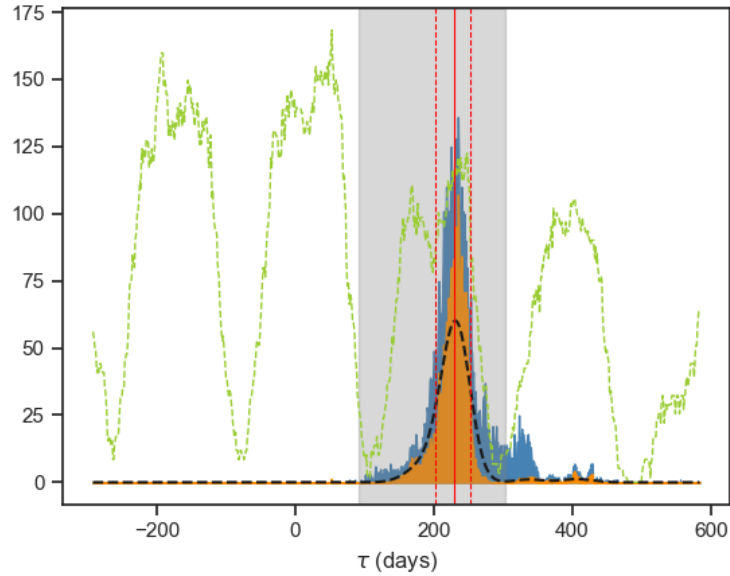
To determine the time lag and its uncertainties of a given CCCD, we followed the same approach of alias identification and removal of Grier et al. (2017). We first computed the fraction of the number of epochs in common between the light curves at a given time lag $N(\Delta t)$ and the number of epochs in common between the light curves at $\Delta t = 0$: $P(\Delta t) = [N(\Delta t)/N(0)]$. For time lags at which there are no epochs in common $P(\Delta t)$ will be equal to zero. Then, we weighted the CCCD by $P(\Delta t)^2$, and smoothed the final distribution by a Gaussian kernel with a width of 15 days (selected by visual inspection of the CCCD as in Grier et al. 2017). After this, we identified the tallest peak of the smoothed distribution, and searched for local minima on both sides of this peak. Then, we rejected all those lags of the original CCCD that fell outside these local minima. Finally, the time lag and its uncertainties were computed using the remaining lags in the CCCD. In the panel (b) of Figure 5.4 we show an example of the final time lag determination of a source located at $z_{spec} = 0.963$.

We also followed the approach of Grier et al. (2017) to determine whether a measured time lag corresponds to a significant detection, using the following criteria:

²<http://ascl.net/1805.032>



(a)



(b)

Figure 5.4: (a): Original Y (blue) and Ks (red) light curves, and Ks light curve after the accretion disk component subtraction (green) of a source located at RA= 149.351562° and DEC= 2.588570°, and at $z_{spec} = 0.963$. The light curves are presented with their flux normalized and shifted by a given value, for visualization purposes (b): Final time determination procedure of the same source in (a). We show the original CCCD (blue distribution), the weighted CCCD (yellow distribution), the smoothed distribution (black dashed lines), and $P(\Delta t)$ (green dashed lines). The gray dashed region shows the lag range used to define the final time lag (red solid line) and its uncertainties (red dashed lines).

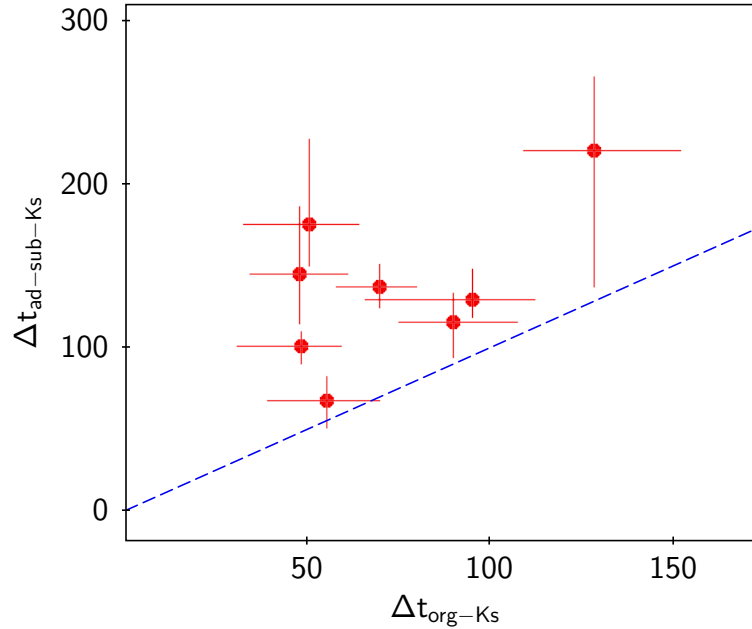


Figure 5.5: Comparison of the time lags measured for the org–Ks ICCF (x–axis) and the ad–sub–Ks ICCF (y–axis), that satisfied the three conditions of Section 5.6.3. The blue dashed line shows the 1:1 relation.

1. The number of successful simulations is equal to or higher than 5000 ($N_s \geq 5000$).
2. The alias identification and removal procedure rejects less than the 60% of the samples.
3. The measured lag is not consistent with zero at a 2σ level ($\Delta t > 2 \times \sigma_{\Delta t}$).

5.6.4 Cross–correlation Results

We computed the CCCD for our 31 sources in the var–type 1 sample. As we mentioned, we used the light curves in the Y band to describe the accretion disk emission, and the light curves in the Ks band to describe the torus emission. We computed the ICCF using the original Ks light curves (org–Ks ICCF), and the Ks light curves after the accretion disk component subtraction (ad–sub–Ks ICCF). 25 of the org–Ks ICCF and 17 of the ad–sub–Ks ICCF satisfied condition 1 of Section 5.6.3. 22 of the org–Ks ICCF and 13 of the ad–sub–Ks ICCF satisfied conditions 1 and 2. And, 10 of the org–Ks ICCF and eight of the ad–sub–Ks ICCF satisfied conditions 1, 2 and 3.

There are five sources with ad–sub–Ks ICCF time lags consistent with zero. Two of them have X–ray luminosities higher than $10^{44.5}$ erg/s, for which the predicted time lag using the relation of Koshida et al. (2014) is $\Delta t > 820$ days, which is too large to be detected with the UltraVISTA light curves. Two have light curves that are very monotonous, which produce poor CC results. And one has $z_{spec} \sim 0.55$, thus the Y band light curve can be contaminated by H α emission. We solved this problem by using the J band to describe the accretion disk emission for this object, as we discuss below.

In Figure 5.5 we compared the eight ad–sub–Ks ICCF time lags that satisfied the three

conditions of Section 5.6.3, with their correspondent org–Ks ICCF time lags. It is clear from the figure that after the accretion disk component subtraction, the computed time lags are larger. Six of these eight time lags are associated to sources with SEDs dominated by the torus, and two have SEDs dominated by the accretion disk. There are two org–Ks ICCF time lags that satisfied all the conditions of Section 5.6.3, but whose correspondent ad–sub–Ks ICCF time lags did not satisfy the conditions. We inspected the Ks light curves obtained after the accretion disk component subtraction of these two sources, and noticed that they are very noisy, which might explain the poor results of the ad–sub–Ks ICCF time lags for these sources. One of these sources has an SED dominated by the torus, and the other one has an SED dominated by the accretion disk. For the rest of the analysis, we consider the eight ad–sub–Ks ICCF time lags that satisfied the three conditions of Section 5.6.3, we called this sub–sample as the “good–lag sample”.

To complement the analysis done with the Y and Ks light curves, we also computed time lags for the J and Ks light curves, since one of the problems of the DRM analysis is the contamination of BLR emission in the optical light curves at some redshifts ranges. In particular, the Y band can present contamination from $H\alpha$ emission at $z_{spec} \sim 0.55$, and from $H\beta$ emission at $z_{spec} \sim 1.1$. On the other hand, the J band can present contamination from $H\alpha$ emission at $z_{spec} \sim 0.91$, and from $H\beta$ emission at $z_{spec} \sim 1.57$. Considering the effective wavelength and width of the Y band, ($1.021 \mu\text{m}$ and $0.093 \mu\text{m}$, respectively), we can say that sources with $1.0 < z_{spec} < 1.2$ or $0.45 < z_{spec} < 0.65$ can have Y band light curves contaminated with BLR emission. Three of the eight sources in the good–lag sample are in this redshift range, so their lag measurements can be affected by the BLR emission, and, as mentioned previously, one source with ad–sub–Ks ICCF time lag consistent with zero has $z_{spec} \sim 0.55$, therefore its Y band emission can be contaminated by $H\alpha$ emission, which might explain its poor CC results. For all these sources, we can use the time lags obtained with J and Ks light curves. Thus, we defined a “final lag” (τ), which correspond to the lag determined with the Y and Ks bands for sources with $z_{spec} < 1.04$, and with the J and Ks bands for sources with $z_{spec} \geq 1.04$. In particular, the results for the source with ad–sub–Ks ICCF time lag consistent with zero located at $z_{spec} \sim 0.55$ improves considerably when we use the J band, therefore we added this source to the good–lag sample.

In Table 5.1 we show the equatorial coordinates in degrees (J2000), the redshift, the dominant component of the SED, the rest–frame Hard band X–ray luminosity band from Marchesi et al. (2016a), the measured time delay for Y and Ks (in days and in the rest–frame), the measured time delay for J and Ks (in days and in the rest–frame), and the final lag (τ ; in days and in the rest–frame), for the nine sources in the good–lag sample.

5.7 Correlations between Time Lags and Optical Luminosity

Suganuma et al. (2006) and Koshida et al. (2014) found that $\Delta t \propto L^{0.5}$. In order to know whether the same result is obtained for sources with $0.5 < z_{spec} < 1.2$, we need information about the optical luminosity of our sources. Suganuma et al. (2006) and Koshida et al.

Table 5.1: Results for the good-lag sample

RA	DEC	z_{spec}	SED	$\log_{10}(L_{\text{HXR}})$	$\Delta t_{(Y-Ks)}$	$\Delta t_{(J-Ks)}$	τ
150.208206	1.87531	1.156	torus	43.732	$220.0^{+44.6}_{-83.3}$	$244.4^{+15.2}_{-76.6}$	$244.4^{+15.2}_{-76.6}$
150.443665	2.0491	0.669	torus	43.907	$115.8^{+17.5}_{-21.5}$	$79.5^{+12.6}_{-7.8}$	$115.8^{+17.5}_{-21.5}$
150.44957	2.24643	0.882	torus	43.907	$129.4^{+19.0}_{-10.8}$	$129.6^{+17.6}_{-15.0}$	$129.4^{+19.0}_{-10.8}$
150.462875	2.00906	0.967	disk	44.264	$136.8^{+13.7}_{-12.5}$	$103.5^{+15.2}_{-17.2}$	$136.8^{+13.7}_{-12.5}$
150.472031	2.41023	0.667	torus	43.591	$68.0^{+14.7}_{-16.8}$	$64.2^{+12.1}_{-13.7}$	$68.0^{+14.7}_{-16.8}$
150.194687	2.06789	0.554	torus	43.095	–	$72.3^{+19.2}_{-28.4}$	$72.3^{+19.2}_{-28.4}$
150.511566	2.40962	0.988	torus	43.881	$100.5^{+8.9}_{-10.9}$	$86.2^{+15.0}_{-12.1}$	$100.5^{+8.9}_{-10.9}$
150.544678	2.50728	1.161	disk	44.349	$174.8^{+52.3}_{-25.8}$	$205.0^{+17.8}_{-38.2}$	$205.0^{+17.8}_{-38.2}$
150.572586	2.49992	1.106	torus	44.269	$144.7^{+41.3}_{-30.5}$	$144.1^{+51.9}_{-28.2}$	$144.1^{+51.9}_{-28.2}$

(2014) obtained the absolute magnitude in the V band (M_V) directly from their images, by subtracting the host galaxy contribution to the total flux in the V band, and computing the average flux in a given season of observations. This technique requires to be able to decompose the source images, into the host galaxy and AGN component. However, since our objects look like point sources in the VISTA images, we could not perform an AGN–host galaxy decomposition.

Marchesi et al. (2016a) provides rest-frame Hard band X-ray luminosities (L_{HXR}). This information can be used to estimate the bolometric luminosity (L_{Bol}) of our sources (since the AGN dominates the Hard X-ray emission), using the bolometric corrections proposed by Marconi et al. (2004) (see their Equation 21). In particular, we solved the equation:

$$0.0015\ell^3 - 0.012\ell^2 + 0.76\ell + (10.46 - \log_{10}L_{\text{HXR}}) = 0,$$

where $\ell = \log_{10}L_{\text{Bol}} - 12$, with L_{Bol} corresponding to the bolometric luminosity. To compare our results with the results from Koshida et al. (2014), we used a bolometric correction of 10 to convert their L_V to L_{Bol} .

In Figure 5.6 we show the final radius (R_{DRM} ; in parsecs), derived from the final rest-frame lags, versus L_{Bol} . We also show the relation between R_{DRM} and L_{Bol} obtained from Koshida et al. (2014) for DRM of local sources, and the results of Bentz et al. (2013) for emission from the BLR ($\text{H}\beta$). We must consider that Koshida et al. (2014) findings are valid for local sources and for sources with $-16 < M_V < -22$, and seven of our nine sources are brighter than $M_V \sim -22$. Therefore our results are useful to test whether we can extrapolate the relation found by Koshida et al. (2014) to brighter sources, and whether it is valid for sources located at higher redshifts. On the other hand, BLR lags are well calibrated for our full range of L_{Bol} .

It can be seen in Figure 5.6 that most of the measured lags are below the relation found by Koshida et al. (2014), and above the relation found by Bentz et al. (2013). There are two sources with lags in agreement with the findings of Koshida et al. (2014), and seven sources with smaller lags than expected. This tells us that we are measuring time lags that are larger than what is expected for the BLR, but smaller than the findings for the dusty torus in the local universe. However, there is evidence of a correlation between L_{Bol} and R_{DRM} for these sources.

A possible explanation for these results is that variability is affecting the value of L_{Bol} when we determine it using the bolometric corrections of Marconi et al. (2004). However, since there is a systematic decrease of the measured lags versus the lags expected from the correlation found by Koshida et al. (2014), we can discard this explanation. Another explanation is that for local sources, either the dust in the torus sublimates at lower temperatures, the dust grains have a lower diameter, or the composition of the grains is different, compared to sources with $z > 0.5$, since, as we shown in Equation 5.1, $R_{sub} \propto L^{0.5} T_{sub}^{-2.8} a^{-0.5}$. We tested different values for T_{sub} and a , and found that for $T_{sub} = 1900$ K and $a = 0.55 \mu\text{m}$, we can reproduce our results. These values correspond to the extreme values expected for graphite grains in the Galactic interstellar medium (Mathis et al., 1977; Netzer, 2013). However, previous studies have proposed that large grains are requested to explain the infrared emission of some AGN (e.g., Maiolino et al. 2001b,a; Gaskell et al. 2004)

Mor & Trakhtenbrot (2011) and Mor & Netzer (2012) studied the SED of a large sample of narrow–line Seyfert 1 galaxies (NLS1) and broad–line Seyfert 1 galaxies (BLS1), and they found that to explain properly the NIR emission of most of their sources, they need a torus with a pure–graphite dust component, located outside the dust–free BLR and within the standard silicate–dust torus. Particularly, our results seems to be in agreement with the results of Mor & Netzer (2012) shown in their Figure 8. Therefore, our results suggest that a torus inner component with a pure–graphite dust composition is required to explain the lags obtained for luminous and high redshift sources.

One of the main implications of these findings is that the results obtained for local sources and sources located at $z > 0.5$ can be different. Therefore, if we want to use the time lag between optical and NIR continuum emission of AGN as standard candle for cosmology, we need to take into account that the correlations found for local sources can be systematically biased when we study high redshift sources, produced by differences in the dust properties of the torus. This makes it difficult to use of the time lag as standard candle, as our results seems to suggest a mix population of torus properties at $z \sim 1$.

We tested whether better results can be obtained when we use the luminosity of the variable component, instead of the total luminosity of each source, as proposed by Hönig et al. (2017). In Section 5.4 we used four different parameters to estimate the variability amplitude ($P_{95} - P_5$, $P_{90} - P_{10}$, Min–Max, and σ_f) to obtain the difference SED of each source. To obtain the luminosity of the variable component, we used the values of these four parameters derived from the Y band light curves, and converted them to rest–frame luminosity in the V band using the results for the spectral slope found by Selsing et al. (2016), where $\alpha_\lambda = 1.70 \pm 0.01$.

In Figure 5.7 we show the results for the final lag (τ) versus the variability amplitude (ΔL_V) computed using the different parameters. We computed the weighted least–squares linear regression (WLS) for each parameter versus τ , and we found that they are all consistent with $\tau \propto (\Delta L)^{0.3 \pm 0.1}$. This is somewhat lower to what is expected for the total luminosity ($\Delta t \propto L^{0.5}$). However in this case, the correlations seems to have less dispersion than the correlation between L_{Bol} and R_{DRM} . One of the advantages of using ΔL_V instead of L_{Bol} , is that ΔL_V can be easily derived from the light curves, making the use of the time lag as standard candle easier.

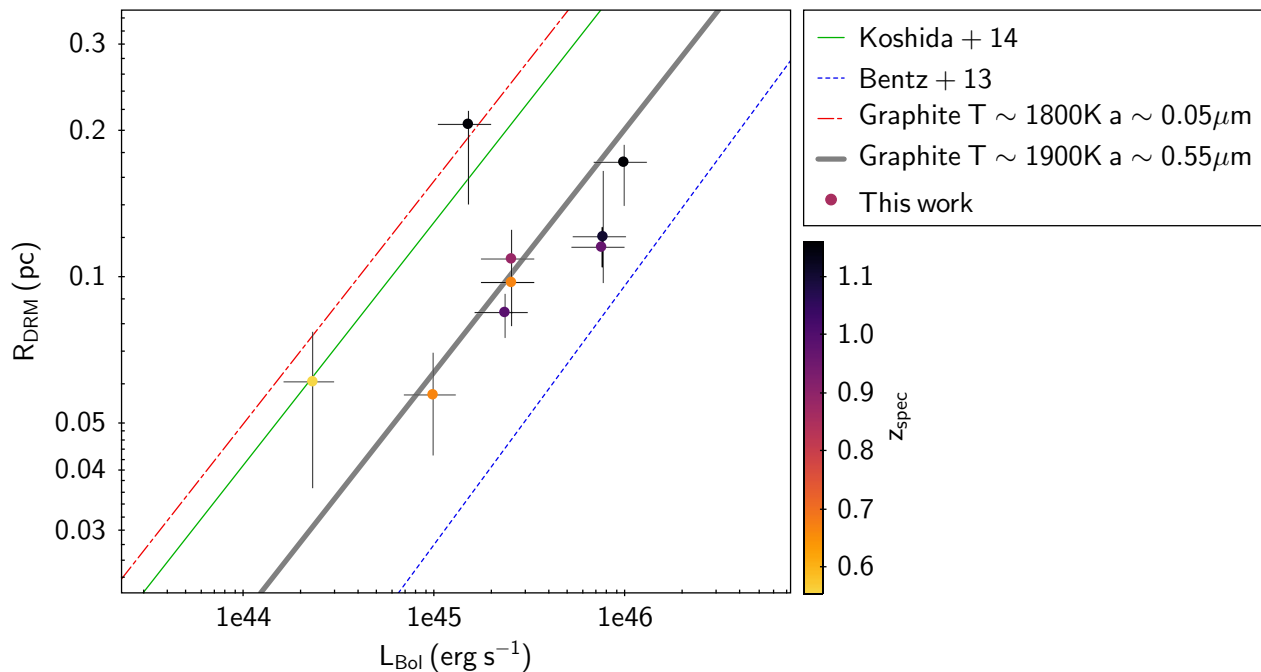


Figure 5.6: Final radius (R_{DRM} ; in parsecs) versus the bolometric luminosity (L_{Bol} ; in erg s^{-1}), obtained from the the Hard band X-ray luminosity, from Marchesi et al. (2016a). The green solid line show the relation obtained by Koshida et al. (2014). The blue dashed line show the result obtained by Bentz et al. (2013) for the BLR ($\text{H}\beta$). The red dot-dashed line shows the sublimation radius of a pure-graphite dust torus with $T_{\text{sub}} \sim 1800\text{ K}$, and a grain size of $a \sim 0.05\mu\text{m}$. And the grey solid line shows the sublimation radius of a pure-graphite dust torus with $T_{\text{sub}} \sim 1900\text{ K}$, and a grain size of $a \sim 0.55\mu\text{m}$.

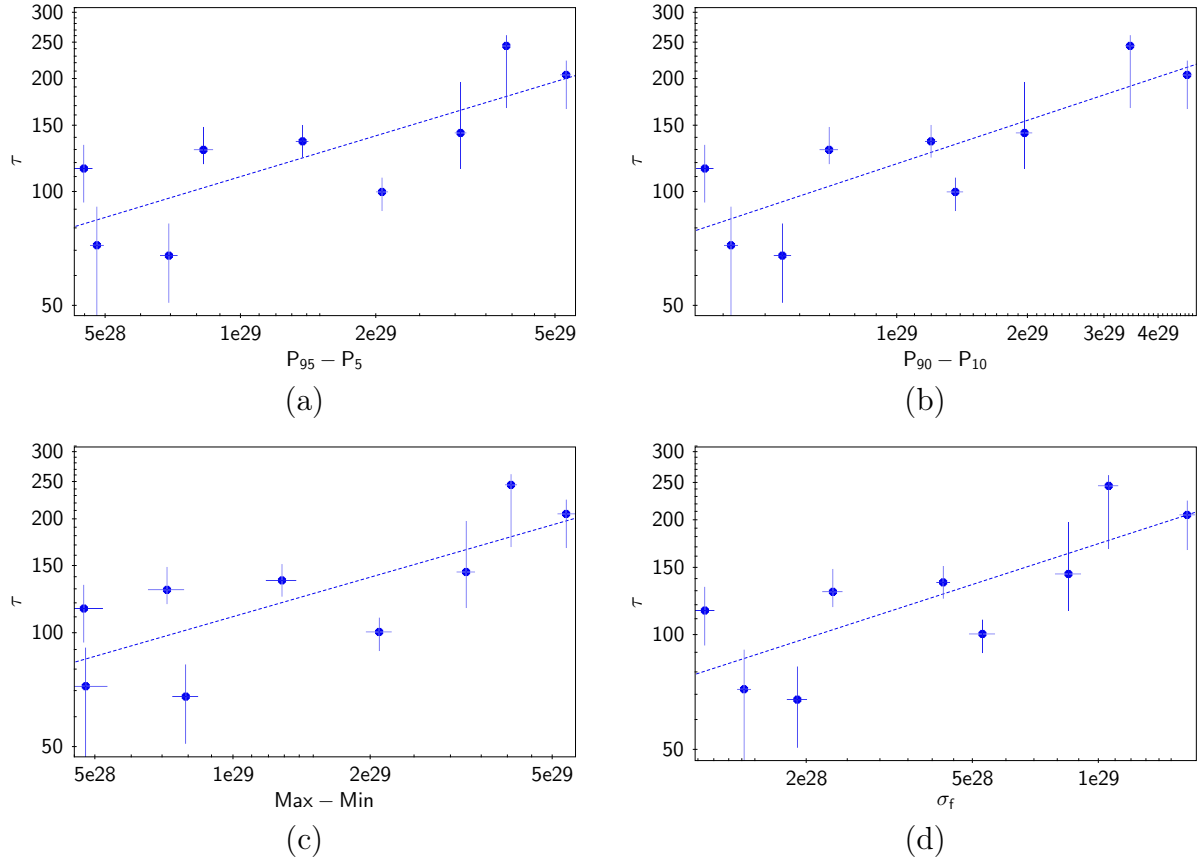


Figure 5.7: Final lag (τ ; in days) vs different variable component luminosity in the V band (ΔL_V ; in $\text{erg s}^{-1} \text{Hz}^{-1}$) measured using different techniques: a) $P_{95} - P_5$, b) $P_{90} - P_{10}$, c) Min-Max, d) σ_f . The blue dashed lines show the weighted least-squares linear regression (WLS).

5.8 Conclusions and Future Work

In this work we used a sample of 31 variable type 1 AGN, to test whether the emission in the NIR of sources at $0.5 < z < 1.2$ is consistent with emission from the dusty torus or the accretion disk, and whether the correlations between the time lag and the optical luminosity found by Suganuma et al. (2006) and Koshida et al. (2014), for local sources, is also seen when we consider sources with $0.5 < z < 1.2$.

From our SED analysis (see Section 5.4) we conclude that 10 of the 31 sources have a variable component dominated by the dusty torus, and 9 have a variable component dominated by the accretion disk. Most of the former sources have $z_{spec} > 0.9$, which might explain the lack of a variable component dominated by the torus.

To obtain the time lags between optical and NIR continuum emission, we used the Y and J light curves to describe the emission from the accretion disk, and the Ks light curves to describe the emission from the torus. We subtracted the accretion disk component from the Ks light curves, following the same approach of Koshida et al. (2014), and then performed a cross-correlation analysis, with ICCF (Peterson et al., 1998, 2004), following the procedure proposed by Grier et al. (2017) to obtain the time lags and their uncertainties. We could obtain secure time lag for nine sources.

We found that seven of our nine time lags (or torus radius) are not in agreement with the findings of Koshida et al. (2014). However, the $\Delta t \propto L^{0.5}$ relation is still present, but requires dust composed by larger grains compared to the typical grain size values used in the literature (e.g., Kishimoto et al. 2007; Koshida et al. 2014; Netzer 2015).

We also tested whether there is a correlation between variable component luminosity (ΔL_V) and the time lag obtained for our nine sources with secure time lags, and found that the variable component luminosity correlates with the time lag, with $\Delta t \propto \Delta L_V^{0.3 \pm 0.1}$.

The correlation found between ΔL_V and Δt seems to provide a better way to use the time lag as standard candle for cosmology. However, we need to test these results with a larger sample. We can increase our sample by including type 1 AGN with known redshifts, regardless of whether they have X-ray information, although the determination of AGN L_{Bol} might be challenging. This should not be hard to do, since there is a huge amount of ancillary data available for the COSMOS field (see Section 2.2.2). We will include an analysis with a larger sample in the paper associated to this work.

Chapter 6

Conclusions

Here I present the most relevant conclusions of this thesis which aims to improve our knowledge about the physics behind AGN variability, and to improve the selection of different AGN populations through variability analyses.

In Chapter 2, I presented our statistical study of NIR variability of X-ray selected AGN in the COSMOS field, using UltraVISTA data. This is the largest sample of AGN light curves in YJHKs bands, making possible to have a global description of the nature of AGN for a large range of redshifts, and for different levels of obscuration. To characterize the variability properties of the sources we computed the Structure Function. The main findings of this Chapter are:

- There is an anti-correlation between the Structure Function A parameter (variability amplitude) and the wavelength of emission, and a weak anti-correlation between A and the bolometric luminosity.
- Broad Line (BL) AGN have a considerably larger fraction of variable sources than Narrow Line (NL) AGN, and that they have different distributions of the A parameter.
- We find evidence that suggests that most of the low luminosity variable NL sources correspond to BL AGN, where the host galaxy could be damping the variability signal.
- For high luminosity variable NL, we propose that they can be examples of “True type II” AGN or BL AGN with limited spectral coverage which results in missing the Broad Line emission.
- The fraction of variable sources classified as unobscured in the X-ray is smaller than the fraction of variable sources unobscured in the optical range. We present evidence that this is related to the differences in the origin of the obscuration in the optical and X-ray regimes.

In Chapter 3, I presented our statistical analysis of the connection between active galactic nuclei (AGN) variability and physical properties of the central SMBH. We constructed optical light curves using data from the QUEST-La Silla survey. To model the variability, we used the structure function, among the excess variance and the amplitude from DRW modeling. For the measurement of SMBH physical properties, we used public spectra from SDSS. Our analysis is based on an original sample of 2345 sources detected in both SDSS and QUEST-

La Silla. For 1473 of these sources we could perform a proper measurement of the spectral and variability properties, and 1348 of these sources were classified as variable (91.5%). The main findings of this Chapter are:

- The amplitude of the variability (A) depends solely on the rest frame emission wavelength and the Eddington ratio, where A anti-correlates with both λ_{rest} and L/L_{Edd} . This suggests that AGN variability does not evolve over cosmic time, and its amplitude is inversely related to the accretion rate.
- The logarithmic gradient of the variability (γ) does not correlate significantly with any SMBH physical parameter, since there is no statistically significant linear regression model with an absolute value of the slope higher than 0.1.
- The general distribution of γ measured for our sample differs from the distribution of γ obtained for light curves simulated from a DRW process. For 20.6% of the variable sources in our sample, a DRW model is not appropriate to describe the variability, since γ differs considerably from the expected value of 0.5.

In Chapter 4, I presented our variability-based AGN selection technique. We used data from the QUEST–La Silla Active Galactic Nuclei (AGN) variability survey to construct light curves for 208,583 sources over ~ 70 deg², with a limiting magnitude $r \sim 21$. Each light curve has at least 40 epochs and a length of ≥ 200 days. We implemented a Random Forest algorithm to classify our objects as either AGN or non-AGN according to their variability features and optical colors, excluding morphology cuts. We tested three classifiers, one that only includes variability features (RF1), one that includes variability features and also $r - i$ and $i - z$ colors (RF2), and one that includes variability features and also $g - r$, $r - i$, and $i - z$ colors (RF3). We obtained a sample of high probability candidates (hp-AGN) for each classifier, with 5,941 candidates for RF1, 5,252 candidates for RF2, and 4,482 candidates for RF3. We divided each sample according to their $g - r$ colors, defining blue ($g - r \leq 0.6$) and red sub-samples ($g - r > 0.6$). We find that most of the candidates known from the literature belong to the blue sub-samples, which is not necessarily surprising given that, unlike for many literature studies, we do not cut our sample to point-like objects. This means that we can select AGN that have a significant contribution from redshifted starlight in their host galaxies. In order to test the efficiency of our technique we performed spectroscopic follow-up, confirming the AGN nature of 44 among 54 observed sources (81.5% of efficiency). From the campaign we concluded that RF2 provides the purest sample of AGN candidates.

Finally, in Chapter 5, I presented our Dust Reverberation Mapping analysis, done to determine the characteristic optical–NIR time lags (Δt), to study whether the emission received in the NIR of high redshift sources is consistent with emission from the dusty torus or the accretion disk, and to determine whether we can use Δt as a standard candle. We used NIR data from the UltraVISTA survey to construct YJHKs light curves for X-ray selected type 1 AGN. We constructed light curves for 79 sources, and 31 of these were classified as variable in the four NIR bands. From our SED analysis, we concluded that 10 sources have a variable component dominated by the dusty torus, and 9 have a variable component dominated by the accretion disk. We performed a cross-correlation analysis with ICCF (Peterson et al., 1998, 2004), and determine secure optical–NIR time lags for nine sources. We estimated bolometric luminosities from X-ray data Marchesi et al. (2016a), and we found that in general our results are not in agreement with what was found by previous studies (e.g., Suganuma

et al. 2006; Koshida et al. 2014), since we require larger grains compared to the typical grain size values used in the literature. We also tested whether the use of the variable component luminosity (ΔL_V) provides better results, finding that $\Delta t \propto \Delta L_V^{0.3 \pm 0.1}$. The results presented in this Chapter are preliminary, and will be included in a forthcoming paper.

Bibliography

- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ArXiv e-prints, arXiv:1801.03181
- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
- Allevato, V., Paolillo, M., Papadakis, I., & Pinto, C. 2013, ApJ, 771, 9
- Angione, R. J., & Smith, H. J. 1972, in IAU Symposium, Vol. 44, External Galaxies and Quasi-Stellar Objects, ed. D. S. Evans, D. Wills, & B. J. Wills, 171
- Antonucci, R. R. J., & Miller, J. S. 1985, ApJ, 297, 621
- Arévalo, P., Uttley, P., Kaspi, S., et al. 2008, MNRAS, 389, 1479
- Arévalo, P., Uttley, P., Lira, P., et al. 2009, MNRAS, 397, 2004
- Barvainis, R. 1987, ApJ, 320, 537
- Bauer, A., Baltay, C., Coppi, P., et al. 2009, ApJ, 696, 1241
- Becker, R. H., White, R. L., & Helfand, D. J. 1994, in Astronomical Society of the Pacific Conference Series, Vol. 61, Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes, 165
- Bellm, E. 2014, in The Third Hot-wiring the Transient Universe Workshop, ed. P. R. Wozniak, M. J. Graham, A. A. Mahabal, & R. Seaman, 27–33
- Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, ApJ, 767, 149
- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
- Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1
- Bondi, M., Ciliegi, P., Schinnerer, E., et al. 2008, ApJ, 681, 1129
- Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, ApJ, 729, 141
- Braccesi, A., Formigini, L., & Gandolfi, E. 1970, A&A, 5, 264
- Breedt, E., Arévalo, P., McHardy, I. M., et al. 2009, MNRAS, 394, 427

- Breedt, E., McHardy, I. M., Arévalo, P., et al. 2010, *MNRAS*, 403, 605
- Breiman, L. 2001, *Machine Learning*, 45, 5
- Brusa, M., Civano, F., Comastri, A., et al. 2010, *ApJ*, 716, 348
- Buisson, D. J. K., Lohfink, A. M., Alston, W. N., & Fabian, A. C. 2017, *MNRAS*, 464, 3194
- Burtscher, L., Davies, R. I., Graciá-Carpio, J., et al. 2016, *A&A*, 586, A28
- Butler, N. R., & Bloom, J. S. 2011, *AJ*, 141, 93
- Buzzoni, B., Delabre, B., Dekker, H., et al. 1984, *The Messenger*, 38, 9
- Caplar, N., Lilly, S. J., & Trakhtenbrot, B. 2017, *ApJ*, 834, 111
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Cartier, R., Lira, P., Coppi, P., et al. 2016, *The Messenger*, 163, 26
- . 2015, *ApJ*, 810, 164
- Chen, C.-T. J., Brandt, W. N., Luo, B., et al. 2018, *MNRAS*, 478, 2132
- Choi, Y., Gibson, R. R., Becker, A. C., et al. 2014, *ApJ*, 782, 37
- Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
- Clemens, J. C., Crain, J. A., & Anderson, R. 2004, in *Proc. SPIE*, Vol. 5492, *Ground-based Instrumentation for Astronomy*, ed. A. F. M. Moorwood & M. Iye, 331–340
- Collier, S., & Peterson, B. M. 2001, *ApJ*, 555, 775
- Covey, K. R., Ivezić, Ž., Schlegel, D., et al. 2007, *AJ*, 134, 2398
- Cristiani, S., Trentini, S., La Franca, F., & Andreani, P. 1997, *A&A*, 321, 123
- Cristiani, S., Trentini, S., La Franca, F., et al. 1996, *A&A*, 306, 395
- Croom, S. M., Rhook, K., Corbett, E. A., et al. 2002, *MNRAS*, 337, 275
- Czerny, B., Doroshenko, V. T., Nikołajuk, M., et al. 2003, *MNRAS*, 342, 1222
- Dalton, G. B., Lewis, I. J., Bonfield, D. G., et al. 2006, in *Proc. SPIE*, Vol. 6269, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 62694A
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, 145, 10
- de Vries, W. H., Becker, R. H., White, R. L., & Loomis, C. 2005, *AJ*, 129, 615
- Denney, K. D., De Rosa, G., Croxall, K., et al. 2014, *ApJ*, 796, 134

Diamond-Stanic, A. M., Fan, X., Brandt, W. N., et al. 2009, *ApJ*, 699, 782

Doi, M., Tanaka, M., Fukugita, M., et al. 2010, *AJ*, 139, 1628

Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, 696, 870

Edelson, R., & Nandra, K. 1999, *ApJ*, 514, 682

Edelson, R., Gelbord, J. M., Horne, K., et al. 2015, *ApJ*, 806, 129

Edelson, R., Gelbord, J., Cackett, E., et al. 2019, *ApJ*, 870, 123

Elitzur, M., & Netzer, H. 2016, *MNRAS*, 459, 585

Emerson, J., McPherson, A., & Sutherland, W. 2006, *The Messenger*, 126, 41

Emerson, J. P., Irwin, M. J., Lewis, J., et al. 2004, in *Proc. SPIE*, Vol. 5493, *Optimizing Scientific Return for Astronomy through Information Technologies*, ed. P. J. Quinn & A. Bridger, 401–410

Emmanoulopoulos, D., McHardy, I. M., & Uttley, P. 2010, *MNRAS*, 404, 931

Enya, K., Yoshii, Y., Kobayashi, Y., et al. 2002a, *ApJS*, 141, 23

—. 2002b, *ApJS*, 141, 31

—. 2002c, *ApJS*, 141, 45

Evans, P. A., Osborne, J. P., Beardmore, A. P., et al. 2014, *ApJS*, 210, 8

Fan, X. 1999, *AJ*, 117, 2528

Fausnaugh, M. M., Denney, K. D., Barth, A. J., et al. 2016, *ApJ*, 821, 56

Feruglio, C., Fiore, F., La Franca, F., et al. 2008, *A&A*, 488, 417

Flesch, E. W. 2015, *PASA*, 32, e010

Francis, P. J., & Wills, B. J. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 162, *Quasars and Cosmology*, ed. G. Ferland & J. Baldwin, 363

F.R.S., K. P. 1901, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559

Gaskell, C. M., Goosmann, R. W., Antonucci, R. R. J., & Whysong, D. H. 2004, *ApJ*, 616, 147

Gawiser, E., van Dokkum, P. G., Herrera, D., et al. 2006, *ApJS*, 162, 1

Gehrels, N. 1986, *ApJ*, 303, 336

- Ginsburg, A., & Mirocha, J. 2011, PySpecKit: Python Spectroscopic Toolkit, Astrophysics Source Code Library, , , ascl:1109.001
- Giveon, U., Maoz, D., Kaspi, S., Netzer, H., & Smith, P. S. 1999, MNRAS, 306, 637
- Glass, I. S. 1992, MNRAS, 256, 23P
- Graham, M. J., Djorgovski, S. G., Drake, A. J., et al. 2014, MNRAS, 439, 703
- Greene, J. E., & Ho, L. C. 2005, ApJ, 630, 122
- Grier, C. J., Trump, J. R., Shen, Y., et al. 2017, ApJ, 851, 21
- Grupe, D., Komossa, S., Leighly, K. M., & Page, K. L. 2010, ApJS, 187, 64
- Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
- Haas, M. 2004, in IAU Symposium, Vol. 222, The Interplay Among Black Holes, Stars and ISM in Galactic Nuclei, ed. T. Storchi-Bergmann, L. C. Ho, & H. R. Schmitt, 267–270
- Heckman, T. M., & Best, P. N. 2014, ARA&A, 52, 589
- Hernán-Caballero, A., Hatziminaoglou, E., Alonso-Herrero, A., & Mateos, S. 2016, MNRAS, 463, 2064
- Hönig, S. F., Watson, D., Kishimoto, M., et al. 2017, MNRAS, 464, 1693
- Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305
- Hopkins, P. F., Hernquist, L., Martini, P., et al. 2005, ApJ, 625, L71
- Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, in Proc. SPIE, Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies, ed. P. J. Quinn & A. Bridger, 411–422
- Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, ArXiv e-prints, arXiv:0805.2366
- Jin, C., Ward, M., & Done, C. 2012, MNRAS, 425, 907
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
- Kasliwal, V. P., Vogeley, M. S., & Richards, G. T. 2015, MNRAS, 451, 4328
- . 2017, MNRAS, 470, 3027
- Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green, R. 1989, AJ, 98, 1195
- Kelly, B. C. 2007, ApJ, 665, 1489
- Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895

- Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, *ApJ*, 788, 33
- Kelly, B. C., Treu, T., Malkan, M., Pancoast, A., & Woo, J.-H. 2013, *ApJ*, 779, 187
- Kim, D.-W., Protopapas, P., Bailer-Jones, C. A. L., et al. 2014, *A&A*, 566, A43
- Kim, D.-W., Protopapas, P., Byun, Y.-I., et al. 2011, *ApJ*, 735, 68
- Kim, M., Ho, L. C., & Im, M. 2006, *ApJ*, 642, 702
- Kirkpatrick, J. A., Schlegel, D. J., Ross, N. P., et al. 2011, *ApJ*, 743, 125
- Kishimoto, M., Hönig, S. F., Beckert, T., & Weigelt, G. 2007, *A&A*, 476, 713
- Kollatschny, W., & Zetzl, M. 2013, *A&A*, 558, A26
- Koshida, S., Yoshii, Y., Kobayashi, Y., et al. 2009, *ApJ*, 700, L109
- Koshida, S., Minezaki, T., Yoshii, Y., et al. 2014, *ApJ*, 788, 159
- Kozłowski, S. 2016, *ApJ*, 826, 118
- . 2017a, *A&A*, 597, A128
- . 2017b, *ApJS*, 228, 9
- Kozłowski, S., Kochanek, C. S., Udalski, A., et al. 2010, *ApJ*, 708, 927
- Krolik, J. H., Horne, K., Kallman, T. R., et al. 1991, *ApJ*, 371, 541
- Kubota, A., & Done, C. 2018, ArXiv e-prints, arXiv:1804.00171
- Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, *ApJS*, 154, 166
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, *ApJ*, 800, 144
- Landt, H., Elvis, M., Ward, M. J., et al. 2011, *MNRAS*, 414, 218
- Lanzuisi, G., Ponti, G., Salvato, M., et al. 2014, *ApJ*, 781, 105
- Lewis, J. R., Irwin, M., & Bunclark, P. 2010, in *Astronomical Society of the Pacific Conference Series*, Vol. 434, *Astronomical Data Analysis Software and Systems XIX*, ed. Y. Mizumoto, K.-I. Morita, & M. Ohishi, 91
- Li, Z., McGreer, I. D., Wu, X.-B., Fan, X., & Yang, Q. 2018, *ApJ*, 861, 6
- Lira, P., Arévalo, P., Uttley, P., McHardy, I., & Breedt, E. 2011, *MNRAS*, 415, 1290

- Lira, P., Arévalo, P., Uttley, P., McHardy, I. M. M., & Videla, L. 2015, *MNRAS*, 454, 368
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
- Luo, B., Brandt, W. N., Hall, P. B., et al. 2015, *ApJ*, 805, 122
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, *ApJS*, 228, 2
- Lusso, E., Comastri, A., Vignali, C., et al. 2010, *A&A*, 512, A34
- Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, *MNRAS*, 425, 623
- Lynden-Bell, D. 1969, *Nature*, 223, 690
- MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, *ApJ*, 721, 1014
- MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, *ApJ*, 728, 26
- Maiolino, R., Marconi, A., & Oliva, E. 2001a, *A&A*, 365, 37
- Maiolino, R., Marconi, A., Salvati, M., et al. 2001b, *A&A*, 365, 28
- Maitra, C., Haberl, F., Ivanov, V. D., Cioni, M.-R. L., & van Loon, J. T. 2019, *A&A*, 622, A29
- Mandal, A. K., Rakshit, S., Kurian, K. S., et al. 2018, *MNRAS*, 475, 5330
- Marchesi, S., Civano, F., Elvis, M., et al. 2016a, *ApJ*, 817, 34
- Marchesi, S., Lanzuisi, G., Civano, F., et al. 2016b, *ApJ*, 830, 100
- Marconi, A., Risaliti, G., Gilli, R., et al. 2004, *MNRAS*, 351, 169
- Martini, P., & Schneider, D. P. 2003, *ApJ*, 597, L109
- Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, *ApJ*, 217, 425
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- McHardy, I. M., Koeding, E., Knigge, C., Uttley, P., & Fender, R. P. 2006, *Nature*, 444, 730
- McHardy, I. M., Cameron, D. T., Dwelly, T., et al. 2014, *MNRAS*, 444, 1469
- McHardy, I. M., Connolly, S. D., Peterson, B. M., et al. 2016, *Astronomische Nachrichten*, 337, 500
- McHardy, I. M., Connolly, S. D., Horne, K., et al. 2018, *MNRAS*, 480, 2881
- McLaughlin, M. A., Mattox, J. R., Cordes, J. M., & Thompson, D. J. 1996, *ApJ*, 473, 763

- Medina, G. E., Muñoz, R. R., Vivas, A. K., et al. 2018, *ApJ*, 855, 43
- Mejía-Restrepo, J. E., Lira, P., Netzer, H., Trakhtenbrot, B., & Capellupo, D. M. 2018a, *Nature Astronomy*, 2, 63
- Mejía-Restrepo, J. E., Trakhtenbrot, B., Lira, P., & Netzer, H. 2018b, *MNRAS*, arXiv:1805.00942
- Mejía-Restrepo, J. E., Trakhtenbrot, B., Lira, P., Netzer, H., & Capellupo, D. M. 2016, *MNRAS*, 460, 187
- Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, *MNRAS*, 437, 3550
- Meusinger, H., & Weiss, V. 2013, *A&A*, 560, A104
- Milvang-Jensen, B., Freudling, W., Zabl, J., et al. 2013, *A&A*, 560, A94
- Mor, R., & Netzer, H. 2012, *MNRAS*, 420, 526
- Mor, R., & Trakhtenbrot, B. 2011, *ApJ*, 737, L36
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, *ApJS*, 206, 8
- Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, *ApJS*, 221, 27
- Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, *ApJ*, 476, 70
- Narayan, R., & Yi, I. 1994, *ApJ*, 428, L13
- . 1995, *ApJ*, 444, 231
- Netzer, H. 2013, *The Physics and Evolution of Active Galactic Nuclei*
- . 2015, *ARA&A*, 53, 365
- Netzer, H., & Trakhtenbrot, B. 2014, *MNRAS*, 438, 672
- Neugebauer, G., Soifer, B. T., Matthews, K., & Elias, J. H. 1989, *AJ*, 97, 957
- Nun, I., Protopapas, P., Sim, B., et al. 2015, *ArXiv e-prints*, arXiv:1506.00010
- Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, *A&A*, 530, A122
- Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, *A&A*, 587, A41
- Panessa, F., & Bassani, L. 2002, *A&A*, 394, 435
- Paolillo, M., Schreier, E. J., Giacconi, R., Koekemoer, A. M., & Grogin, N. A. 2004, *ApJ*, 611, 93

Pâris, I., Petitjean, P., Aubourg, E., et al. 2017a, ArXiv e-prints, arXiv:1712.05029

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017b, *A&A*, 597, A79

Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, *ApJ*, 811, 95

Peterson, B. M. 1993, *PASP*, 105, 247

Peterson, B. M. 2001, in *Advanced Lectures on the Starburst-AGN*, ed. I. Aretxaga, D. Kunth, & R. Mújica, 3

Peterson, B. M., Wanders, I., Horne, K., et al. 1998, *PASP*, 110, 660

Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, *ApJ*, 613, 682

Pettitt, A. N. 1976, *Biometrika*, 63, 161

Rahman, M., Mendez, A. J., Ménard, B., et al. 2016, *MNRAS*, 460, 163

Rakshit, S., & Stalin, C. S. 2017, *ApJ*, 842, 96

Ramos Almeida, C., & Ricci, C. 2017, *Nature Astronomy*, 1, 679

Rengstorf, A. W., Brunner, R. J., & Wilhite, B. C. 2006, *AJ*, 131, 1923

Rengstorf, A. W., Mufson, S. L., Andrews, P., et al. 2004, *ApJ*, 617, 184

Ricci, C., Bauer, F. E., Arevalo, P., et al. 2016, *ApJ*, 820, 5

Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, *AJ*, 123, 2945

Richards, G. T., Nichol, R. C., Gray, A. G., et al. 2004, *ApJS*, 155, 257

Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, *ApJS*, 180, 67

Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, *ApJ*, 733, 10

Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, *ApJS*, 199, 3

Ross, N. P., Ford, K. E. S., Graham, M., et al. 2018, *MNRAS*, 480, 4468

Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250

Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61

Samus', N. N., Kazarovets, E. V., Durlevich, O. V., Kireeva, N. N., & Pastukhova, E. N. 2017, *Astronomy Reports*, 61, 80

Sánchez, P., Lira, P., Cartier, R., et al. 2017, *ApJ*, 849, 110

Sánchez-Sáez, P., Lira, P., Mejía-Restrepo, J., et al. 2018, *ApJ*, 864, 87

Schinnerer, E., Carilli, C. L., Scoville, N. Z., et al. 2004, *AJ*, 128, 1974

Schinnerer, E., Smolčić, V., Carilli, C. L., et al. 2007, *ApJS*, 172, 46

Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, *ApJS*, 188, 384

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525

Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. 2010, *ApJ*, 714, 1194

Schmidt, M., & Green, R. F. 1983, *ApJ*, 269, 352

Secrest, N. J., Dudik, R. P., Dorland, B. N., et al. 2015, *ApJS*, 221, 12

Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J.-K. 2016, *A&A*, 585, A87

Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, *AJ*, 134, 2236

Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337

Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, *ApJ*, 788, 48

Shemmer, O., Brandt, W. N., Anderson, S. F., et al. 2009, *ApJ*, 696, 580

Shemmer, O., Brandt, W. N., Netzer, H., Maiolino, R., & Kaspi, S. 2008, *ApJ*, 682, 81

Shemmer, O., & Lieber, S. 2015, *ApJ*, 805, 124

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, *ApJ*, 680, 169

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, *ApJS*, 194, 45

Simm, T., Salvato, M., Saglia, R., et al. 2016, *A&A*, 585, A129

Smartt, S. J., Valenti, S., Fraser, M., et al. 2015, *A&A*, 579, A40

Smith, K. L., Mushotzky, R. F., Boyd, P. T., et al. 2018, *ArXiv e-prints*, arXiv:1803.06436

Smith, R. J., Croom, S. M., Boyle, B. J., et al. 2005, *MNRAS*, 359, 57

Smolčić, V., Ciliegi, P., Jelić, V., et al. 2014, *MNRAS*, 443, 2590

Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163

Stern, D., McKernan, B., Graham, M. J., et al. 2018, *ApJ*, 864, 27

Suberlak, K., Ivezić, Ž., MacLeod, C. L., Graham, M., & Sesar, B. 2017, *MNRAS*, 472, 4870

Suganuma, M., Yoshii, Y., Kobayashi, Y., et al. 2006, *ApJ*, 639, 46

- Sun, M., Grier, C. J., & Peterson, B. M. 2018, PyCCF: Python Cross Correlation Function for reverberation mapping studies, Astrophysics Source Code Library, , , ascl:1805.032
- Sutherland, W., Emerson, J., Dalton, G., et al. 2015, *A&A*, 575, A25
- Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, *ApJS*, 172, 9
- Taniguchi, Y., Kajisawa, M., Kobayashi, M. A. R., et al. 2015, *PASJ*, 67, 104
- Tie, S. S., Martini, P., Mudd, D., et al. 2017, *AJ*, 153, 107
- Timmer, J., & Koenig, M. 1995, *A&A*, 300, 707
- Tohline, J. E., & Osterbrock, D. E. 1976, *ApJ*, 210, L117
- Trevese, D., Kron, R. G., Majewski, S. R., Bershad, M. A., & Koo, D. C. 1994, *ApJ*, 433, 494
- Troyer, J., Starkey, D., Cackett, E. M., et al. 2016, *MNRAS*, 456, 4040
- Turner, T. J., George, I. M., Nandra, K., & Turcan, D. 1999, *ApJ*, 524, 667
- Uomoto, A. K., Wills, B. J., & Wills, D. 1976, *AJ*, 81, 905
- Uttley, P., Edelson, R., McHardy, I. M., Peterson, B. M., & Markowitz, A. 2003, *ApJ*, 584, L53
- van Dokkum, P. G. 2001, *Publications of the Astronomical Society of the Pacific*, 113, 1420
- Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, *ApJ*, 601, 692
- VanderPlas, J., Connolly, A. J., Ivezić, Z., & Gray, A. 2012, in *Proceedings of Conference on Intelligent Data Understanding (CIDU)*, pp. 47-54, 2012., 47–54
- VanderPlas, J. T. 2018, *ApJS*, 236, 16
- Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & vanden Berk, D. E. 2008, *MNRAS*, 383, 1232
- Wold, M., Brotherton, M. S., & Shang, Z. 2007, *MNRAS*, 375, 989
- Wu, J., Brandt, W. N., Anderson, S. F., et al. 2012, *ApJ*, 747, 10
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Young, M., Brandt, W. N., Xue, Y. Q., et al. 2012, *ApJ*, 748, 124