



SAPIENZA
UNIVERSITÀ DI ROMA



PH.D. IN ASTRONOMY, ASTROPHYSICS AND SPACE
SCIENCE

CYCLE XXXIII

Variability selected AGNs in the Chandra Deep Field South region

Andrea Bonchi

A.Y. 2020/2021

Supervisor: Dr. Andrea Grazian
Coordinator: Prof. Paolo De Bernardis
Deputy Coordinator: Prof. Nicola Vittorio

Contents

1	Introduction	5
1.1	What is the Reionization Era	5
1.1.1	When was the Universe re-ionized?	6
1.1.2	A global view	13
1.1.3	An analytic picture	15
1.2	Who reionized the Universe?	17
1.2.1	Star Forming Galaxies	19
1.2.2	AGNs	22
1.2.3	How to find high redshift AGNs	24
1.3	The adopted strategy	29
1.3.1	The Chandra Deep Field South (CDF-S)	29
1.3.2	The Visible Multi-Object Spectrograph (VIMOS) instrument	31
2	The Data Reduction	33
2.1	Input data	33
2.1.1	Reorganize input data	34
2.1.2	Epoch selection	35
2.2	Pipeline Description	36
2.2.1	General Structure	37
2.2.2	Overview	38
2.3	Some more details on the pipeline	42
2.3.1	Calibration Data creation	42
2.3.2	Instrument Calibration	45
2.3.3	Masking	47
2.3.4	Sky subtraction	49
2.3.5	Stacking Images	52
2.4	Output Products	55

3	Detection of variables AGNs	57
3.1	Catalogue extraction	57
3.1.1	Photometric calibration	58
3.1.2	Stars and Galaxies separation	61
3.2	Creation of the catalogue for the analysis	66
3.2.1	Catalogue instrumental magnitude limit	68
3.2.2	Selection for number of detections	70
3.3	MAD calculation and variables selection	71
3.4	MAD* and the binning method	73
3.4.1	obj_num: variable bin with fixed number of sources	74
3.4.2	mag_step: fixed bin and variable source selected number	76
3.4.3	obj_num and mag_step: a mixed method	78
3.5	Catalogue of variables	80
4	Result analysis	82
4.1	Visual inspection	83
4.2	Comparison with variability-selected AGNs from literature	85
4.2.1	Reference Catalogues	85
4.2.2	Cross-match results	90
4.2.3	Characterization of common variable sources	96
4.3	Comparison with X-ray catalogues	96
4.4	Comparison with multi-wavelength survey catalogues	99
4.5	Variables catalogue characterization	101
4.6	Some examples of variables from this work	103
4.6.1	Some bright known variable AGNs: the 9795 and the 18162	103
4.6.2	High redshift variables: the 15548 and the 15936	106
5	Summary & Conclusions	109
5.1	A brief summary	109
5.2	Results & Conclusions	112
5.3	What about the future	114
A	Calibration	119
A.1	<i>SExtractor</i> configuration file	119
B	Catalogues	123
B.1	Catalogue of Variables: main measures	123
B.2	Catalogue of Variables: cross-match results and characterization	126
	Bibliography	128

Abstract

The Reionization of the Universe is a landmark event in the history of cosmological structure formation and marks the end of the Dark Ages and the beginning of the so called Cosmic Dawn.

Measurements of Gunn-Peterson effects on luminous $z > 6$ quasar spectra (Fan et al. 2006; Becker et al. 2015; McGreer, Mesinger, and D’Odorico 2015) and of the Thomson optical depth $\tau = 0.054 \pm 0.007$ from the CMB (Planck Collaboration et al. 2020b), set the boundaries of Reionization Era in the redshift interval $z \approx 6 - 8.5$.

It is generally thought that the Intergalactic Medium (IGM) is kept ionized by the integrated UV emission from Active Galactic Nuclei (AGNs) and Star Forming Galaxies (SFG), but the relative contribution of these sources as a function of cosmic epoch are poorly known at present.

The numerous population of high redshift SFGs suggested them as main candidates for the emission of the UV ionizing radiation. However, measures of low Lyman Continuum escape fraction $f_{esc}(LyC)$ (e.g. Grazian et al. 2017) for these sources questioned their capability to ionize the Universe.

AGNs have, indeed, a very high ionizing power, but their space density shows a steep decrease for $z > 3$.

The search for high- z AGNs at $z > 4$ has been carried out mainly through selections based on X-ray or UV-optical colors, but recent works (Schindler et al. 2019b; Schindler et al. 2019a) have shown that these criteria are not 100% complete, and they could be complemented by other selection criteria, e.g. based on variability.

I developed a variability-selection method, based on the statistical parameter Median Absolute Deviation (MAD), to search for a population of faint AGNs that have not been detected by other selection techniques.

I used all available observations of Chandra Deep Field South (CDF-S) obtained with the VIMOS/VLT instrument in R -band to create a set of

stacked images, one for each year, with a time coverage extending from year 2007 to 2015. The R -band dropout limits the exploitation of this method to the $z < 5$ redshift boundary, after Reionization has been completed, but it shows its potential in a well known and well studied field as the CDF-S.

I extracted a catalogue from each of the single epoch images and compared them to each other. I could then study the magnitude variability of each source among different epochs.

The method described here selected a sample of 73 variable sources in the CDF-S brighter than a 90% completeness limit of $R \leq 26$ magnitude.

Five of these variable sources were previously identified by other four variability-based catalogues from the recent literature. Small overlaps were also found by comparing reference catalogues to each other. This confirms a strong dependency of variability-selection results from the method characteristics such as the photometric bands observed, the sampling time-scale and the statistical methods adopted. Therefore, results obtained via a single instrument/dataset cannot be considered exhaustive for variable objects selection.

A small overlap were also found between the catalogue of variable sources discussed in this work with X-ray surveys from *Chandra* and *XMM*. Only 10 of the 73 variable sources show a detectable X-ray emission.

The source characterization throughout a cross-match with the multi-wavelength surveys Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) and Multiwavelength Survey by Yale-Chile (MUSYC) provides a redshift value (spectroscopic or photometric) for the majority of variable sources of this work.

The distribution of source magnitudes with redshift shows that X-ray emitter sources are just the *tip of the iceberg* of a more numerous AGN population, selected through variability, that is typically fainter in the optical band. The optical variability is able to identify a large number of low-luminosity AGNs, especially at high redshifts, including those that remain undetectable even at the deepest observations in the X-rays.

New instruments for surveys are in their final phase of construction, as the *Vera Rubin Observatory* and *Euclid*. Variability will be one of the most powerful tool to detect AGNs in their data.

They will provide multi-epoch observations in the optical and near infrared bands, that are needed to explore up to $z \sim 7$, at the Reionization epoch.

Chapter 1

Introduction

1.1 What is the Reionization Era

According to the current Λ CDM cosmological model, after the nucleosynthetic yields of the Big Bang, the diffuse hydrogen was hot, fully ionized and optically thick due the Thomson scattering.

The subsequent *Recombination phase* decoupled neutral hydrogen from the photon field, leaving a last scattered light, the cosmic microwave background (CMB). The cosmic gas “recombined” and became neutral around $z \approx 1100$, when the Universe was 400,000 years old.

The formation of the first PopIII stars in minihalos at $z \sim 20 - 30$ (Trenti and Stiavelli 2009) and then the first galaxies at $z \sim 10$ (Barkana and Loeb 2001; Bromm and Yoshida 2011) started the emission of a UV background (UVB). The radiation transformed the cold neutral intergalactic hydrogen into highly ionized warm plasma. The resonant absorption in spectra of distant quasars (e.g. Fan et al. 2006; Becker et al. 2015; McGreer, Mesinger, and D’Odorico 2015) describes an IGM completely reionized at $z \sim 5 - 6$.

This phase, often identified also as *Cosmic Dawn*, marks the end of cosmic Dark Ages and the beginning of the Age of the Galaxies. It is what we identify as *Reionization Era*, a turning point in the history of structure formation in the Universe.

Figure 1.1 shows a graphical representation of the Universe history. It is possible to notice how the beginning of the Reionization Era coincides with appearance of the first Stars and Galaxies. Indeed, according to the current acknowledged scenario, the UV radiation of first objects started to dig ionized hydrogen (HII) bubbles around them in a “sea” of neutral hydrogen

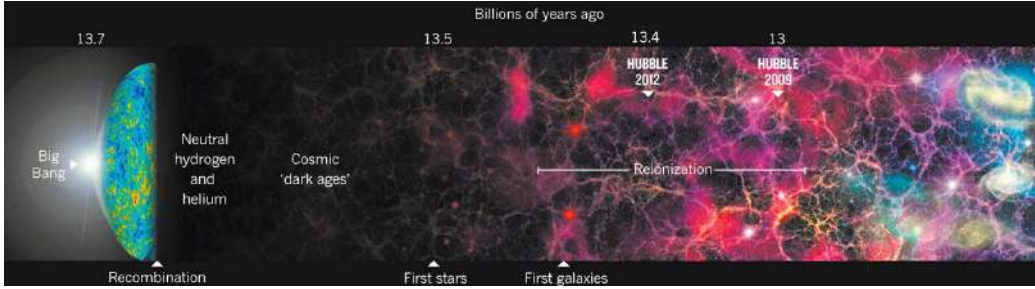


Figure 1.1: Artistic representation of Universe history. The figure illustrates deepness reached by Hubble Deep fields and the limits of Reionization era. *Illustration: Nik Spencer; Source: NASA/WMAP Science Team; R. Ellis (Caltech)*

(HI). In a first long phase sources drive the expansion of isolated Strömgren spheres, that eventually overlap and accelerate to complete the ionization of the Universe at $z < 6$.

With the depletion of the last filaments of neutral hydrogen the Universe appears now fully ionized (the neutral fraction is $x_{HI} \equiv \frac{n_{HI}}{n_{HI} + n_{HII}} \sim 10^{-5}$ at $z = 0$).

The main questions about Reionization are when and how, after the Dark Ages, did the Universe get ionized? What was the typical radiation field and, how homogeneous is it and what was producing it? In the next section we will answer the first question, the only one where there is a clear convergence.

1.1.1 When was the Universe re-ionized?

A first question is to identify when the Reionization happened, when it started and how it evolves. Recent measures on CMB and high redshift quasars observations set the reionization period between $6 < z < 8.5$.

The end of Reionization at $z \sim 6$

[Gunn and Peterson 1965](#) first proposed the use of Ly α resonance absorption in the spectrum of distant quasars as a direct probe to the neutral hydrogen density in the IGM at high-redshift. For objects beyond Reionization, neutral hydrogen in the IGM creates complete Gunn-Peterson (GP) absorption troughs in the quasar spectrum blueward of Ly α emission. Observations of the GP effect directly constrain the evolution of neutral hydrogen fraction and the ionization state of the IGM.

The *Gunn-Peterson optical depth* to Ly α photons is:

$$\tau_{GP} = \frac{\pi e^2}{m_e c} f_\alpha \lambda_\alpha H^{-1}(z) n_{HI} \quad (1.1)$$

where f_α is the oscillator strength of the Ly α transition, $\lambda_\alpha = 1216 \text{ \AA}$, $H(z)$ is the Hubble constant at redshift z and n_{HI} is the density of neutral hydrogen in the IGM.

For a uniform IGM at high redshift, adopting the present-day baryon density Ω_b and $H(z) \propto h\Omega_m^{1/2}(1+z)^{3/2}$, the GP optical depth can be expressed as:

$$\tau_{GP}(z) = 4.9 \times 10^5 \left(\frac{\Omega_m h^2}{0.13} \right)^{-1/2} \left(\frac{\Omega_b h^2}{0.02} \right) \left(\frac{1+z}{7} \right)^{3/2} \left\langle \frac{n_{HI}}{n_H} \right\rangle \quad (1.2)$$

Even a tiny neutral fraction $x_{HI} \sim 10^{-4}$ gives rise to a complete Gunn-Peterson absorption. This test is only sensitive at the end of reionization, when IGM is already mostly ionized, but it saturates for higher neutral fraction in an earlier stage.

The neutral hydrogen fraction, and therefore the GP optical depth, depends on the local density of the IGM. By studying the evolution of the average transmitted flux, or effective optical depth, it is possible to trace the evolution of the UV ionizing background and neutral fraction of the IGM, taking into account that at high redshift the IGM is highly clumpy.

The *transmitted flux ratio* can be defined as the average ratio of the observed flux to that in absence of absorption:

$$\mathcal{T}(z_{abs}) \equiv \left\langle \frac{f_\nu^{obs}}{f_\nu^{int}} \right\rangle \quad (1.3)$$

and therefore the *effective Gunn-Peterson optical depth* can be defined as:

$$\tau_{GP}^{eff} \equiv -\ln(\mathcal{T}) \quad (1.4)$$

Using a sample of 19 high redshift, $z_{em} > 5.7$, QSOs [Fan et al. 2006](#) measured the mean opacity of the IGM to Ly α in discrete regions along the line of sight towards individual objects.

Looking at the spectra they noticed a strong redshift evolution of the transmission of the IGM: transmitted flux is clearly detected in spectra of quasars $z < 6$ and blueward of the Ly α emission line; the absorption troughs deepen for high redshift QSOs and complete GP absorption begins to appear along lines of sight at $z > 6.1$.

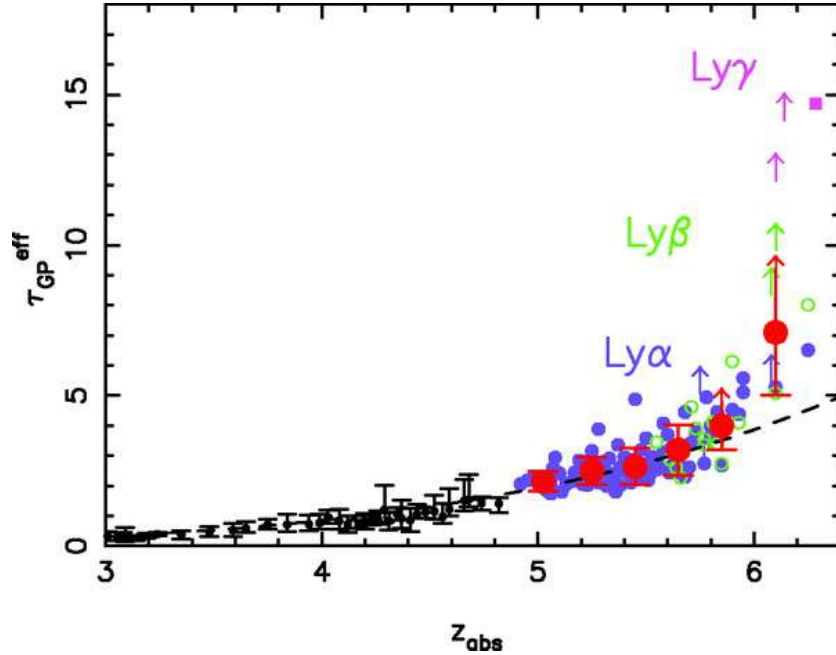


Figure 1.2: From [Fan et al. 2006](#) Figure 5. Evolution of the optical depth with combined Ly α , Ly β and Ly γ results.

Figure 1.2 shows the evolution of the optical depth with redshift averaged over intervals $\Delta z = 0.15$ along each line of sight. In complete GP, troughs where no flux is detected, the limit is indicated with an arrow.

The dash line represents the best fit power law for $z_{abs} < 5.5$:

$$\tau_{GP} = (0.85 \pm 0.06) \left(\frac{1+z}{5} \right)^{4.3 \pm 0.3} \quad (1.5)$$

while there appears to be an accelerated evolution at $z \gtrsim 5.7$ where the best fit evolution has a behaviour $\tau_{GP} \propto (1+z)^{>11}$.

There is also a rapid increase in the sample *variance* with redshift along different lines of sight for $z > 5.7$: this is due to the fact that some lines of sight have complete GP troughs, while others have optical depths only mildly higher, compared to the power-law extrapolation from low redshift. This measurements dispersion supports the interpretation that the IGM is close to the end of a clumpy reionization.

Fan et al. studied also the evolution of HII regions around the quasars, noticing that the sizes decrease significantly with redshift, while the distribution of dark gaps in QSOs spectra show a similar trend to the GP optical depth.

All those measures can be used to derive the ionization state of the IGM

using parameters as the mean UV ionizing background, the mean free path of UV photons and the neutral hydrogen fraction x_{HI} .

From $z \sim 5$ to $z > 6$ there is an order of magnitude increase for the neutral fraction with the evolution accelerated at $z > 5.7$. At the same time the dark gaps evolution sets an upper limit on the IGM neutral fraction $x_{HI} < 50\%$ at $z \sim 6.4$ (McGreer, Mesinger, and D’Odorico 2015).

A contribution to this topic was done also by Becker et al. 2015 who found evidence of patchy hydrogen reionization.

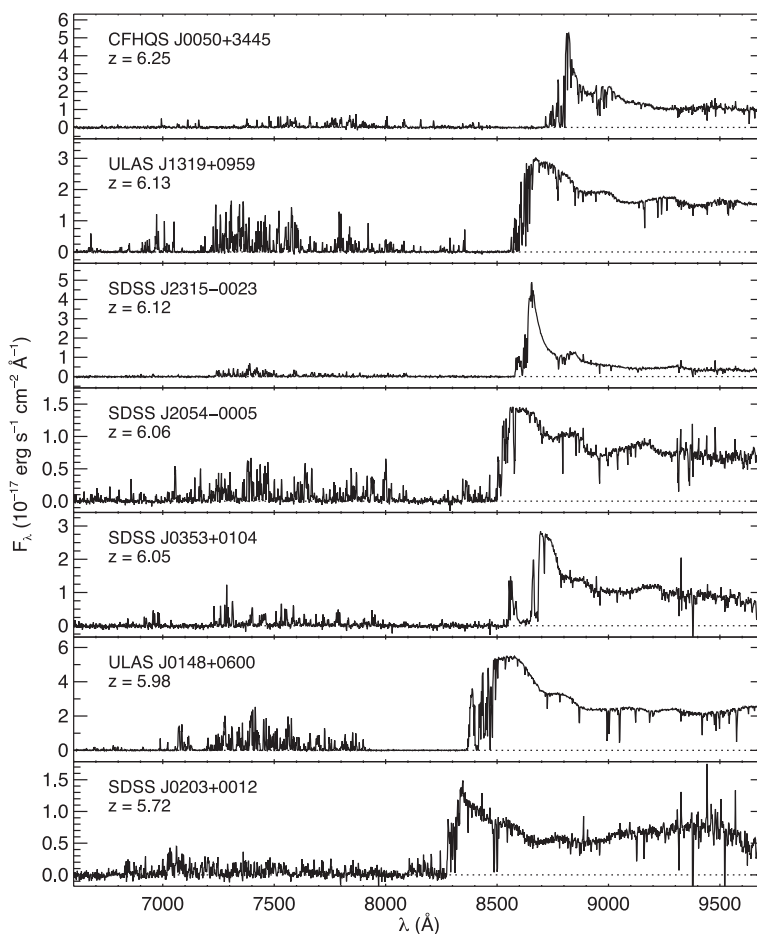


Figure 1.3: From Becker et al. 2015 Figure 1. Spectra of $z \sim 6$ quasars analysed in the work.

They report 7 new high resolution spectra of QSOs at $z > 5.7$, in Figure 1.3, to integrate the 19 QSOs of Fan et al. 2006 and other 16 lower redshift quasars from the literature spanning the redshift range $4.5 < z_{em} < 5.4$. With

this larger sample they were able to evaluate the evolution of the Ly α opacity, including its scatter between lines of sight, in a self-consistent manner over the entire redshift range $3.8 < z < 6.3$.

This work was triggered by the spectra of ULAS J0148-0600, second to last in Figure 1.3. Its Ly α forest displays an unusually long ($\sim 110\text{Mpc } h^{-1}$) Ly α trough over $5.523 < z < 5.879$, which is in contrast with other quasar spectra showing abundant transmitted flux towards other lines of sight at the same redshift.

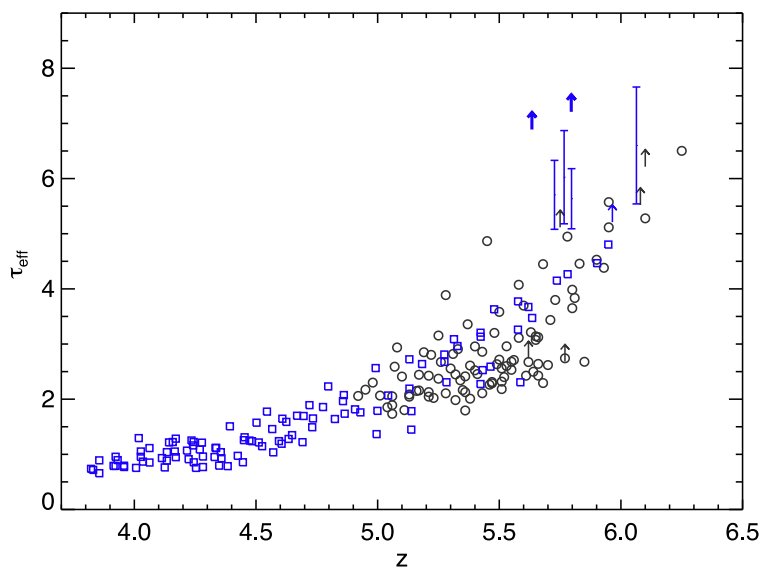


Figure 1.4: From [Becker et al. 2015](#) Figure 6. Ly α τ_{eff} measurements. Black circles and arrows (lower limits) are from [Fan et al. 2006](#). Blue squares, error bars, and arrows are from [Becker et al. 2015](#). The two blue arrows at $z \simeq 5.63$ and 5.80 with $\tau_{eff} \gtrsim 7$ are from the line of sight towards ULAS J0148+0600.

Authors provided an updated description of the end of Reionization Era in terms of Thomson optical depth τ_{eff} . Figure 1.4 shows all available measures of τ_{eff} as function of redshift, as reported by [Becker et al. 2015](#). The two blue arrows are from the line of sight towards ULAS J0148+0600. Here it is evident an overall increase of τ_{eff} with redshift and an enhanced scatter at $z > 5$.

Becker et al. compared those data with simulations of simple models of the UV background to understand if they are able to reproduce sightline-to-sightline variations. They found that, at lower redshift $3.9 < z < 4.9$, models can track data while, at higher redshift $5.7 < z < 5.9$, only a small percentage of data can be reproduced with models. They argued that spatial

variations in the mean free path must be present at these redshift. This is in line with a patchy reionization picture.

Some other recent works confirmed this hypothesis as [Bosman et al. 2018](#), [Eilers, Davies, and Hennawi 2018](#), [Kulkarni et al. 2019](#), [Eilers et al. 2019](#) and [Meiksin 2020](#).

The evolution of Reionization

Due to the high opacity at $z > 6$ it is not possible to determine when Reionization started using high redshift quasars.

It is instead possible to study a more distant probe, the Cosmic Microwave Background (CMB): this radiation is indeed affected by the total column density of free electrons along each line of sight, parametrized by its Thomson scattering optical depth τ . Reionization leaves then imprints in the CMB power spectra, both in polarization at very large scales and in intensity via the suppression of temperature angular power spectrum TT power at higher multipole moment ℓ (angular scale).

In addition to the effect on the polarized large-scale CMB anisotropies, reionization generates CMB temperature anisotropies through the *kinetic Sunyaev-Zeldovich (kSZ)* effect ([Sunyaev and Zeldovich 1980](#)), caused by the Doppler shift of photons scattering off electrons moving with bulk velocities. It is composed by two component: the “homogeneous” component when the reionization was complete and the “patchy” one, which arises during the process of reionization, from the proper motion of ionized bubbles around emitting sources.

In a recent release of “Planck results papers”, (i.e. [Planck Collaboration et al. 2020a](#); [Planck Collaboration et al. 2020b](#)) a study on constraints from CMB data on reionization history was reported. They combined the CMB anisotropy data in temperature with the low-multipole polarization data to fit the different Λ CDM models.

At smaller scales than the horizon size at reionization ($\ell \gtrsim 10$), free electrons generated during the reionization can scatter and partially damp the CMB anisotropies. This leads to a most scaled-independent suppression of the above $\ell \approx 10$ by a factor of $e^{2\tau}$. Therefore the amplitude of the bump scales like τ^2 .

The results from different model for reionization history are shown in Figure 1.5. The `FlexKnot` method result represents the best model-independent estimate of the optical depth, resulting $\tau = 0.054 \pm 0.007$. This method reconstructs any arbitrary history using an interpolation function between a varying number of knots ([Millea and Bouchet 2018](#)). Here the model param-

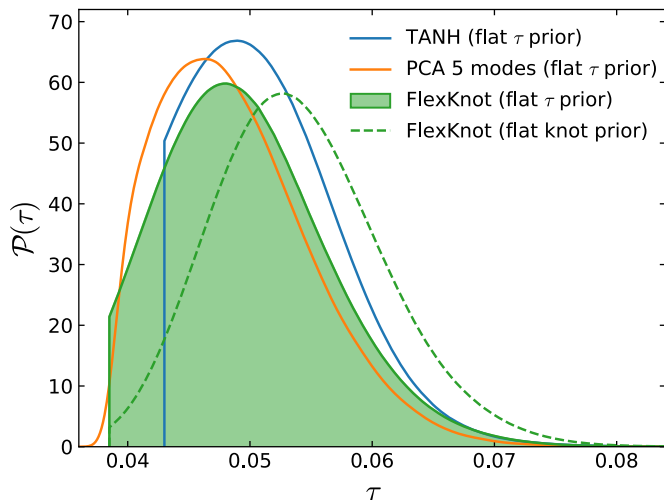


Figure 1.5: From [Planck Collaboration et al. 2020b](#) Figure 44 (Top): marginalized constraints on the optical depth to reionization from lowE alone, assuming different models of reionization and different priors over the model parameters. Only reionization parameters are varied here. The solid lines use a flat prior on τ , while the dashed line uses a flat prior on the knot amplitudes.

eters are directly tied to the physical quantity of the ionization fraction.

The total integrated optical depth to reionization τ is related to the free electron fraction $x_e(z) \equiv n_e^{reion}(z)/n_H(z)$ by

$$\tau = n_H(0)\sigma_T c \int_0^{z_{max}} \frac{(1+z)^2 dz}{H(z)} x_e(z) \quad (1.6)$$

where $n_e^{reion}(z)$ is the number density of free electron from reionization, σ_T is the Thomson scattering cross-section and $n_H(0)$ is the number density of the hydrogen nuclei today.

Figure 1.6 shows the free electron fraction $x_e(z)$ constrained from Planck along with astrophysical constraints. The shaded bands are middle 68th and 95th percentiles from **FlexKnots** method.

The estimate of $\tau = 0.054 \pm 0.007$ suggests a mid-point reionization redshift $z_{re} = 7.7 \pm 0.7$. This is consistent with astrophysical observations of quasar lines and models in which reionization happened relatively fast and late.

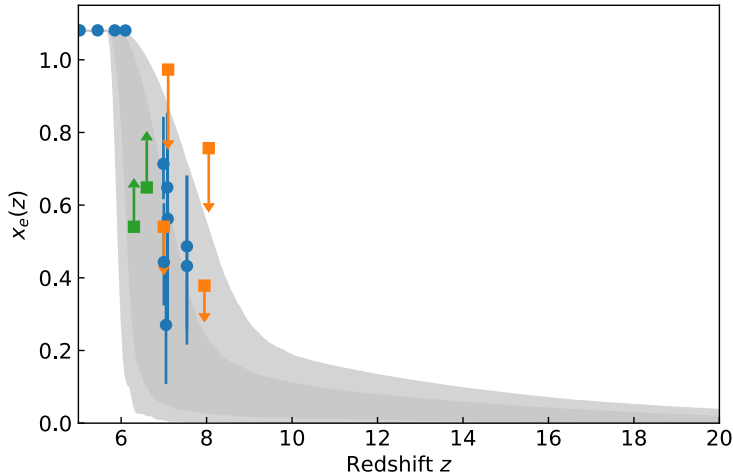


Figure 1.6: From [Planck Collaboration et al. 2020a](#) Figure 36 : Free electron fraction, $x_e(z)$, constrained from Planck using the `FlexKnots` method of [Planck Collaboration et al. 2020b](#) and [Millea and Bouchet 2018](#), along with astrophysical constraints, as tabulated in [Bouwens et al. 2015](#), updated to include [Greig et al. 2017](#), [Bañados et al. 2018](#), [Mason et al. 2018](#), and [Davies et al. 2018](#). The shaded bands are middle 68th and 95th percentiles

1.1.2 A global view

The ones described in previous sections are only some of the main observational techniques to constrain cosmic reionization and the evolution of x_{HI} . Some other estimation methods provide more independent probes, that were collected together, as described below.

A global view is reported in Figure 1.7, from the [Ota et al. 2017](#) paper, where all known measures of neutral fraction x_{HI} are collected together.

Dark damping wings observed on some $z > 5.9$ γ -ray burst (GRB) spectra are associated with the presence of neutral IGM. It was then possible to set some constraints on x_{HI} evolution with redshift ($x_{HI} = 0.12 \pm 0.05$ at $z = 6.33$ ([Chornock et al. 2014](#)) and to determine that reionization might not yet be complete at $z > 6$.

Another probe of reionization are Ly α emitters (LAEs). Their Ly α luminosity function (LF) could decline as neutral hydrogen absorbs or scatters Ly α photons from LAEs. From the observed lack of UV LF evolution of LAEs at $z = 5.7 - 6.6$ it is possible to estimate a neutral fraction of $x_{HI} = 0.4$ at $z = 6.6$.

The strength of the Lyman alpha absorption from the IGM depends on the spatial distribution of galaxies and cosmic neutral hydrogen patches

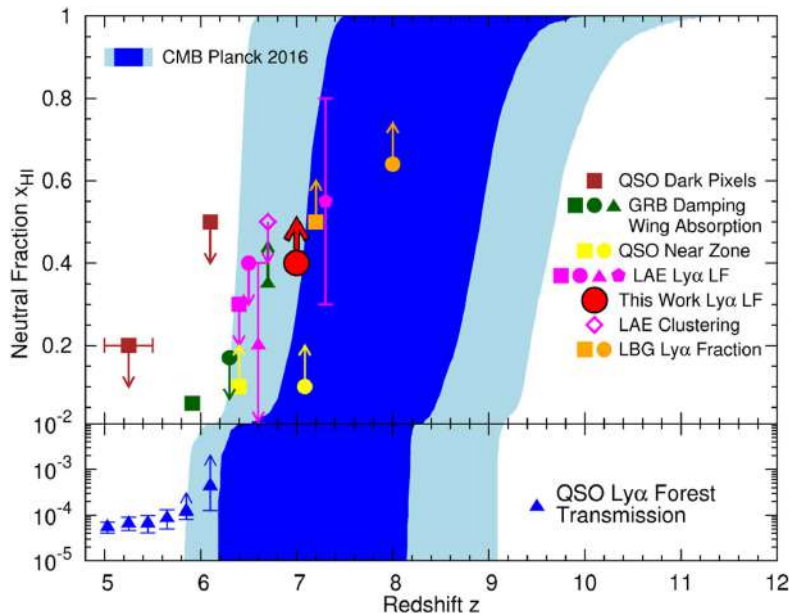


Figure 1.7: From [Ota et al. 2017](#) Figure 20 (Left). Cosmic reionization history (neutral fraction x_{HI} as a function of redshift) constrained by their $z = 7$ LAE survey and previous studies using various probes. The blue and light blue shaded regions show the 68% and 95% allowed intervals of reionization history constrained by the analysis of the pre-2016 Planck CMB observations data by [Planck Collaboration et al. 2016](#).

([Furlanetto, Zaldarriaga, and Hernquist 2006](#); [McQuinn et al. 2007](#)). Strong clustering of LAEs can be due to large patches of neutral hydrogen (HI).

Some Lyman break galaxies (LBGs) are known to exhibit strong Ly α emission ($EW_0 > 25 \text{ \AA}$) while others do not. The fraction of Ly α -emitting LBGs increases from $z \sim 3$ to 6, but suddenly decreases from $z \sim 6$ to 7–8. This could be due to attenuation of Ly α emission by neutral hydrogen and the rapid evolution of the neutral fraction from $z \sim 6$ to 7.

Model independent constraints to x_{HI} can be provided from studies on QSO dark Ly α and Ly β forest pixels. [McGreer, Mesinger, and D’Odorico 2015](#), [Becker et al. 2015](#) and [Bosman et al. 2018](#) found indication that Reionization was almost complete at $z \sim 5.5 - 6.0$.

The size of near-zone QSO, measured by the transmission blueward of Ly α due to the strong QSO radiation, give an estimate of the neutral hydrogen $x_{HI} > 0.1$ at $z = 7.1$ ([Mortlock et al. 2011](#); [Bolton et al. 2011](#)).

Combining together all these probes we obtain a scenario, as the one

reported in Figure 1.7, of a late and fast Reionization.

1.1.3 An analytic picture

The definition of boundaries of the Reionization process is the first step to unveil more details about this phase. From the current scenario a detailed model is very challenging and requires numerical simulations to take into account all relevant physical process as self-gravity, star formation, feedback, radiative transfer and dark matter. In addition, it is not yet clear which sources contributed to reionization and their interaction with the environment.

A zeroth order description requires to properly consider the production and absorption of ionizing LyC radiation as a function of epoch in a clumpy, expanding medium. In his paper [Madau 2017](#) proposes an updated version of his previous *reionization equation* taking into account the presence of optically thick Lyman-limit systems (LLSs) that are known to determine the mean free path of ionizing radiation after overlap.

At $z < 6$ the diffuse IGM is observed to be highly ionized, with only a small residual amount of neutral hydrogen set by the balance between radiative recombinations and photoionizations from a nearly uniform UV radiation background, and provides negligibly small HI photoelectric absorption.

The *continuum opacity* is instead dominated by the optically thick (to ionizing radiation) Lyman-limit systems, high density regions that trace non-linear and collapsed structures, occupy a small portion of the volume, and are able to keep a significant fraction of their hydrogen in neutral form.

To briefly summarize, the model for the total absorption probability per unit length $\langle \kappa_\nu \rangle$ is composed by two factors:

$$\langle \kappa_\nu \rangle = \langle \kappa_\nu^{IGM} \rangle + \langle \kappa_\nu^{LLS} \rangle \quad (1.7)$$

where $\langle \kappa_\nu^{IGM} \rangle^{-1}$ is the photon proper mean free path at the hydrogen Lyman edge caused by the LLSs. In the limit of optically thick, Poisson-distributed clouds it can be represented by the average spacing between LLSs (“picket-fence” absorption).

The $\langle \kappa_\nu^{LLS} \rangle$ represents, indeed, the absorption coefficient for photons in the IGM and it can be expressed in terms of the neutral fraction $\langle x_{HI} \rangle = (1 - Q)$ with Q as volume-averaged hydrogen ionized fraction.

The updated version of the *reionization equation* is:

$$\frac{dQ}{dt} = \frac{\langle \dot{n}_{ion} \rangle}{\langle n_H \rangle (1 + \langle \kappa_{\nu L}^{LLS} \rangle / \langle \kappa_{\nu L}^{IGM} \rangle)} - \frac{Q}{t_{rec}} \quad (1.8)$$

where $\langle \dot{n}_{ion} \rangle$ is the emission rate into the IGM of ionizing photons per unit proper volume and t_{rec} is the effective recombination timescale of a clumpy IGM.

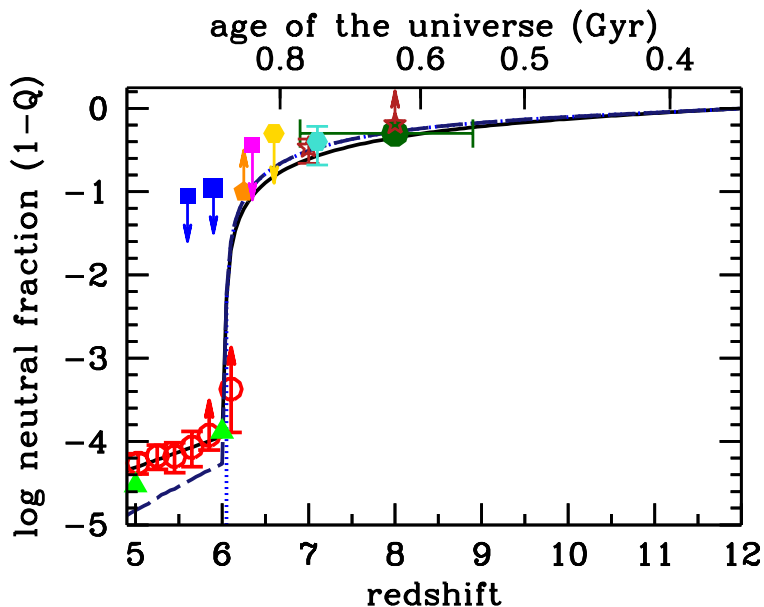


Figure 1.8: From [Madau 2017](#) Figure 1 (Left). Curves represent the average neutral fraction $(1 - Q)$ of an IGM that recombines according two different models of t_{rec} and the contribution of LLSs. The data points represents constraints on the ionization state of the IGM from a set of independent measures: i.e. Gunn-Peterson optical depth, Ly α forest opacity, dark pixel statistics and all the other available at paper time.

Figure 1.8 shows the reionization history as predicted by integrating equation (1.8) with a constant emission rate of ionizing photons per hydrogen atom $\langle \dot{n}_{ion} \rangle / \langle n_H \rangle = 2.9 \text{ Gyr}^{-1}$. In the same figure, all available measures from both GP high redshift QSO and Planck are shown.

The result obtained from equation (1.8) allows to identify two main phases: pre and post overlap, when the opacity changes source.

Prior to overlap, when $Q < 1$ and $\langle \kappa_{\nu_L}^{LLS} \rangle \ll \langle \kappa_{\nu_L}^{IGM} \rangle$ the volume-averaged photoionization rate becomes independent of the mean free path and equal to the ionizing photon emission rate into the IGM per hydrogen atom.

After the overlap $Q \rightarrow 1$ and $\langle \kappa_{\nu_L}^{LLS} \rangle \gg \langle \kappa_{\nu_L}^{IGM} \rangle$ the volume-averaged photoionization rate is reduced by the finite mean free path of LLSs and becomes smaller as the neutral fraction of the IGM decreases. The integration of equation (1.8) therefore provides a link between the pre-overlap and post-overlap phases of the reionization process.

Photoionizations dominate over recombinations until just before overlap. As overlap is approached, the neutral fraction of the IGM dives below 10^{-4} and the photoionization source term drops rapidly.

This work can be used to test different models and parameters: the comoving ionizing photon emissivity into the IGM $\langle \dot{n}_{ion} \rangle$ can be expressed as a function of the comoving luminosity density ρ_{UV} and f_{esc} , the *escape fraction*, i.e. the globally-averaged relative fraction of *LyC* photons that escape from individual galaxies and make it into IGM.

1.2 Who reionized the Universe?

While boundaries on Reionization seem to be determined by stringent limits from Gunn-Peterson effect on high z QSOs and from Planck CMB observations, there is still a large debate about who was able to ionize the whole Universe in few hundreds of Myrs.

It is generally thought that the IGM is kept ionized by the integrated UV emission from Active Galactic Nucleus (AGN) and Star Forming Galaxies (SFG), but the relative contribution of these sources as a function of epoch is poorly known.

The main ingredients to determine which kind of sources was able to ionize the Universe are the objects density and their capability to ionize hydrogen, that can be tracked respectively by the *Luminosity Function* (LF) and the *Lyman continuum escape fraction* f_{esc} .

The LF describes the distribution per luminosity interval and redshift; f_{esc} is the ratio between the photons produced at $\lambda \leq 912 \text{ \AA}$ rest-frame and those that are able to reach the intergalactic medium i.e. that are not absorbed by neutral hydrogen or by the dust of the IGM.

Figure 1.9 shows the comoving 1 Ryd emissivity (13.6 eV) according the “*minimal reionization model*” proposed by [Haardt and Madau 2012](#). For $z > 3$ the reducing population of bright AGNs ($M_{1450} \leq -27$) appears to

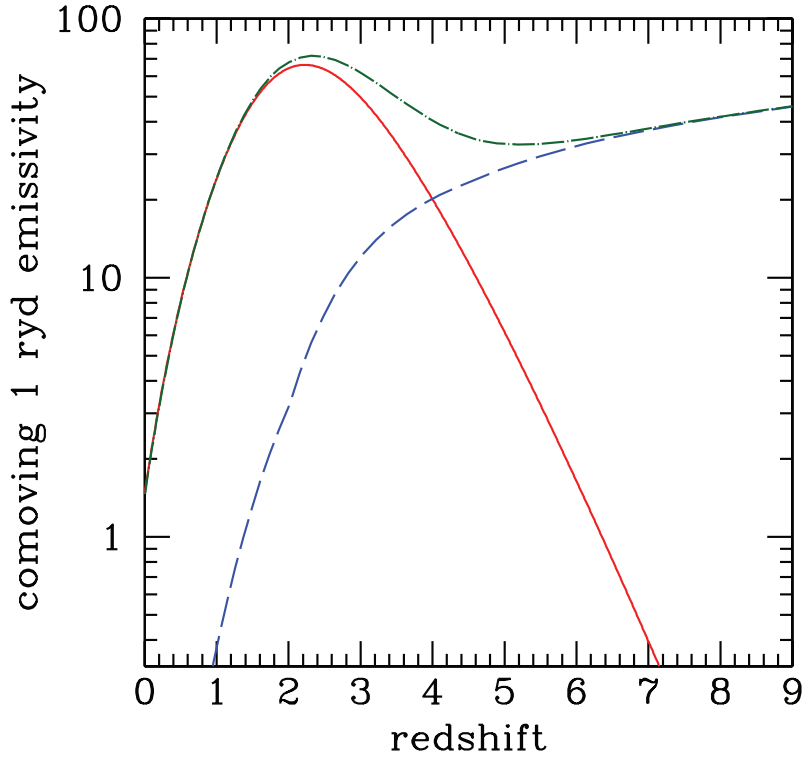


Figure 1.9: From [Haardt and Madau 2012](#) Figure 7 (Right): comoving galaxy emissivity (in units of $10^{23} \text{ erg s}^{-1} \text{ Mpc}^{-3} \text{ Hz}^{-1}$) of 1 ryd photons escaping into the IGM (blue dashed line). The solid red line shows the best-fit quasar emissivity for comparison, while the dot-dashed line shows the total quasars + galaxies emissivity.

make a small contribution to the radiation background and it is believed that massive stars in galactic and sub-galactic systems provide the additional ionizing flux needed at early time.

The relative contribution of SFGs and AGNs to UV ionizing background is still under discussion. Therefore in the next section I will introduce, for both SFGs and AGNs, their *Luminosity Function* (LF) evolution with redshift, i.e. how many sources are there and at which luminosity, and the *escape fraction* f_{esc} to determine “the ionizing power” of each class of sources.

1.2.1 Star Forming Galaxies

The high redshift SFG are the main candidates to be the ionizing sources because of the lack of a numerous population of $z > 3$ AGNs. Luminosity functions of different works give us a scenario of a rich population of star forming galaxies at $z > 4$ (Finkelstein et al. 2015; Bouwens et al. 2015; Finkelstein et al. 2019).

A lot of discrepancies are instead found in the estimated value of f_{esc} for this kind of sources, especially for high redshift, where we expect an increase in the efficiency to engender ionizing bubble in their neighborhood.

As an example, in their model Haardt and Madau 2012 had to assume a rapid increase of SFGs f_{esc} with redshift, because of the small escape fraction measured at $z = 3.3$ by Boutsia et al. 2011

$$\langle f_{esc} \rangle = 1.8 \times 10^{-4} (1 + z)^{3.4} \quad (1.9)$$

to ensure the UV background needed to ionize hydrogen.

Some measurements on local galaxies indicate an escape fraction within 6 – 13%, but for higher redshift $3 < z < 4.5$, before the IGM absorption prevents this kind of measures, we still lack reliable estimates.

Grazian et al., restricting the research to $z \sim 3.3$, estimated the *relative escape fraction* f_{esc}^{rel} from the ratio of fluxes in the U and R bands (Grazian et al. 2016 and Grazian et al. 2017).

Indeed at that redshift the U and R bands correspond to rest-frame wavelengths of 900 and 1500 Å respectively. Therefore, it is possible to estimate the relative escape fraction defined as the fraction of emitted Lyman continuum photons, IGM corrected, relative to the observed fraction of photons at 1500 Å rest frame.

$$f_{esc}^{rel} = \frac{(L_{1500}/L_{900})_{int}}{(F_R/F_U)_{obs}} \exp(\tau_{900}^{IGM}) \quad (1.10)$$

where $(L_{1500}/L_{900})_{int}$ is the ratio of the intrinsic luminosity at 1500 Å and 900 Å rest frame, $(F_R/F_U)_{obs}$ is the ratio of the observed flux in the R and U bands and τ_{900}^{IGM} is the optical depth of the IGM at 900 Å rest frame.

In the paper Guaita et al. 2016 was explored the dependency of intrinsic luminosity ratio $(L_{1500}/L_{900})_{int}$ from different population synthesis models. On the base of their results Grazian et al. decided to adopt a $(L_{1500}/L_{900})_{int}$ value of 3.

To estimate the IGM coefficient, authors used U -band filter of LBC imager at the Large Binocular Telescope (LBT) and adopted the procedure

outlined in [Boutsia et al. 2011](#). They obtain a mean IGM transmission $\langle \exp(\tau_{900}^{IGM}) \rangle = 0.28$ for $z \sim 3.3$.

For each galaxy they could, finally, compute the observed flux $(F_R/F_U)_{obs}$, using the R -band detection and measured the flux, or at least an upper limit, in U -band images, assuming a galaxy profile¹ and correcting for spatial resolution differences between bands ([Grazian et al. 2016](#)).

From equation (1.10) they could, then, estimate the f_{esc}^{rel} for a set of galaxies at $z \sim 3.3$.

Authors proceed with the stacking of LBC U and R bands of the clean sample in different luminosity intervals. They stacked source cut-outs masking all nearby sources detected in the R band and leaving the central region unmasked. Using the same procedure previously described, the authors were able to measure f_{esc}^{rel} in different luminosity intervals.

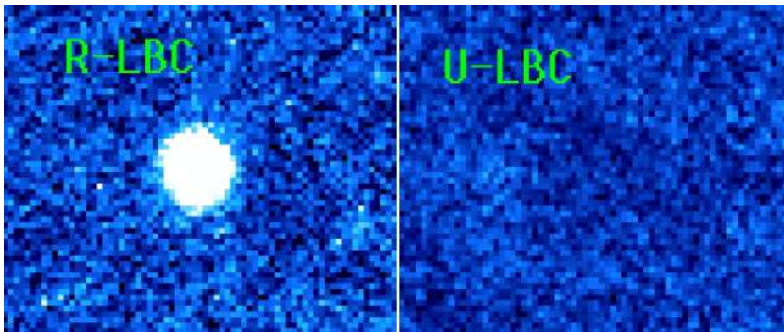


Figure 1.10: From [Grazian et al. 2017](#) Figure 4: Stack of the 69 galaxies with $3.27 \leq z_{spec} \leq 3.40$ in the R (left) and U bands (right) for the COSMOS, EGS, and CANDELS/GOODS- North fields. The size of each image is 16 arcsec by 14 arcsec.

In Figure 1.10 the stacking of all 69 galaxies is shown in U and R bands resulted in a non detection in the U -band to a level of $U = 31.74$ at 1σ and a clear detection in the R -band $R = 24.73$. This measure corresponds to an upper limit to the relative escape fraction of $f_{esc}^{rel} \leq 1.7\%$ at 1σ level at $z = 3.3$, the most stringent for SFGs at $z > 3$.

Figure 1.11 shows measurements and upper limits obtained from single galaxy observations (indicated with asterisks and small arrows) and by stacking the galaxies of the sample in different luminosity intervals. The gray,

¹For the U -band authors assumed the same profile of the galaxy in the R -band through the ConvPhot software ([De Santis et al. 2007](#))

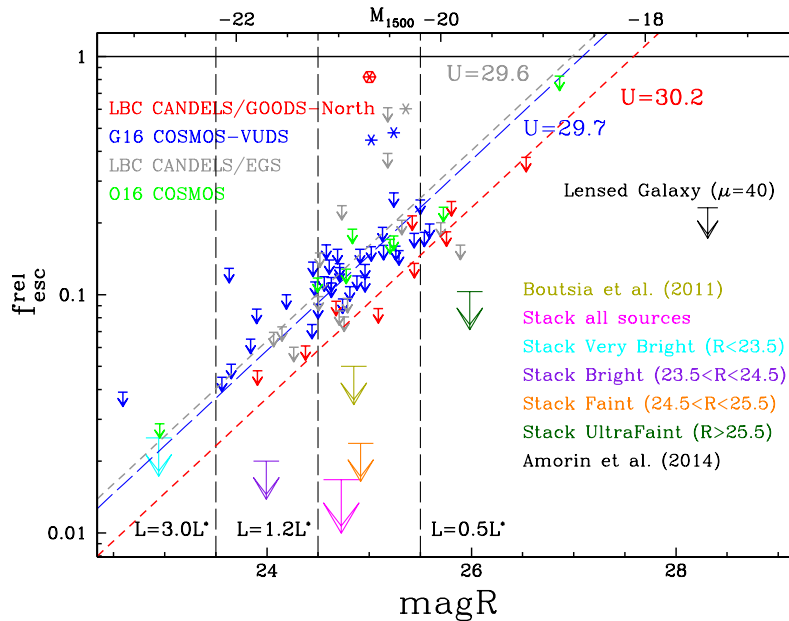


Figure 1.11: From [Grazian et al. 2017](#) Figure 3: Measured values and the upper limits at 1σ for the LyC relative escape fraction of galaxies at $z \sim 3.3$ in the COSMOS (blue, green), EGS (gray), and CANDELS/GOODS-North fields (red). Small descending arrows show the upper limits of f_{esc}^{rel} associated with individual galaxies. The large arrows (cyan, purple, magenta, orange, gold, dark green) represent the limits to f_{esc}^{rel} derived by stacking the $z \sim 3.3$ galaxies in different luminosity intervals.

blue and red dashed lines show the escape fraction versus R -band magnitude corresponding to the 1σ depth of the U -band in different fields using the equation (1.10).

Dividing the sample in different luminosity intervals, the resulting limits are $f_{esc}^{rel} \leq 2.5\%$, 2.0% , 2.4% and 10.3% at $M_{1500} = -22.8$, $M_{1500} = -21.7$, $M_{1500} = -20.8$ and $M_{1500} = -19.7$ respectively, represented by large arrows on Figure 1.11. The apparent increasing trend for fainter sources is related to a limit in depth for U band more than a real trend.

Finally, the magenta arrow, the largest one, is the value of $f_{esc}^{rel} \leq 1.7\%$ found from the stacking of all images.

[Grazian et al. 2017](#) adopted an alternative approach, based on relative escape fraction to derive the LyC emissivity (that is at 900 \AA rest-frame) ρ_{900}^{esc} . This quantity can be translated in terms of HI *photoionizing background* Γ_{-12} in units of 10^{-12} photons per second.

Authors evaluated Γ_{-12} values for different parametrization of f_{esc}^{rel} and luminosity functions and compared them with the expected contribution by galaxies required in order to explain the measured UV background (UVB). They can conclude that $z \sim 3$ galaxies brighter than $M_{1500} = -17.4$ have difficulties providing the UVB measured in the Lyman forest at $z \sim 3$, given that their LyC escape fraction is below 10%.

Independent estimates of the escape fraction in GRB host galaxies at $1.6 < z < 6.7$ by [Tanvir et al. 2019](#) set a stringent upper limit of 1.5% for f_{esc} of faint star forming galaxies, indicating that for the hypothesis that such galaxies dominated the reionization of the Universe could be problematic.

1.2.2 AGNs

The other candidates for ionizing the Universe are AGNs. It is well known that bright QSOs ($M_{1450} \leq -26$ at $z \geq 3$) are efficient producers of HI ionizing photons, and they can ionize large bubbles of HI even at distances up to several Mpc out to $z \sim 6$. However, their space density is too low at high- z to provide the cosmic photoionization rate required to keep the IGM ionized at $z > 3$. The bulk of ionizing photons could, then, come from a population of fainter AGNs, but the luminosity function at high redshift is not sufficiently well known to test this hypothesis.

It is possible to find high redshift AGNs first selecting galaxies at $z > 4$ in deep NIR H -band catalogues, that correspond to UV-rest frame galaxy luminosity. Then, from this sample, AGN sources can be identified on the detection of significant X-ray emission in the source positions. This method was recently proposed by [Fiore et al. 2012](#).

Applying this procedure to some well observed fields, [Giallongo et al. 2019](#) could extract 32 faint AGNs from a population of near-IR selected galaxies at photometric $z > 4$ showing X-Ray detection in deep *Chandra* images available. The authors were able to explore the UV luminosity function in the redshift interval $4 < z < 6.1$ including results from an analogous work on COSMOS field, [Boutsia et al. 2018](#), and other bright QSO samples.

They studied space densities and emissivities of AGNs at $z > 4$ to evaluate their contribution to the cosmological ionizing emissivity.

The *galaxy volume density* $\phi(M, z)$ in a given $(\Delta z, \Delta M)$ bin can be com-

puted as

$$\phi(M, z) = \frac{1}{\Delta M} \sum_{i=1}^n \left[\sum_j \omega(j) \int_{z_{low}}^{z_{lim}^{(i,j)}} \frac{dV}{dz} dz \right]^{-1} \quad (1.11)$$

where $\omega(j)$ is the area in units of steradians corresponding to the region j , n is the number of objects in the bin and dV/dz is the comoving volume element per steradian.

Rest frame UV 1450Å absolute magnitudes M_{1450} were computed for each redshift from apparent magnitudes in the filters closer to this rest frame wavelength.

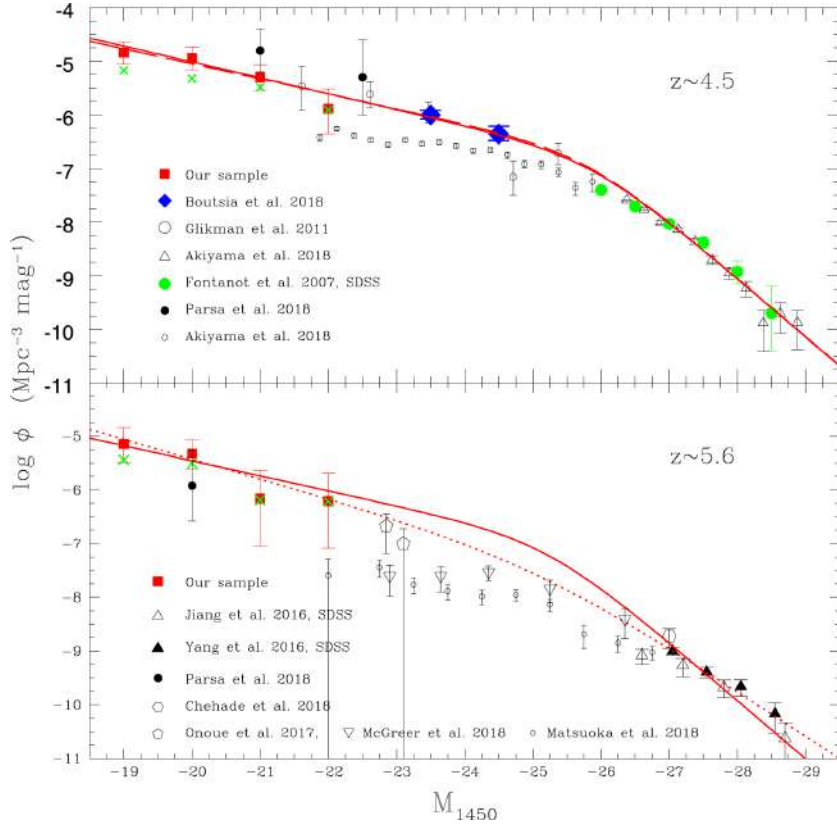


Figure 1.12: From [Giallongo et al. 2019](#) Figure 4: UV 1450 Å AGN luminosity functions in two redshift intervals. Different symbols represent $1/V_{max}$ data points from different surveys as explained in the figure box. Dashed and continuous curves are LF best fits obtained using different models and sub-samples.

Figure 1.12 shows the AGN volume densities as a function of M_{1450} , z in two redshift intervals $\Delta z = 4-5$ and $\Delta z = 5-6.1$. Lines represent LF best-fit

derived for different models. Only CANDELS and SDSS with $M_{1470} < -27$ samples have been used for the analysis. Other volume densities derived by color-selected surveys at intermediate magnitudes are shown for comparison.

The luminosity function at UV 1450 Å rest frame were modelled with a double power-law with a break magnitude M_{break}

$$\phi = \frac{\phi^*}{10^{0.4(M_{break}-M)(\beta-1)} + 10^{0.4(M_{break}-M)(\gamma-1)}} \quad (1.12)$$

For the first redshift interval, $z \sim 4.5$, authors determine values for break magnitude M_{break} and slopes, β and γ . For the bin $z \sim 5.6$, given the shortage of data, they could not derive a global shape for the LF with a good accuracy, obtaining different results for several sub-samples.

The global LFs have been used to predict ionizing emissivities and photoionization rates from the global AGN population at $4 \lesssim z \lesssim 6$. It was assumed a full escape fraction ($f_{esc} = 1.0$) due to the AGN activity, supported by measures of high f_{esc} of ionizing photons in AGNs of intermediate luminosity (Grazian et al. 2018).

From the obtained results, they suggested a significant contribution to the ionizing UV background by the AGN population in the scenario of a late Reionization, as proposed by Planck data. The AGN activity in star-forming galaxies could add the required boost of ionizing photons escaping into the IGM and could play a significant role in the ionizing history of the IGM.

1.2.3 How to find high redshift AGNs

The research of a faint high z AGN population is a puzzling problem.

In general AGN can be identified in several ways, the most obvious being through **spectroscopy**. AGNs can either be identified through extremely broad lines or by determining line ratios. The former has the problem that it only selects the small subset of broad-lined Type I and misses Type II AGNs as well as weak-lined objects. Both have the problem that spectroscopy is extremely costly especially for faint sources.

A commonly used method for AGN selection is through their **optical colors**. This method was first used by Markarian 1967 who selected objects with excess UV flux and created the first catalogue of nearby AGNs. For sources at higher redshift, the excess has to be searched in the bands that correspond to the UV rest-frame (i.e. the R -band for $z \sim 3$ and the H -band for $z \sim 4$). The main problem of this method is the contamination from Star

Forming Galaxies that can create UV excess. A color comparison could be used to detect high redshift AGNs.

Another AGN selection method, originally presented by [Zamorani 1985](#), was recently revived by [Fiore et al. 2012](#) as anticipated in Section 1.2.2. From a sample of sources, detected in optical band, AGNs are selected to be the ones with a X-Ray flux above a certain threshold.

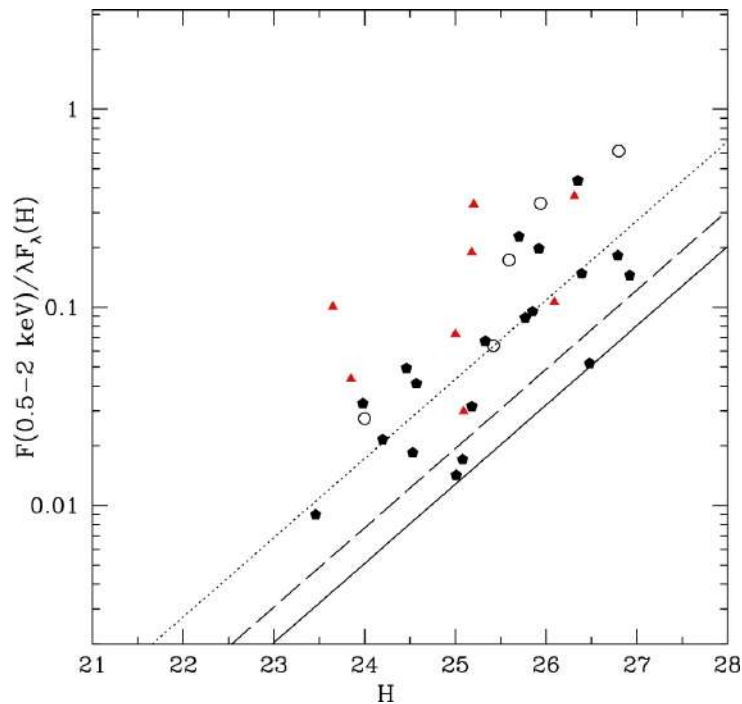


Figure 1.13: From [Giallongo et al. 2019](#) Figure 3: Dimensional X/H as a function of the H -band magnitude for some AGN candidates. Straight lines represent the adopted X-Ray flux limits of the three fields at the 40% completeness level adopted in their work

Figure 1.13 represents the **X-Ray/Optical flux ratio** method applied by [Giallongo et al. 2019](#) where H band was used. The straight lines indicate the locus of the constant X-Ray flux limits adopted. Only sources up to those lines are selected as AGNs.

Samples selected with this method are biased against relatively weak X-Ray emitters.

It is possible to select AGNs through the **Radio excess**, but only about 10% of AGNs show considerable radio emission. X-ray surveys have similar

problems to most radio surveys, the resolution is generally rather poor, causing problems with confusion.

Mid-infrared emission is thought to originate from warm dust in the obscuring dust torus surrounding the AGNs. However, also star-forming galaxies such as ultra-luminous infrared galaxies (ULIRGs) show excess mid-infrared emission and therefore selected samples are contaminated by SFGs, especially at higher redshifts.

Variability as detection option

AGNs show time-variable emission in every waveband in which they have been studied. The characteristic time-scales of variability ranges from hours to years, with the shortest time-scales being associated with shorter emission wavelengths. An alternative and promising AGN selection method is, therefore, using **variability** as search technique, i.e. detect all variable sources and later on select AGNs from other features.

The physical explanation of AGN variability was studied in the the last decades looking for links between variability properties and physical parameters of the system. One example is the study of [Sánchez-Sáez et al. 2018](#) who analysed connections of variability with nuclear black hole physical properties.

The short timescale - short wavelength relation can be understood in the context of current AGN models, where ultraviolet (*UV*) and optical emission are originated in an accretion disk around a super-massive black hole (SMBH), and non-thermal X-ray emission is produced in an inner hot plasma component (corona), which is geometrically much smaller and more concentrated than the accretion disk, and therefore able to show more rapid variability.

In their work [Sánchez-Sáez et al. 2018](#) found no dependencies between variation amplitude A and redshift z_{spe} , that means no evolution over cosmic time. The detected anti-correlation between A and λ_{rest} 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. On the other side, to explain the anti-correlation between A and L/L_{Edd} , confirmed in their work, different hypotheses were proposed (i.e. age effects, hotter disk), but each one presents some limitation and it is still on debate.

Variability selection has been used as a powerful tool for selecting AGN, in particular due to the emergence of large surveys in the last decades. It has enabled us to identify objects of intrinsically low X-ray to optical ratio, which would be otherwise missed by X-ray surveys and those with unusual spectral energy distribution (Trevese et al. 1989; Giallongo, Trevese, and Vagnetti 1991; Hook et al. 1994; Cristiani et al. 1996; Trèvese and Vagnetti 2002).

In general, the variability method may be complementary to the most common surveys based on optical colours or X-Ray emission, which suffer different biases.

Trevese et al. 2008 analysed the properties of AGN candidates selected in Chandra Deep Field South (CDF-S) on the basis of their variability from a photometric campaign, whose sampling rate and duration were studied especially for the detection of supernovae. They produced a good purity sample and was followed by a spectroscopic follow-up presented in Boutsia et al. 2009. Most of the variable sources are broad line AGNs (BLANGs), confirming the efficiency of the method; using color method on this sample some were missing by the X-Ray survey, while some other have extended images, and then have been missed by standard colour selection technique, which is restricted to point like sources.

Different statistical test methods were compared by Villforth, Koekemoer, and Grogin 2010 to select AGNs through variability in the GOODS fields, to verify false positive and completeness. Villforth, Sarajedini, and Koekemoer 2012 identified, by observing the SEDs of host galaxies of a variability selected sample, a high percentage of disturbed host morphologies and massive starbursts, demonstrating that commonly used selection methods would have missed a large number of these objects. Combining all those measurements together, an analysis of possible AGN triggering mechanisms was done, suggesting that minor mergers and triggering through first passage likely play a role for lower luminosity AGN (Villforth, Sarajedini, and Koekemoer 2012).

Variability can also be applied to extended objects, as for example low luminosity AGNs (LLAGNs), that cannot be detected by the (non-stellar) colour selection since their SED is contaminated (or even dominated) by the light of the host galaxy.

To support this hypothesis it is interesting the Baldwin-Phillips-Terlevich (BPT) diagram reported in Figure 1.14 from the work of Baldassare, Geha, and Greene 2018 who identified AGNs in local ($z < 0.15$) low-mass galaxy using variability selection on Sloan Digital Survey (SDSS) spectroscopy.

BPT diagrams, proposed by Baldwin, Phillips, and Terlevich 1981, are

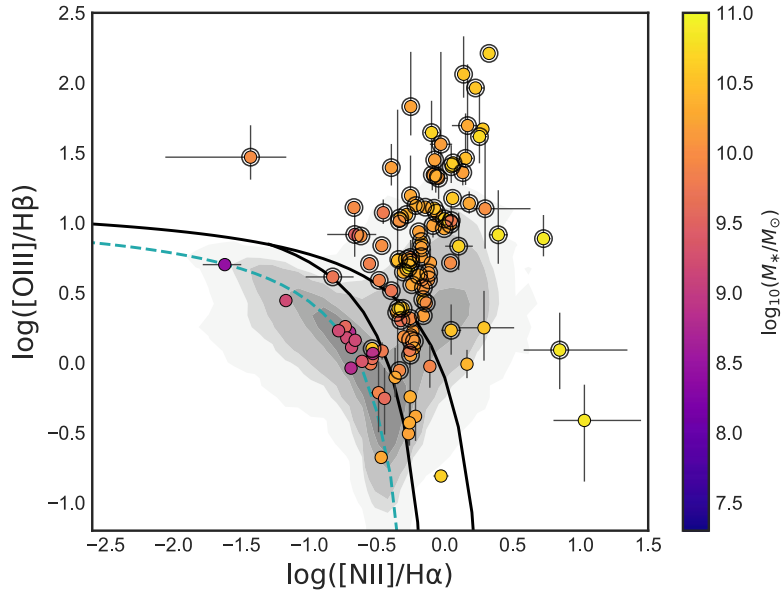


Figure 1.14: From [Baldassare, Geha, and Greene 2018](#) Figure 8: BPT diagram showing the positions of the variability-selected AGNs. The black lines show the traditional BPT diagram separation lines ([Baldwin, Phillips, and Terlevich 1981](#); [Kewley et al. 2006](#)), and the green dashed line shows the star forming main sequence [Kewley et al. 2013](#).

a set of nebular emission line diagrams used to distinguish the ionization mechanism of nebular gas. It can be used as empirical method to identify AGN and starburst based on line ratios. These diagrams were studied in numerous works and the separation lines have been developed and adapted as a function of the ionization models and/or observations available ([Kewley et al. 2006](#), [Kewley et al. 2013](#)).

The variability sample, selected by [Baldassare, Geha, and Greene 2018](#), is reported in a BTP diagram in Figure 1.14. The black lines show the traditional BPT diagram separation lines ([Baldwin, Phillips, and Terlevich 1981](#); [Kewley et al. 2006](#)), and the green dashed line shows the star forming main sequence [Kewley et al. 2013](#)). Data points are color-coded by galaxy stellar mass, obtained from the NASA-Sloan Atlas (NSA). 100 galaxies reside in the AGN or composite regions of the BPT diagram, while 20 are in the star-forming region.

The figure shows that for low-mass galaxies the host SED dominates over the AGN contribution showing typical SFGs color. Using common color se-

lection methods they are often wrongly discarded.

This capability to detect low-mass AGNs is very promising for the research of faint high redshift AGNs population.

1.3 The adopted strategy

In the previous sections we had an overview on the Reionization era and on some open questions about it. Here I describe the idea behind this work.

The idea is to test the efficiency of a variability detection method to find AGN, in particular faint AGNs, not previously detected with other selection methods.

Indeed variability seems a promising method to contribute to the high z AGN luminosity function definition and the consequent reionization open questions.

Considering the photometric dropout, to explore reionization redshifts ($z > 6$) with variability methods a set of observations in near-IR bands are needed. Those kind of surveys are not still available, even if some very interesting projects and instruments are in the final phases (i.e. *Rubin Observatory* and *Euclid Telescope*). I had to move then to the optical bands, the most diffuse and long time observed.

The purpose was to develop a robust variability method to select AGNs from available observations. The goal is then to test its results against the most diffuse AGN selections method (X-Ray), to evaluate its capability to identify sources and, eventually, characterize them.

The method described here is intended to be applied to future near-IR surveys to explore the reionization time.

I decided therefore to search for a very well known field of which are available as many as possible number of observations. The choice fell on the Chandra Deep Field South (CDF-S) that was observed many times along the last decades by the instrument Visible Multi-Object Spectrograph (VIMOS) mounted on VLT.

1.3.1 The Chandra Deep Field South (CDF-S)

The Chandra Deep Field South (CDF-S) is one of the deepest observations obtained with the X-Ray *Chandra* satellite, which stared at the same path

of the sky for 11 days.

The (first) final image was the result of the co-addition of 11 individual Chandra ACIS-I exposures obtained in 1999-2000, for a total exposure time of 1 Ms (Giacconi et al. 2002) showed in Figure 1.15.

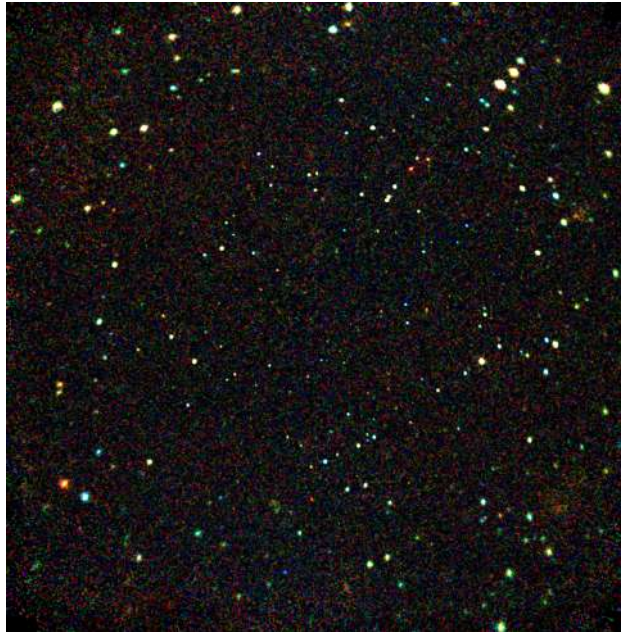


Figure 1.15: The CDF-S 1Ms image (Giacconi et al. 2002) located in the Southern Hemisphere constellation of Fornax. Most of the objects seen in the Chandra Deep Field-South are active galaxies and quasars powered by massive black holes. The intensities of the X-rays in this image are depicted, from lowest to highest energies, by red, yellow, and blue respectively

The nominal aim point of the CDF-S is $RA=3^h 32^m 30^{sec}$, $Dec=-27.0^\circ 48' 00''$ (J2000), covering 0.11 square degree.

The Chandra telescope has repetitively observed the CDF-S for more than 15 years, reaching a total observing time of 7 Ms. Different catalogues have been released during this period with gradually increasing observing time: 2 Ms Luo et al. 2008, 4 Ms Xue et al. 2011 and 7Ms Luo et al. 2017.

After the first observations, the CDF-S has become the center of the most comprehensive multi-wavelength campaign ever carried out with ground-based and space observatories. The two largest surveys are the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) and the Great Observatories Origins Deep Survey (GOODS).

The CANDELS is the largest project in the history of the HST, with 902

assigned orbits of observing time. It was carried out between 2010 and 2013 with two cameras on board Hubble – Wide Field Camera 3 (WFC3) and Advanced Camera for Survey (ACS) – and aims to explore galactic evolution in the early Universe.

The GOODS unites extremely deep observations from NASA’s Great Observatories, *Spitzer*, *Hubble*, and *Chandra*, ESA’s *Herschel* and *XMM-Newton*, and from the most powerful ground-based facilities (*VLT*, *VLA*, *ALMA*), to survey the distant universe to the faintest flux limits across the electromagnetic spectrum.

Both of these surveys include the CDF-S field and provide a very wide wavelength coverage from Radio to X-Ray.

Researchers used the CDF-S data in combination with data from CANDELS and GOODS, to study galaxies and black holes between one and two billion years after the Big Bang.

1.3.2 The Visible Multi-Object Spectrograph (VIMOS) instrument

VIMOS is a visible (360 to 1000 nm) wide field imager and multi-object spectrograph mounted on VLT UT3. The instrument is made of four identical arms each with a field of view of $7' \times 8'$ with a shift between each quadrant of $\sim 2'$.

VIMOS operates in three different modes: Imaging (IMG), Multi-Object Spectroscopy (MOS) and with Integral Field Unit (IFU).

The CCD size in Imaging mode, which I am interested in, is 2048×2440 pixels, with a pixel size of $0.205''$.

The Imaging mode is offered with a set of 6 broad band filters: U' , Bessel $BVRI$, Gunn z . Each arm of VIMOS is equipped with its own set of filters and some minor difference may exist in the transmission curve of the 4 quadrants, as shown in Figure 1.16.

After a deep analysis of CDF-S observations availability on the ESO Archive, I choose the VIMOS R -band for this work. This band offers the largest number of pointings with the longest time range. Indeed it was usually adopted as pre-imaging band for VIMOS spectroscopic observations, that ensured a good sampling along years.

In 2010 the instrument was equipped with a new deep depletion set CCDs, with exactly the same format, but with higher throughput in the red and

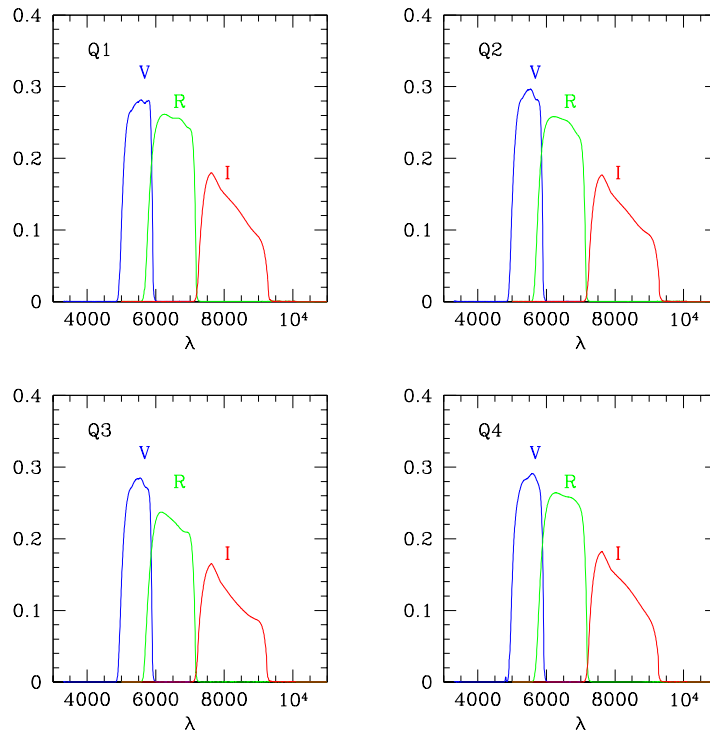


Figure 1.16: Bandpass transmission curves for VIMOS filters V , R and I for the four quadrants. The transmission value includes contribution from filters, CCD and telescope optics.

reduced fringing. This update provided, in our images, an improvement in the signal to noise ratio for source detection and the reduction of number pixels affected by instrument features.

Chapter 2

The Data Reduction

The purpose of this work is to select AGNs in the Chandra Deep Field South (CDF-S) region through their variability.

I have used the deep imaging database of the CDF-S that is available in the ESO Science Archive¹. In particular, I have chosen to analyse the *R*-band observations carried out with the VIMOS instrument at VLT, since these observations covers an extended time interval of 14 years, from 2002 to 2015.

In this chapter it is explained the implemented procedure to create a collection of CDF-S region images observed with VIMOS, for different years.

The procedure adopted to download VIMOS data is presented in Section 2.1 with the description of input data format.

Section 2.2 reports a brief overview of the calibration pipeline I wrote for those data, describing the pipeline general structure and its main phases. All single steps are, then, detailed explained in Section 2.3, starting from the creation of calibration data products up to the final image stacking phase.

A brief description of the produced stacked image is finally provided in Section 2.4.

2.1 Input data

The ESO Archive Facilities page provides an interface to make queries on ESO's instrument images.

I intended to retrieve all available CDF-S raw observations from VIMOS *R*-band, with corresponding raw calibration frames. I submitted a query for

¹https://archive.eso.org/eso/eso_archive_main.html

pointing within $15'$ from the center of CDF-S with the list of parameters proposed in Table 2.1.

Field	Value
RA	$3^h 32^m 30^{sec}$
DEC	$-27.0^\circ 48' 00''$
SearchBox	$0.0^\circ 15' 00''$
Imaging	VIMOS
Category	Science
Type	OBJECT
Mode	IMAGE
Filter	R

Table 2.1: List of parameters submitted to Eso Archive Facilities to query for VIMOS *R*-band pointing $15'$ around the center of CDF-S

I obtained a list of 1136 images observed between 2002 to 2015.

It is worth nothing that VIMOS files consist of single chip images, and therefore each pointing is split in 4 files.

A list of single BIAS and twilight FLAT frames, to properly calibrate data, were downloaded with the raw observations. I decided to discard observations from year 2002 because those calibrations files were not available.

Table 2.2 summarizes, for each year, the number of observations available, the pointing exposure time range in seconds and the period during the year when it was observed.

Some of the pointings were observed specifically for some Large Program, as the 167.D-0492 dedicated to Supernova (SN) research, and have an exposure time of 480 s with position angles -26° and -64° . Some other are indeed obtained as pre-imaging for spectroscopic programs with exposure time ranging from 150 to 530 s and different position angles.

2.1.1 Reorganize input data

As anticipated downloaded VIMOS files consist of single chip images. Therefore each pointing is separated in 4 files.

The pipeline for data calibration, that I have developed in collaboration with the staff from the Astronomical Observatory of Rome (INAF-OAR), was thought to work using one file for each observation. I had, therefore, to combine the four single chip images in a single Multiple Extensions FITS (MEF) file.

Year	N_{obs}	Exptime range (s)	Period
2003	38	200 - 300	Nov - Dec
2004	89	200 - 480	Jul - Jan
2005	32	200 - 480	Ago - Nov
2006	4	150	Oct
2007*	5	300 - 530	Sep
2008*	3	200	Jul
2009*	12	250 - 1950	Ago - Sep
2010	4	175 - 600	Oct
2011*	9	365 - 600	Ago
2012*	9	360	Nov - Dec
2013*	25	180 - 600	Ago
2014*	31	240 - 550	Sep - Oct
2015*	15	550	Jul - Ago

Table 2.2: Number of observation for each year, the exposure time range of the images and coverage period during the year. Asterisks indicates years I was able to calibrate and use for my analysis.

From the analysis instrument data flow it was evident that observations were initially saved in a single file containing all chips. Later on it was separated in single chip images to be stored in the Archive saving original filename in `ORIGFILE` file header keyword.

To do this I wrote a Python script, `organizeData.py`, to reconstruct MEF files combining single chip images on the base of `ORIGFILE` keyword value and `DET.CHIP1.ID`, that stores the CCD name, to identify their position/order. It is worth nothing that, due to the CCD upgrade in 2010, the CCD name changed.

This script runs on a list of files and produces MEFs only for “complete” observations, meaning when all four chips are available.

2.1.2 Epoch selection

In case of scientific pointings all the four chips were always available and I could easily reconstruct 276 MEF files in the epoch interval 2003 - 2015.

This is not true for calibration images `BIAS` and `FLAT`. Indeed some of them were dedicated to spectroscopic calibrations, or the observation target was in a single chip and therefore calibration images were provided just for one chip.

Consequently for some epoch I was able to reconstruct just a fraction of the calibration images. In case of 2006, for example, no a single raw **FLAT**, associated with CDF-S observations, was provided complete (with all 4 chips) by the ESO Archive. The 4 available images for this year are pre-imaging for spectroscopic observations and were acquired **FLAT** frames only for some of the detectors. Due to the lack of these calibration files I could not use images of some epochs.

I had to discard the first 4 years, preserving pointings from 2007 to 2015 but 2010, for a total of 8 epochs, pointed with an asterisk on Table 2.2.

2.2 Pipeline Description

The purpose of the pipeline is to create a final stacked image using raw scientific observations. In this section I'll present an overview on the pipeline structure and data flow. Some more advanced details on each pipeline step will be provided in the next section.

For the calibration process of scientific images two calibration data product are needed, the MasterBias (MB) and the MasterFlat (MF), to track the electronic noise and the telescope wave-front distribution respectively. Those have to be created from raw calibration frames. The dark current for VIMOS *R*-band is negligible and no **DARK** frames were measured for these observations.

The CDF-S field was typically observed, by VIMOS, from July up to December, as reported in Table 2.2. During this period instrument features can change and therefore it is needed to create different MBs and MFs to track the instrument behaviour internally in the year. This is especially needed in case of the MF, whose features are extremely variable.

I created separated lists of raw **BIAS** and raw **FLAT-SKY**, according to their creation dates, to produce different data products. Each scientific image was then associated to the closest MasterBias and MasterFlat.

Therefore, for each epoch, I collected data in three different “kinds” of MEF files lists:

- lists of raw **BIAS** frames to create MasterBiases
- lists of raw **FLAT-SKY** frames to create MasterFlat
- list of raw **OBJ** frames, that are the scientific images

The pipeline software can manage more input lists at once. It was then possible to calibrate and stack together all scientific frames for each year, associating different MasterBias and/or MasterFlat as calibration data.

2.2.1 General Structure

The pipeline is composed by different steps (or tasks) that can be activated manually through a configuration file.

So it can be executed step by step or all at once.

Some steps are based on a collection of tools for imaging and spectra calibrations, hosted at servers of Osservatorio Astronomico di Roma (OAR).

The collection is composed by the most diffuse suites for astronomy, as *AstrOmatic*² and *AstrometriC*³, and some other tools developed by Diego Paris, Researcher at OAR.

Software algorithms are optimized for a better performance on large images, or set of images, with a dedicated memory management.

This collection provides some useful procedures for image analysis and calibration. They are used in the calibration pipeline for Large Binocular Telescope (LBT) data and other instruments and are, therefore, a good starting point for my purpose.

The pipeline I wrote to calibrate VIMOS data is composed by two files:

- i. a **Python file** with the **pipeline software**. Here I wrote methods to manage each single step and a main function dedicated to the interface and data managements, that calls all the steps consequentially
- ii. an **xml file** used as **configuration** to set parameters for each task. Its structure reflects the pipeline one and contains a *main tag* dedicated to each step. The steps can be activated with a common tag **active** present for all steps.

In that way it's possible to execute the entire pipeline or just single steps on different data through a configuration file.

²Suite of astronomy tools developed by E. Bertin <https://www.astromatic.net>

³A C++ package for the astrometric calibration of wide-field images developed by Mario Radovich of INAF - Osservatorio Astronomico di Padova

```

1 <flat description="Masterflat creation. 'active' tag switches
   on/off the procedure. 'blank' is the output value to be
   used when there are no pixels (i.e. clipped, see below). '
   lower' and 'upper' are respectively the low and the high
   threshold to be applied to the output pixels. Pixels above
   this threshold are set to blank value. [...] Obviously
   input masterdarks prefixes number, output masterflat
   prefixes number and output masterflat bad pixel mask
   prefixes number must match input flat lists number.">
2 <active type="boolean">no</active>
3 <verbose type="boolean">yes</verbose>
4 <blank type="string">1</blank>
5 <lower type="string">0.85</lower>
6 <upper type="string">1.5</upper>
7 <reject type="string">sigclip</reject>
8 <nsigma type="string">3.0</nsigma>
9 <masterdarks type="string">masterbias.1.fits</masterdarks>
10 <flatlists type="file">flatlist.1.txt</flatlists>
11 <masterflats type="string">masterflat.1.fits</masterflats>
12 <masterflatbpms type="string">masterflat.1.bpsm.fits</
   masterflatbpms>
13 </flat>

```

Listing 2.1: Example of pipeline configuration file: main tag containing MasterFlat creation step parameters

The code in Listing 2.1 is an example of a *main tag* for the MasterFlat creation step. Its first tag, `active`, common to all steps, determines whether it should be executed or not. To run the pipeline, therefore, I had just to activate steps I'd like to execute and pass these parameters through a configuration file.

The Python main script checks which steps are activated and calls the relative functions. Each step retrieves input parameters and file-lists from the corresponding *main tag*, verifying their integrity and compatibility, and finally calling executables to perform task activity.

2.2.2 Overview

The purpose of the pipeline is to create a calibrated stacked image from a collection of scientific pointings and relative raw bias and flats frames.

The process flow can be divided in 5 different phases.

In Figure 2.1 all the pipeline steps are collected in dashed lines that identify the different phases. Red color was used for calibration pipelines and blue for the scientific one.

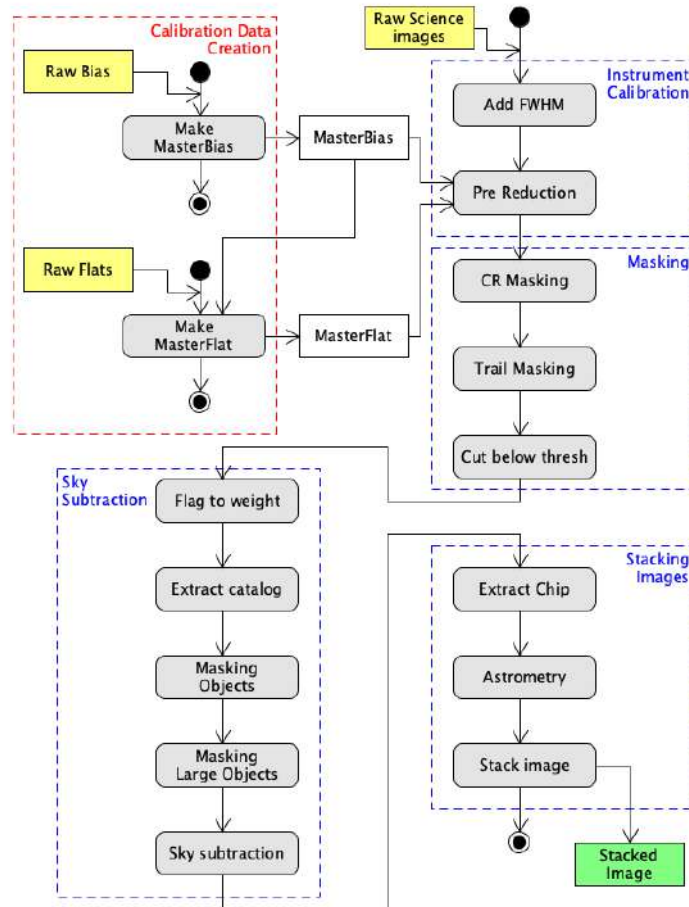


Figure 2.1: Representation of Data calibration pipeline for VIMOS instrument

Next a brief overview of phases content and meaning.

Calibration Data Creation

The first phase is composed by two independent calibration pipelines for MasterBias and MasterFlats production.

The first gets as input a list of raw BIAS frames and creates a **Master-Bias**, rejecting bad pixels with a σ -clipping. BIAS images are observed with daily cadence to measure signal due to the electronics.

The second task is dedicated to **MasterFlat** image creation. It gets a list of raw FLAT-SKY, removes the bias noise, trims out outliers, stacks

images and normalizes the final products. MasterFlats are produced from raw FLAT-SKY, meaning flats obtained observing the sky during the twilight. Therefore they reflect full instrument contribution to the pixel response, both detector quantum efficiency and full optics contribution.

Instrument Calibration

This is the first phase of scientific pipeline dedicated to instrumental effects removal. The source FWHM value of each image is estimated from metadata stored in the input configuration file, and it's written in the relative file keyword. Successively images are subtracted of the MasterBias, to remove electronic noise, and divided for the MasterFlats, to calibrate for the single pixel efficiency.

A new set of calibrated images are produced at the end.

Masking

This phase gathers some tasks to flag contaminated pixels from different contributions.

The Cosmic Ray Masking task calls iteratively a software, based on [van Dokkum, Bloom, and Tewes 2012](#) algorithm (*L. A. Cosmic: Laplacian Cosmic Ray Identification*), to detect and flag pixels with spike signals due to a CR hit. A new `bad pixel mask` (bpm) file is created to store flags.

Satellite trails are flagged manually. These features are not easy to be detected automatically, but, on the other hand, they are simple to be identified at first sight and are not so recurring. Therefore I decided to detect them visually, find their coordinate and automatize the flagging phase. So, after the first time, this task can be executed automatically.

Due to the complexity of instrument optics the focal plane is not flat and it is slightly vignetted on the corners outside 8'. The MasterFlat ratio creates on the corners and along borders, some spurious signals. To avoid this feature I developed a task to mask pixels that present a signal lower than a fraction of the median image value.

Sky Subtraction

Up to this point images are calibrated for instrumental effects and some sky effects easily detectable. This phase is dedicated to the evaluation and subtraction of the sky contribution to the images, masking all sources.

As first I converted the flag image in a weight map, containing a value of 1 for good pixels and 0 for flagged ones. It is used by the software *SExtractor* during the extraction of the source catalogue.

I created an `object mask` starting from the *SExtractor* segmentation map, produced during the source extraction. A dilatation factor around each detected object is applied to mask conservatively all the pixels belonging to sources.

For very large objects a greater dilatation parameter has to be applied in a second step. They are selected according to their estimated dimensions and a more extended area is masked around them. Parameters for this steps have to be tuned to avoid to masking also the sky background.

The `object mask`, together with the `bad pixel mask`, identify all pixels that are not usable to estimate the sky contribution. I merged them together in a new single mask.

With this mask the pipeline runs *SExtractor* again to obtain the sky background image. It's possible, then, to subtract the sky image from the scientific exposure to obtain a `sky-subtracted image`, where only source contribution is present.

Stacking images

At this stage images are photometrically corrected and the sky was removed, leaving only sources signal. Images are then ready for the stacking.

The astrometric calibration step has to precisely reconstruct the observations coordinates to proceed with a correct stacking step.

Thanks to the wide instrument field of view, a large number of sources are detected on each detector, enough for a good astrometry. I decided to split MEF files into single-detector files, to allow to the software a fine astrometric calibration. In that way I removed constrains due to the detector positions on the field of view according the instrument model. Each detector is then calibrated independently.

The astrometric calibration is performed in some steps: a new catalog is extracted from detector images and it is compared with a reference catalogue (USNO, GSC, UCAC or SDSS). From cross match results between extracted and reference catalogues, a better astrometric solution is calculated to minimize distances between sources. The final astrometric solution is, at last, written on the final files according to the World Coordinate System (WCS).

The stacking process starts with the creation of a common pixel grid.

Each detector is arranged in the right position on that grid according to the WCS metadata and it is then resampled. The stacking of these resampled images creates the final image.

This task can be run using the sky subtracted images or even with the original images. In this last case the resampling step is preceded by the creation of an RMS mask.

2.3 Some more details on the pipeline

Here I will report a detailed description of pipeline phases and their internal tasks.

2.3.1 Calibration Data creation

The first two tasks are dedicated to the production of calibration data.

They are, technically, two independent pipelines: they use a list of raw **BIAS** and raw **FLATS** as input and produce **MasterBias** and **MasterFlats** as output.

It's necessary to create first **MasterBiases** and next **MasterFlats**, that use them as input.

MasterBias

The task to create **MasterBias** gets as input a list of **BIAS** files. Bias is the term used to describe a CCD camera's pixel-to-pixel variation in signal at zero exposure time. Each pixel has a slightly different base value that is estimated through bias frames. A bias frame is, ideally, an exposure of zero length, and therefore the **EXPTIME** in the header is set up as 0.0.

The bias signal does not vary a lot with time, so I decided to combine together all data available for the year in a single product. Usually were observed sets of 5 **BIAS** frames at a time, so I had stacked from 5 to 40 bias frames for each epoch.

The **MasterBias** function calls the executable `mkmasterdark`, developed by Diego Paris, that applies a σ -clipping ($< 3\sigma$) to remove outliers and combine **BIAS** images to create a final **MasterBias** with a median stacking.

MasterFlat

MasterFlat data product provides an estimate of the single pixel response. It tracks the pixel-to-pixel difference in measured value from an identical signal.

The purpose of MasterFlat data product is to equalize the whole image to measure the same signal independently where it is located on the detectors.

Assuming the same input light flux, the single pixel response depends on its quantum efficiency and on the fraction of light that reaches the pixel. This latter contribution depends on the complexity of the instrument optic.

To measure this feature images with a uniform illumination are used. Those images are taken observing calibration lamps or the dome wall or the sky during twilight, as in the case of VIMOS. This kind of flats, called FLAT-SKY, consist in short expositions during twilight, when the sky light is flat enough. They permit to estimate the contribution from the entire instrument, including the most external parts of the optic.



Figure 2.2: The first detector of a raw Sky Flat of 2009

Some features are clearly visible on the raw FLAT-SKY detector in Figure 2.2. I'm referring to the bad pixel column and a lower responsive circular area on the top of the detector. The VIMOS focal plane is not strictly flat because of the complexity of the optical elements involved. The gradient from the top left corner to the right bottom is due to this distortion.

Even with the very short exposure time (~ 1 sec) some sources are clearly

visible, in particular on the left side of the image. To delete their contribution in the final MasterFlat, `FLAT-SKY` are observed in set of 6 with a small dither between them.

Because of the large number of contributions to the single pixel response, the MasterFlat can change with time. To track this temporal evolution I created, for each epoch, a set of different MF. Then, each scientific observation will be associated with the closest MF.

The executable `mkmasterflat`, developed by Diego Paris, implements an optimized algorithm. The Python function parses input parameters from the configuration file and calls this task. An example of configuration *main tag* is reported in Listing 2.1 with a brief task description.

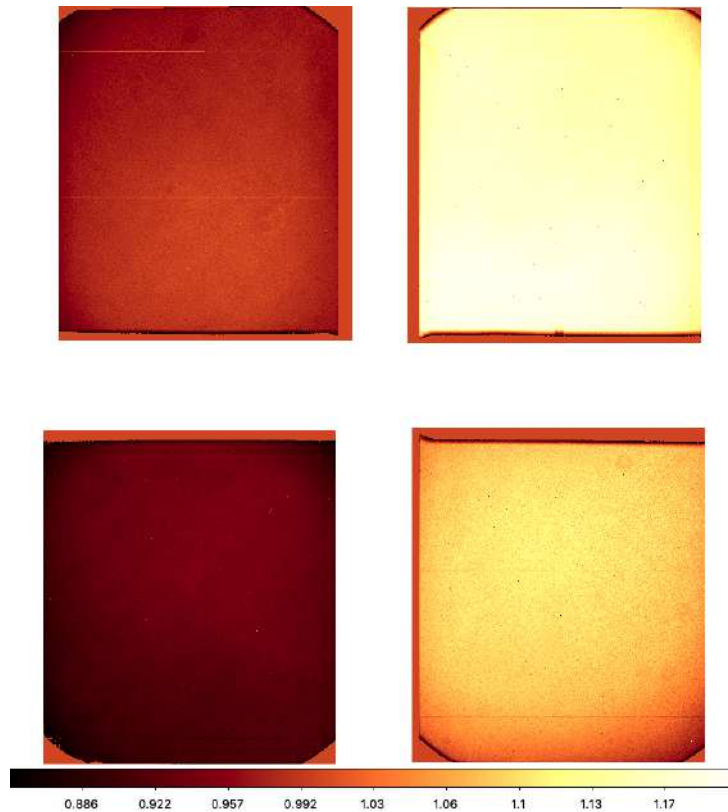


Figure 2.3: One of the MasterFlats created for year 2009

First the MasterBias is subtracted from the raw `FLAT-SKY`.

A `rejection` σ -clipping method around the median is used to identify outliers. It flags all pixels above and below `nsigma=3.0`, excluding them

from the final combination: it avoids CR hits and sources contribution to MasterFlat. Bad Pixels result to be outliers in all pointings, so they are flagged with the `blank` value specified in the configuration file, (1 in our case).

After the flagging phase all single `FLAT-SKY` frames are median stacked together and normalized to 1. Finally all pixels with a value below `lower` or higher than `upper`⁴ are flagged and treated as Bad Pixel. MasterFlats are then produced together with their relative `bad pixel masks (bpm)`.

One MasterFlat for 2009 is shown, as example, in Figure 2.3. In this image detectors have different colors due to the median difference among detector response. This feature hides some other global features, as the light concentration in the center, that is instead visible in Figure 2.2 due to the optics. We can even recognize small scale features, as the lower efficiency spot on the top of the first detector. The detector borders and the lower efficiency zones are flagged in the `bad pixel mask` and their value is set to 1.

2.3.2 Instrument Calibration

This is the first phase of Scientific Calibration pipeline. It gets as input a list of raw `OBJ` scientific pointings.

In Figure 2.4 a raw pointing of the year 2009 is showed as example. This image contains some of the most common features that the pipeline corrects or flags.

The black frame all around detector borders are due to the 50 pixels unexposed for bias estimation, while corners are affected by vignetting. The dark area on the corner of top right detector is due to vignetting, but is visually amplified by the color scale adopted. A satellite trail passes through the two detectors below and some ghosts and diffuse lights are clearly visible in the image.

Add FWHM

The first task is dedicated to estimate the Full Width at Half Maximum (FWHM) from metadata.

⁴`lower` and `upper` refer to the MasterFlat configuration file as in Listing 2.1

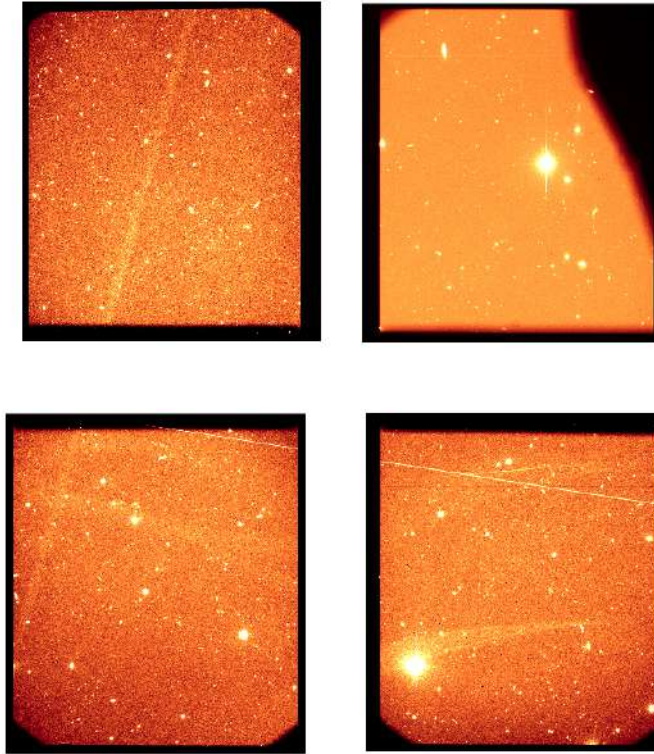


Figure 2.4: Raw pointing observed in 2009, August 18th

The expected FWHM is derived from the DIMM seeing⁵ measurements that are stored in the header of the science frames,. The keywords `FWHMSTART` and `FWHMEND` store DIMM seeing measurements at start and at the end of the exposure, respectively.

I used them to have a first estimate of mean FWHM.

$$\text{FWHM} = \frac{\text{FWHM}_{\text{START}} + \text{FWHM}_{\text{END}}}{2 * \text{PIXEL_SCALE}}$$

The DIMM seeings measured is typically $\sim 1''$ while the VIMOS `PIXEL_SCALE=0.205''`. I expect therefore a FWHM value ~ 5 pixels.

⁵The DIMM seeing is defined as the Full Width Half Maximum of a stellar image observed with a perfect large telescope, at 500nm wavelength and at zenith. It is measured over periods of a few minutes all night long.

Pre-Reduction

The pre-reduction phase collects all the processes that have to be applied to an exposure to correct for instrumental effects due to electronics, CCD and telescope optic. They are the first bunch of steps in a calibration pipeline for any instrument.

For this pipeline I used the software `fors2prereduce`, developed by the LBC Team, that includes all needed passages with a single executable.

These are the main elaboration steps:

- i. images are trimmed to exclude the bias section around detectors
- ii. MasterBias is subtracted
- iii. all pixels with a value above a saturation level (65000 ADU in our case) are flagged
- iv. images finally are divided for MasterFlats

At the end of this task images are corrected for effects due to detector electronics and a new set of `pre-reduced images` is created together with a `bad pixel mask (bpm)`.

2.3.3 Masking

During this phase the mask associated to each image is updated to keep track of pixels affected by features. This phase includes the detection of Cosmic Ray hits and trails left by satellite passages during the observations. At the end also detector portions with a low signal are masked.

CR Masking

The identification of Cosmic Ray hits on images is typically a very demanding task due to the intensive calculus needed. For this step I used the executable `crmask` developed by LBC Team.

The algorithm is based on L.A. Cosmic⁶ algorithm proposed by [van Dokkum, Bloom, and Tewes 2012](#) that can be iterated more times for a deeper research. I choose to apply a double iteration sufficient to detect large fraction of CR hits on my images.

The flag `8` is used to identify CR hits on `bpm` image.

⁶cfr <http://www.astro.yale.edu/dokkum/lacosmic/>

Trail masking

Satellite passages on the field of view during the observations left a long bright and straight trail along detectors, often more than one (see Figure 2.4). They are clearly identifiable by a visual inspection, but they are not so trivial to be automatically detected with an algorithm.

For that reasons they are generally masked manually. I created a task to automatize the flagging phase.

Satellite trails pass through the whole detector, from one side to the other, leaving a straight line. Therefore to mask it completely I need just to set 3 values: the coordinates of two points belonging to the trail and a value for the mask line thickness.

The first time that I reduce a new image I have to stop the pipeline and visually analyse it. I had to identify trails, retrieving the coordinates of two points, and choose a value for the mask thickness according the trail brightness. Then I could add these informations in the configuration file. After this first iteration the task can be re-executed automatically.

The task reads those metadata and calls the `imreplace` executable that masks the line identified by pixel coordinates with the thickness, in pixel, indicated.

```
1 <trailmasking description="The procedure to flag trails">
2   <active type="boolean">no</active>
3   <verbose type="boolean">yes</verbose>
4   <trailflag type="string">1</trailflag>
5   <files description="list of files">
6     <file>
7       <name type="string">VIMOS_IMG_OBS227_0018.p.fits</name>
8       <detector type="string">3</detector>
9       <command type="string">-T [16:2094,1007:2251,10]</
10      command>
11     </file>
12     ...
13 </files>
14 </trailmasking>
```

Listing 2.2: Example of *main tag* for trail masking

An example of trail flag *main tag* on configuration file is showed in Listing 2.2. For each detector affected by satellite trail I added a tag `file` containing the filename, the detector number and the command to be passed to `imreplace` task. It contains points coordinates and line thickness. Trails are flagged with **1** value on **bpm** image.

The `-T` at the beginning of the command indicates the line shape that I used to mask trails. With other parameters this task can be, potentially, used to mask features with other shapes.

Cut below threshold

A third masking task is dedicated to a different feature, not originated by external sources. The focal plane of VIMOS is not strictly flat because of the complexity of the optical elements involved and some portions of images show a low signal. When the MasterFlat is applied some unexpected features can be created.

To avoid those features I created a new task to mask image pixels with a signal lower than a threshold.

```
1 <cutbelowthresh description="This step flag all values below
  a specific thershold respect median; it searches for the
  object mask in the same folder to exclude bad pixel and
  object from median count">
2   <active type="boolean">no</active>
3   <inputlists type="file">scilist1.p.txt , scilist2.p.txt ,
  scilist3.p.txt , scilist4.p.txt</inputlists>
4   <threshold type="string">0.90</threshold>
5   <belowflag type="string">16</belowflag>
6 </cutbelowthresh>
```

Listing 2.3: Example of configuration tag for masking below a threshold

It gets as input the images and the corresponding bad pixel masks and calculates the median. Then the task masks pixels with values below a certain fraction of the median. I decided to mask all pixels with a value $< 10\%$ of the median.

These pixels are flagged with value **16**.

This task can also be executed in a more advanced pipeline phase, after the object mask creation. In that case it can use the object mask for the median calculation.

2.3.4 Sky subtraction

This phase is dedicated to sky estimation and subtraction.

I had to identify all pixels not usable to estimate the sky (bad pixels, cosmic rays, trails, sources) to create a mask. I could then evaluate pure sky background, that I intended to subtract from images, masking those pixels.

Flat to weight

During previous steps it was created a mask to track bad pixels, CR hits, diffuse light and low flux pixels.

This simple step transforms the mask pixel in a binary weight map (**wht**). I used the `flag2weight` executable of LBC team to set as 0 all flagged pixel and as 1 all the others. It can be used as weight map for source extractions.

Extract catalogues

I expect, then, that not flagged pixels could just contain sources or pure sky. My purpose was to estimate the pure sky contribution, so I extracted a catalogue to find also pixels “affected” by sources. The idea is to create a mask to flag them.

Using the weight maps as input I run the software *SExtractor* on the image to identify objects and create a catalogue.

Next I report an example of extracting catalogue *main tag* to manage *SExtractor* configuration parameters and other metadata for this task:

```
1 <makingsexcatandupdateFWHM description="This step creates
  sextractor catalogs, and updates a more accurate FWHM for
  each image of input lists. The 'active' tag switches on/
  off the procedure and 'updateFWHM' switches on/off the
  procedure to update the FWHM into the fits headers.
  Multiple input file lists must be separated by commas. '
  detect_sigma' determines the threshold used by sextractor.
  A default value for FWHM must be given to avoid errors. '
  back_size' and 'back_filtersize' and 'sexconf', '
  sexcatparam', 'sexnnw' configurations files also must be
  given. WARNING: the 'updateFWHM' tag needs a reference
  catalog (supposed to be created indipendently). [...] If
  images have no astrometric solution, please skip this step
  .">
2 <active type="boolean">no</active>
3 <inputlists type="file">scilist1.p.txt,scilist2.p.txt,
  scilist3.p.txt,scilist4.p.txt</inputlists>
4 <detect_sigma type="string">3.0</detect_sigma>
5 <fwhm_key type="string">FWHM</fwhm_key>
6 <fwhm_default type="string">5.0</fwhm_default>
7 <back_size type="string">256</back_size>
8 <back_filtersize type="string">3</back_filtersize>
9 <sexconf type="conf">/scisoft/Pipeline/scripts/config/sex.
  conf</sexconf>
10 <sexcatparam type="conf">/scisoft/Pipeline/scripts/config/
  sexcatparam.txt</sexcatparam>
```

```

11 <sexnnw type="conf"/>/scisoft/Pipeline/scripts/config/sexnnw
    .txt</sexnnw>
12 <updateFWHM type="boolean">no</updateFWHM>
13 <referenceCatalog type="conf">SELSTAR.dat</referenceCatalog
    >
14 <radius type="string">3.0</radius>
15 <nsigma type="string">3.0</nsigma>
16 <niter type="string">16</niter>
17 </makingsexcatandupdateFWHM>

```

Listing 2.4: Example of configuration tag for sources extraction

As reported in the *main tag* description, the set of parameters in Listing 2.4 can be tuned to optimize extraction.

It is possible activate to the option to update the FWHM source value at runtime using measures from extracted sources . An estimated FWHM was previously stored in the FITS files header (cfr. Section 2.3.2).

To activate this feature an astrometric solution for images is needed, but we did not have yet. For that reason I did not update the FWHM.

Masking Objects

To create a mask for sources I needed to estimate how much the sources are extended. I decided to use the segmentation map produced by *SExtractor* and apply a dilatation factor to include all pixels affected by source light.

Then I merged this source mask with the bad pixel one to obtain a mask that identifies sky pixels, that are pixel with just the sky contribution.

The step procedure can be summarized in the next points:

- i. the executable `mkkernel` convolves a kernel with the FWHM value. This result has to be used during the extraction as source shape
- ii. a tuned run of *SExtractor* produces a catalogue with the corresponding segmentation source map. The configuration parameters are similar to the one described in the Extraction Catalogue paragraph
- iii. with the executable `sgm2flag` the segmentation map is converted into a mask where all pixels belonging to the sources are flagged
- iv. the segmentation map does not include all pixels affected by source light. Therefore the sources flags are extended, using the `dilate` executable, to include neighbouring pixels. A **source mask** (`msk`) is created at the end

- v. finally an **object-bad pixel mask (obm)** is created merging the **msk** and **bpm**, to obtain a mask to select only sky pixels

In case of very large objects the previous dilatation step is not sufficient to mask completely the source influence area. A second dedicated task is then used only for very large objects whose detection area (larger then 2000 pixels in our case) has to be dilated using a greater parameter. The aim of using the standard procedure together with this procedure for large objects is to perform a differential object detection.

The **msk** and **obm** masks were finally updated.

Sky subtraction

At this point the pipeline has created the **obm** mask to identify both bad pixels and sources. The **obm** was converted in weight map, with 0 where there was a flag and 1 where not. In that way pixels with only sky contribution were selected.

A run of *SExtractor* using this weight map creates the **sky** image.

This **sky** image is then subtracted from the scientific frame to create the **sky subtracted image (sks)**.

2.3.5 Stacking Images

Last steps removed sky from images leaving only source contribution. To proceed with stacking I needed first to find the correct astrometric solution for **sks** to align sources.

Extract chips

Up to this task the 4 detectors of a single pointing were stored together in a single MEF file. This approach was very useful for the data management. I now had to look for a fine astrometric solution and it was more convenient to work on a single detector.

This procedure splits chips of the **sks** MEF and associated weight maps. It produces 4 files from each image and names them combining the object observed and the filter band.

Images are extracted and grouped according to the same rotation angle, in order to compute an independent astrometric solution.

Astrometry

This step evaluates the astrometric solutions for **sks** images. It compares the catalogue extracted from an image with a reference catalogue and derives the geometrical solution to minimize distances. The solution is provided in a set of keywords, according to the WCS standard, to store the set of polynomial coefficients that describes the optical distortion.

The procedure implemented gives the opportunity to query for a reference catalogue that is automatically retrieved from archives (USNO, UCAC, GSC, SDSS). I decided to use the USNO catalog, its CDF-S portion was downloaded from servers with an automatic query on image pointings coordinates.

The astrometric solution is calculated comparing the reference catalogue with the one extracted and cleaned⁷ from each image.

I used some executables from the software suite *AstrometriC*, developed by M. Radovich. Different tools are here provided to cross match catalogues, minimize sources distances and evaluate distortion parameters.

The procedure can be summarized in the next steps:

- i. it makes cross-match between the image catalogue and the reference catalogue
- ii. it finds a preliminary solution chip by chip minimizing distances
- iii. it finds a general solution for the whole mosaic at once
- iv. it writes individual solutions for the four chip on new header files

As output this procedure produced a list with names of file for which a good astrometric solution was found. Images with a bad astrometry are finally rejected. For my sample this procedure was able to find a good astrometric solution for all images.

Headers containing astrometric solutions are saved as separate files (with extension *.head*) that can be used by resampling programs as *SWarp*.

Stacked images

The step is dedicated to image resampling and to the creation of a final stacked image.

⁷in the cleaned catalogue I considered only sources with `FLAGS < 4`, `MAG_AUTO < 99` and `ELLIPTICITY ≤ 0.5`

The resampling and stacking (coadding) process is executed with the software *SWarp* of Astromatic suite. A dedicated configuration file `swarpconf` must be provided together with the input image list.

It is possible to choose between two alternative resampling and coadding algorithms according to the map used: the `weight map` or the `RMS map`.

In the first case `weight maps` are used for resampling and stacking. The procedure consists in a sequence of *SWarp* runs with different configurations:

- i. on the base of images WCS a common grid is defined for the output frame. Its definition is stored in an header file to be used as reference for resampling
- ii. images and weight maps are resampled according to the common grid with the same instrument pixelscale of 0.205”
- iii. resampled images are coadded using relative weights maps and two different algorithms defined in `COMBINE_TYPE`:
 - (a) `MEDIAN` the output is a median of all pixels values scaled with non-zero weights. This is a robust coaddition method for data polluted by spikes and noise
 - (b) `WEIGHTED` the output is a weighted average of input values. This combination is the most appropriate for detecting and measuring faint sources on properly weighted and homogeneous images

Two image are created at the end, a first “safe” image to check results in noisy parts, and a second one that is intended to be more interesting for scientific purpose.

In the case of `RMS map` algorithm the procedure needs an input image that stores absolute standard deviation of pixel values. This rms map has to be built from raw images and MasterFlat(s).

The procedure is similar to the previous case with two main differences:

1. at the beginning from images and MasteFlat(s) it is calculated the RMS map containing the absolute standard deviation of pixel values
2. images are resampled and coadded using the RMS map for accurate weighting phase

A final stacked image is produced at the end.

In both cases within the stacked image a corresponding flag weight map is provided.

I choose the second option for the pipeline.

2.4 Output Products

As described on Section 2.1.2, some of the epochs downloaded were provided with a incomplete set of calibration data. I could use only the 8 complete epochs, from 2007 to 2015 except from 2010.

I executed the described pipeline for those epochs combining all observations available to create a single stacked image per year. This procedure, as described later, flats all fluxes variations within the year that I will not able to detect. On the other hand it guarantee us images enough deep to detect faint sources, that I am interested in.

In Figure 2.5 the final result of the calibration pipeline for the year 2009 is presented. On the top is reported the stacked image, combination of 12 observations. On the bottom the corresponding weight map is shown.

Images were obtained coadding frames with different depth. As a result the depth for stacked frame is not homogeneous. Some image portions are more faint and their weight is very low, while the most deep parts have an higher weight.

Due to this large inhomogeneity, during all my analysis I considered only sources extracted with a signal-to-noise ratio $(S/N) > 5$. This conservative approach ensured us a right source magnitude association and minimized the risk to introduce false variability detections.

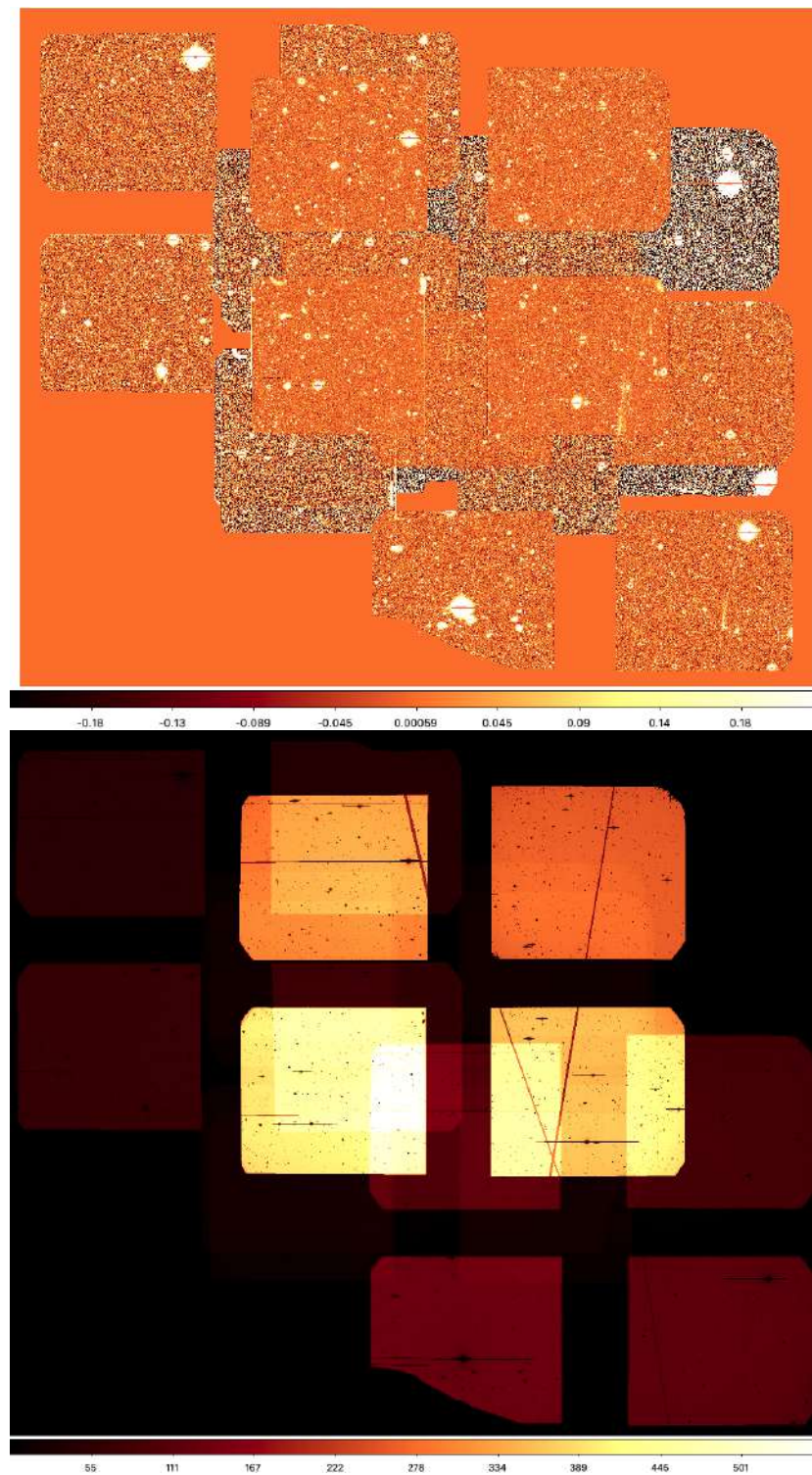


Figure 2.5: Top: Stacked image for the year 2009. Bottom: corresponding weight map

Chapter 3

Detection of variables AGNs

In this chapter it is explained the implemented procedure to identify variable AGNs in the Chandra Deep Field South (CDF-S) region.

At this aim, I have used the deep imaging database of the CDFS that is available in the ESO Science Archive. In particular, I have chosen to analyze the *R*-band observations carried out with the VIMOS instrument at VLT, since these observations covers an extended time interval of 14 years, from 2002 to 2015.

The detailed procedure adopted to reduce the raw science images of VIMOS, that I downloaded from the ESO Science Archive, has been described in detail in Chapter 2. For each year of observation, a stacked image has been produced by summing all the available VIMOS images of the CDFS field for years from 2007 to 2015.

A source catalogue has been extracted for each stacked image, as I will describe in detail in the next Section 3.1. Then the catalogues were merged to create a single catalogue with multi epoch detection for each source, as described in Section 3.2.

Section 3.3 and Section 3.4 are indeed dedicated to the statistical method implemented to identify variable sources.

Finally the catalogue of variables identified with my method is described in detail in Section 3.5.

3.1 Catalogue extraction

Catalogues were extracted from stacked images using the software *SExtractor*¹ of the *AstrOmatic* suite. A detailed description of this tool is available

¹For more info <https://www.astromatic.net/software/sextractor>

on the paper [Bertin and Arnouts 1996](#). This is a very powerful detection software that can be highly tuned according to images properties and extraction desiderata.

A *configuration file* to set input parameters and a *parameter file* to select the catalogue output columns are needed to process images.

The *SExtractor* input for my analysis is composed by:

- **images/maps:**
 - **scientific:** stacked image
 - **weight map:** image with weight pixel value
 - **flag map:** image with flags for not valid pixels (saturated, bad pixel, CR hits, trails)
- **configuration file:** an ascii file containing input parameters to configure the extraction
- **parameter file:** list of columns requested in the output catalogue

The configuration file contains a list of parameters to tune the different phases of catalogue production like *Extraction*, *Weighting*, *Flagging*, *Background estimation*, *Photometry*. Most of them can be set according to instrument and image characteristics, while few others need a more advanced analysis.

The full configuration file used for this extraction is reported in Appendix A.1. In the next two subsections I present the analysis dedicated to two of those parameters, one for the photometric calibration and a second for the separation between point-like and extended sources.

3.1.1 Photometric calibration

The magnitude of an arbitrary object is defined as:

$$mag = -2.5 \log_{10} F + ZP$$

where F is the “flux” calculated by some tool (in that case *SExtractor*) on the base of the source counts and the image exposure time.

ZP is the so called *Zero-Point*, defined as the magnitude of an object that produces one count per second. Then I had to determine the ZP value for each stacked image to photometrically calibrate our data.

The ZP is estimated from the comparison between source magnitudes measured and the values reported in a reference catalogue with a good photometry. The choice of this reference catalogue has to be done carefully, because source magnitudes depend on the instrument and on the band observed.

To optimize this calibration I decided to extract by myself the reference catalogue from a reference image, already calibrated, of the same field, observed with the same instrument in the same band. So that I could extract catalogues from both reference and our images and compare resulting magnitudes. This approach ensures us homogeneity and avoids magnitudes conversions between different bands.

The deep GOODS-South image observed with VIMOS R -band instrument, published by [Nonino et al. 2009](#), and reported in Figure 3.1, was perfect for my purpose.



Figure 3.1: Nonino stacked image in R -band of VIMOS instrument

It was composed stacking together a set of selected VIMOS observations of GOODS-South in the interval of years 2004 to 2006, obtaining a deep image in R -band up to $R_{lim} \approx 29$ mag (AB, 1σ , $1''$ radius aperture).

The input zero point was provided to *SExtractor* through `MAG_ZEROPOINT` parameter. For the reference image I used the value of $ZP = 27.49$ mag, as

suggested by the paper author.

I filtered the resulting catalogue from not perfectly extracted sources², meaning the one with `FLAGS` $\neq 0$ or `MAG_AUTO` = 99, obtaining 59198 objects.

Then I extracted a catalogue from each of the stacked images using the same parameter set and a dummy value for `MAG_ZEROPOINT`. I made a cross-match ($< 0.5''$) between reference and extracted catalogue calculating the difference in magnitude, that corresponds to the correction I had to apply to the input `MAG_ZEROPOINT`.

To reduce the noise I decided to limit photometric calibration to brighter sources up to 23 mag. Moreover I applied a σ -clipping routine (3σ limit) to exclude outliers and wrong cross-matches.

For each epoch I iteratively corrected the input `MAG_ZEROPOINT` up to reach a negligible median difference in magnitude within the errors.

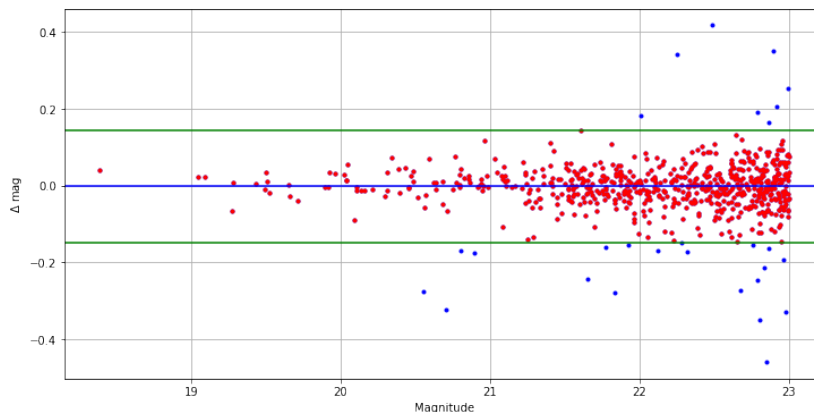


Figure 3.2: Magnitude difference between reference and 2009 catalogues for a `MAG_ZEROPOINT` = 27.20 mag. Green lines are at 3σ from median value

Figure 3.2 shows the difference in magnitude for year 2009 using a `MAG_ZEROPOINT` of 27.20 mag. Green lines indicate the limits of 3σ adopted in σ -clipping procedures that selected only red sources.

The resulting zero-points for each year obtained with this procedure are reported in the Table 3.1, with the number of cross-matched sources and the σ value. The typical error found for `MAG_ZEROPOINT` is $< 5\%$ for all epochs.

²*SExtractor* flags indicate some features found during the extraction that can affect the measure in some way. When the software cannot determine the source magnitude it sets `MAG_AUTO`=99

Year	MAG_ZEROPOINT	# cross-match	σ
2007	27.40 mag	368	0.045
2008	27.30 mag	170	0.049
2009	27.20 mag	614	0.047
2011	27.40 mag	510	0.044
2012	27.51 mag	466	0.065
2013	27.37 mag	485	0.068
2014	27.47 mag	550	0.055
2015	27.45 mag	417	0.055

Table 3.1: Zero-point values with number of cross-matched sources and σ value

3.1.2 Stars and Galaxies separation

To characterize variable objects it should be important to distinguish between point-like and extended sources. Point-like sources have, ideally, negligible extent and their shape is mainly tracked by the image PSF. This geometrical characteristic identify typically stars and very far galaxies that cannot be spatially resolved, like quasars.

Should be useful anticipate that the section below reports a detailed description of analysis I did to tune the parameter dedicated to the stars and galaxies separation. Due to the large depth inhomogeneity of our images and the consequent not complete reliability of the parameter, I decided to avoid selections on my catalogue on the base of it. This will be more extensively explained in the section conclusions.

The **CLASS_STAR (CS)** produced by *SExtractor*, is a “measure” of how much a source is point-like. It is therefore a good option for my purpose. It can assume a value between 0, for very irregular sources, and 1, in case of perfect points. This parameter relies on a multilayer feed-forward neural network trained using supervised learning, to discriminate galaxies, that usually look more extended or fuzzier than stars and QSOs.

When a source is a point, its physical extension on the focal plane is due to the optical PSF, the seeing of the atmosphere and other instrumental effect, as the pixel cross-talk. So that the **CLASS_STAR** classification mainly depends on two input parameters: the **PIXEL_SCALE**, (i.e. the pixel dimensions on the sky) and the **SEEING_FWHM**, an evaluation of the Full Width Half Maximum of stellar images in arcsec.

If the **PIXEL_SCALE** is not provided, it is automatically calculated by *SExtractor* using World Coordinate System (WCS) information in the FITS file.

Instead the `SEEING_FWHM` has to be somehow estimated and provided in the input parameters list.

It is usually possible to measure sources PSF, and therefore its FWHM, using different tools, as for example *Skycat*³.

Otherwise my situation is very complicated, because often images are largely not homogeneous. As previously described, images are produced stacking together all observations available for each year. Therefore the result can show different depths, signal-to-noise ratios (S/N) and sources PSFs along the image. The stacked image of year 2009, in Figure 2.5, is a clear example: different depths zones are here clearly visible.

Therefore it is not possible provide a single PSF FWHM for the entire image.

I decided to study the `CLASS_STAR` classification to optimize the selection for image zones that can provide deepest and most interesting sources.

Tuning of `SEEING_FWHM` input parameter

To better study the `CLASS_STAR` on the base input `SEEING_FWHM` parameter used for the extraction, first I identified the deepest and the shallowest parts of 2009 image.

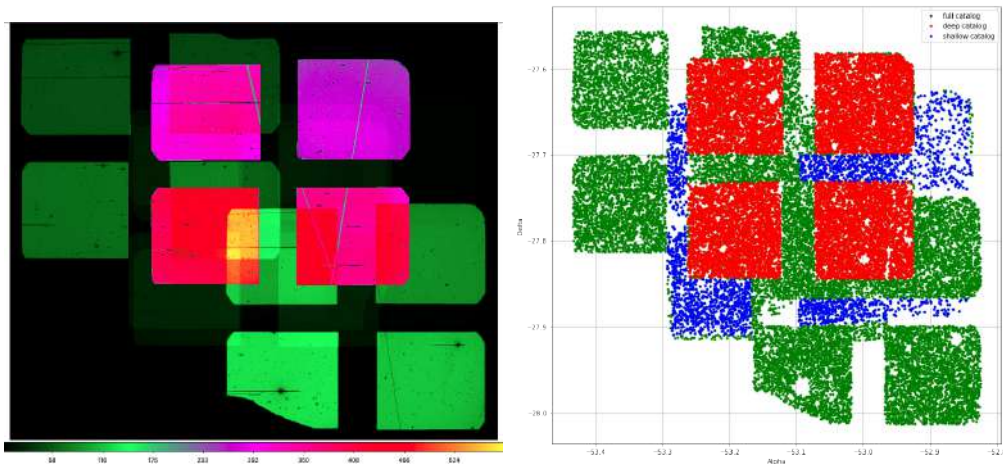


Figure 3.3: *Left image:* Weight map for year 2009. *Right image:* Catalogues of year 2009 divided according image weight. In green the full catalogue, in red deep areas and in blue the most shallow areas

³for more information <http://www.eso.org/sci/observing/tools/skycat.html>

In Figure 3.3 are shown, on the left, the weight map for image of 2009 and, on the right, the catalogue extracted, where sources with different depths are identified. The weight map is the same of the bottom image in Figure 3.3, but in a different color scale to better highlight differences.

Sources extracted from the four “purple” (deep) detectors on left are reported in red on right. Blue sources on the catalogue figure are, instead, from the shallowest ones and their weight is barely visible on the left image. All the other sources are identified in green.

I run *SExtractor* on the 2009 image varying only the `SEEING_FWHM` parameter between $0.5''$ to $2''$, with steps of $0.1''$. The result is a collection of 16 identical catalogues that differ just for `CLASS_STAR` classification.

The selection of point-like sources was made applying a threshold on the `CLASS_STAR` value.

Depending on the purpose of the study, 2 threshold values are typically applied: `CLASS_STAR` > 0.5 for a light selection, that excludes just very irregular sources, and `CLASS_STAR` > 0.9 for a more strict point source identification. For my analysis I decided to test both of them plus the `CLASS_STAR` > 0.7 to track the intermediate behaviour.

In the 3 plots on Figure 3.4 are reported the fraction of sources selected as stars for the 3 `CLASS_STAR` selections. In green are shown fractions for the complete catalogue, in red for the deep and in blue for the shallow.

In case of `CLASS_STAR` > 0.5 selection (the top one on Figure 3.4) all sources of the shallow catalogue are identified as stars for `SEEING_FWHM` greater than about $1.0''$, while complete and deep catalogues have a peak around `SEEING_FWHM` $\sim 1.2''$.

For more strict `CLASS_STAR` selections the ratio presents a peak around a value that increases with depth.

CLASS_STAR selection effects

The shallow part of the catalogue is the less interesting for our purpose. Therefore I decided to limit the analysis on the complete catalogue and its most deep part.

To evaluate the impact of `SEEING_FWHM` on `CLASS_STAR` classification I compared 2 meaningful examples of the 16 catalogues:

- `SEEING_FWHM` = $0.9''$: value measured on the image using the SkyCat tool. It corresponds also to the peak value of shallow catalogue

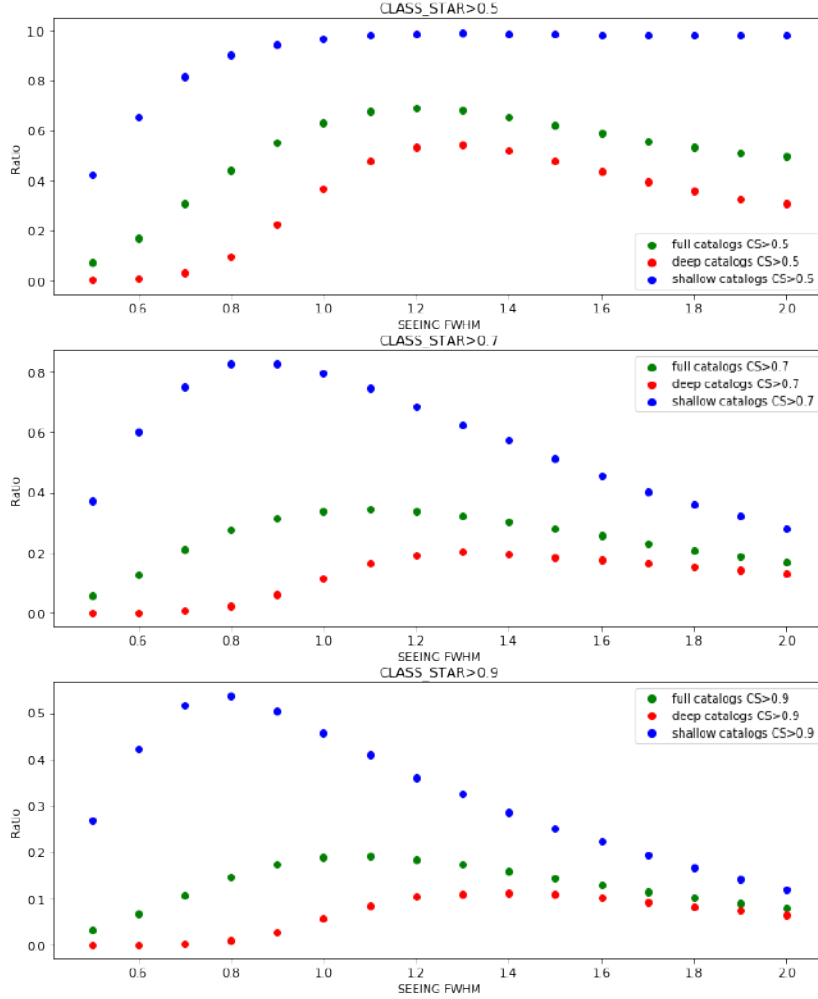


Figure 3.4: Ratio of sources selected as stars with three thresholds ($\text{CLASS_STAR} > 0.5, 0.7$ and 0.9) for full, deep and faint catalogues

- $\text{SEEING_FWHM} = 1.3''$: peak value for deep catalogue

Figure 3.5 shows the fraction of sources identified as stars with the strict selection $\text{CLASS_STAR} > 0.9$ for complete and deep values. Fraction values for $\text{SEEING_FWHM} = 0.9''$ are encircled in gray, while the blue corresponds to a $\text{SEEING_FWHM} = 1.3''$.

Limiting the analysis to the deep area, a value of $\text{SEEING_FWHM} = 0.9''$ identifies as stars only a small fraction of the sources ($\sim 2.5\%$). Using, instead, a value of $\text{SEEING_FWHM} = 1.3''$ about 10% of sources have a $\text{CLASS_STAR} > 0.9$. This is a reasonable value assuming an homogeneous distribution.

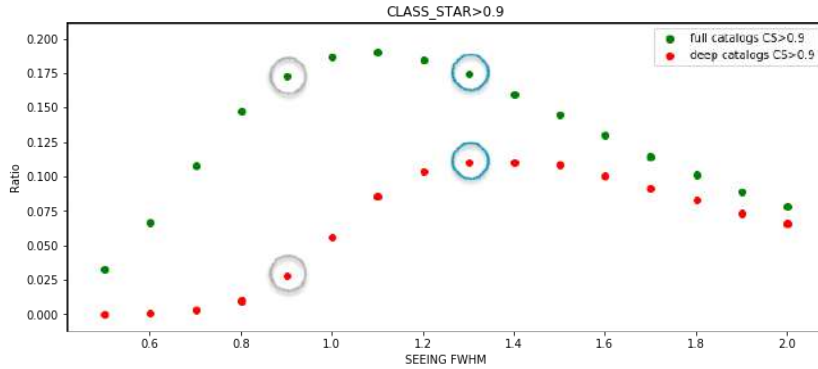


Figure 3.5: Ratio of sources selected as stars ($\text{CLASS_STAR} > 0.9$) for full and deep catalogue. In blue are underlined values for $\text{SEEING_FWHM} = 1.3''$ and in gray a reference value of $\text{SEEING_FWHM} = 0.9''$

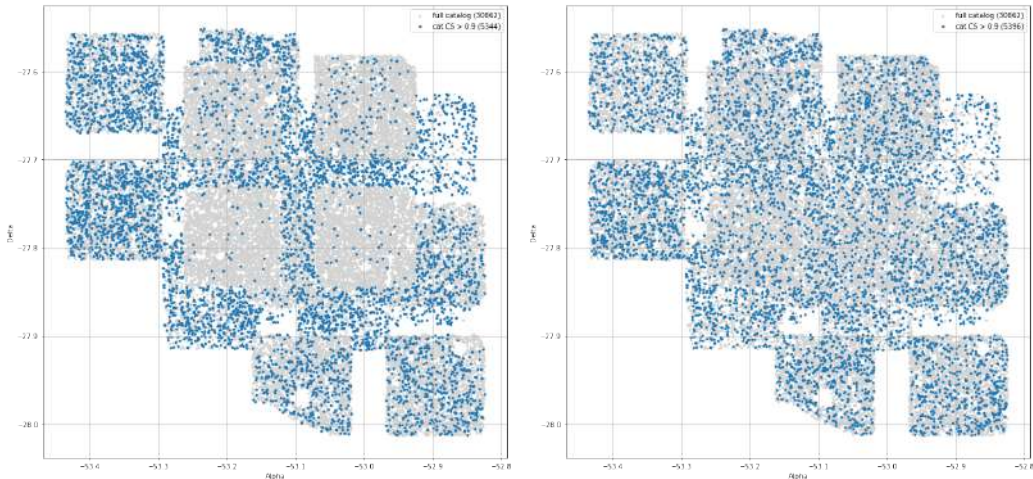


Figure 3.6: In blue are represented sources identified as stars ($\text{CLASS_STAR} > 0.9$) over the full catalogue (in gray). On the *left image* for $\text{SEEING_FWHM} = 0.9''$ while on the *right* the one with $\text{SEEING_FWHM} = 1.3''$

The difference between the two choices is clearly visible in Figure 3.6 where catalogues on the sky are shown. Sources identified as stars are reported in blue, over the full catalogue in gray.

On the left the value of $\text{SEEING_FWHM} = 0.9''$ was used during the extraction and on the right $\text{SEEING_FWHM} = 1.3''$. In the first case large part of the “stars” are in the most shallow section for the image, while only few sources are identified as stars in the deep region. A value of $\text{SEEING_FWHM} = 1.3''$ identify stars more homogeneously on the entire image. This result seems more reasonable.

Advertisements and conclusions on CLASS_STAR selection

The analysis here presented highlighted a great depth inhomogeneity inside images and the consequent large range of `SEEING_FWHM` on different image areas.

I decided to adopt a value of `SEEING_FWHM = 1.3''` for the catalogue extraction, because it is the best value for the deeper area and identifies stars more homogeneously on the entire image, as shown in Figure 3.6. I underline that the `SEEING_FWHM` parameter affects only the `CLASS_STAR` value; moreover this classification is not considered completely reliable by the software author himself and has been superseded by the `SPREAD_MODEL` estimator in more recent *SExtractor* versions. Unfortunately this last estimator was not available in software version I used.

Consequently I decided to do not separate point-like sources from extended ones and maintain a single catalogue. The `CLASS_STAR` value associated to each source has to be considered “a realistic indicator” just for sources that belong to the deep image area.

3.2 Creation of the catalogue for the analysis

At this point I could extract catalogues from the 8 photometrically calibrated stacked images.

Year	Effective Area (deg ²)	Magnitude limit	# sources	# match 2009
2007	0.1048	25.9	15697	9449
2008	0.0571	25.4	8641	5341
2009	0.1811	26.0	30862	-
2011	0.1058	25.7	13266	8074
2012	0.0871	26.0	11581	7811
2013	0.1037	26.1	21073	10420
2014	0.1258	26.6	22398	10797
2015	0.0866	26.4	16365	8684

Table 3.2: List of catalogues with effective area covered in deg², catalogue magnitude limit, number of sources detected and number of cross match with the 2009 catalogue

On Table 3.2 are reported some useful information about images and the corresponding catalogues. As *Effective Area* is intended the image extension calculated considering contribution only from pixel with weight $\neq 0$. The

Magnitude limit is indeed calculated from the catalogue magnitude distributions in Figure 3.8, as explained in the next section. This is related with images depth.

The third column is the number of sources detected in each image, that is determined by the combination of extension and depth. So that as example the 2015, despite it covers the same effective area of 2012, it has $\sim 40\%$ more sources because it is more deep.

The most extended and populated catalogue is the 2009, that I choose as reference. I made a cross-match between 2009 and the other catalogues within $0.5''$, 2.5 times the pixel dimension that corresponds to $\sim 1/2$ of the typical measured images seeing. The 2009 is not the deepest catalogue available, but it is the best choice to be used as reference thanks to its large effective area covered that ensured the largest overlap with the other catalogues.

The last column on Table 3.2 reports the number of sources cross-matched for each epoch. The small fraction of sources in common among catalogues, typically slightly more then half, is due to the fact that pointings are often not well aligned.

To evaluate the correct sources associations I visualized the difference of magnitudes between 2009 and the cross-matched year.

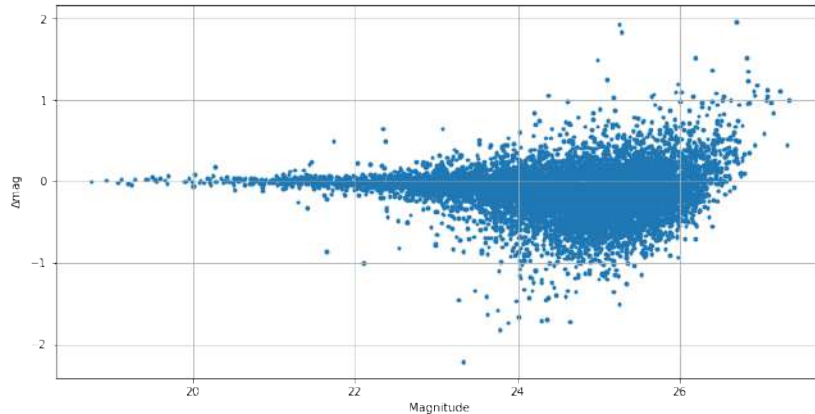


Figure 3.7: Sources magnitude difference between 2009 and 2011 catalogues

In Figure 3.7 the example of the year 2011 is shown. 8074 sources had a good association with the reference catalogue, $\sim 60\%$ of 2011 catalogue. Residuals are close to zero for brightest sources and their dispersion increases for weaker sources, as expected. In any case they are distributed around zero,

to confirm the correct relative photometric calibration.

The **merged catalogue** is composed by 30862 sources. For each source detected in 2009 from 1 to 7 detections in the other epochs were found.

Because of the idea is to compare sources magnitudes in different epochs should be useful discuss the choice I did for the source magnitude estimation. I decided to adopt the *SExtractor* **MAG_BEST** value, that is usually equal to **MAG_AUTO** photometry, but if the contribution of other sources exceeds the 10% it is **MAG_ISOCORR**. In any case I filtered detections with **FLAGS =0**, therefore I excluded all blended sources.

I'm conscious that the seeing variation along the image can influence the source shape estimated, but **MAG_BEST** is the best compromise to characterize both extended and point-like sources.

More sophisticated analysis can be pursued using fluxes measured with different apertures (**MAG_APER**) or using a template/PSF fitting photometry. Those procedures need a good separation between point-like and extended sources that have to be treated independently. Unfortunately, as previously explained, this separation is not possible up to now.

My aim is to postpone those tests to future analysis.

3.2.1 Catalogue instrumental magnitude limit

An interesting parameter to be evaluated is the catalogue **instrumental magnitudes limit**. This is the magnitude that the instrument can reach, in a specific configuration, without loosing in completeness. This is a limit due to the instrument and to the image depth. I can consider my catalogue “*complete*” up to this limit, while, for greater magnitudes, it is affected by an instrumental selection.

Usually the instrumental magnitude limit is considered the peak of a catalogue magnitude distribution.

I evaluated this value for all years from the distribution visualized in Figure 3.8, where a red dashed line indicates the magnitude limit for the 2009 at ~ 26 mag. Instrumental magnitudes limits for all catalogues are reported in the third column of Table 3.2.

The magnitude distribution for the **merged catalogue** is visualized in Figure 3.9, where the instrumental magnitude limit is the red dashed line.

The number of sources grows from bright to faint sources until they reach the magnitude limit of ~ 26 mag. For fainter sources the catalogue loses its completeness: there is, indeed, a step decrease in the number of sources

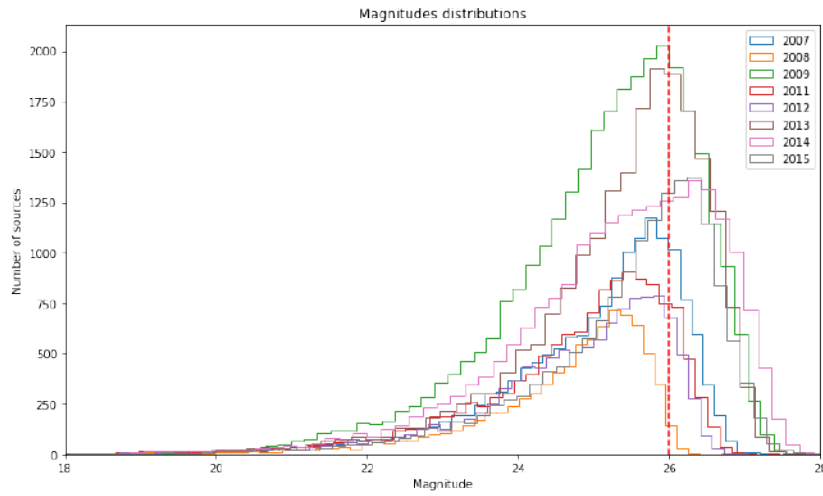


Figure 3.8: Magnitude distributions for the 8 catalogues. The red dashed line indicates the instrumental magnitude limit for year 2009 used as reference.

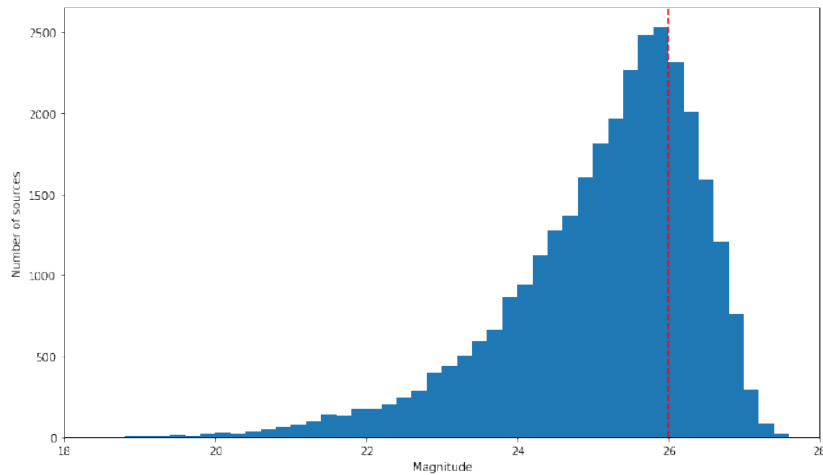


Figure 3.9: Merged catalogue source magnitude distribution and instrumental limit at 26 mag

detected due to the instrumental limit.

I did not filter the merged catalogue for this value, because I was interested also in studying fainter sources. Sources with magnitudes > 26 mag have to be treated carefully, because their variability identification should be affected by a catalogue selection effect.

This value will therefore be reported on graphs to identify the complete sample.

3.2.2 Selection for number of detections

The purpose was to analyse the variation of sources magnitude. Therefore I needed to select only sources with a sufficient **number of detections (ND)** to be statistically valid.

The ND distribution is visualized in Figure 3.10. Stacked images have a slightly different field coverage, due to the different coordinates of observation pointings. So that the corresponding catalogues does not overlap perfectly, this is the reason why large part of sources has few detections.

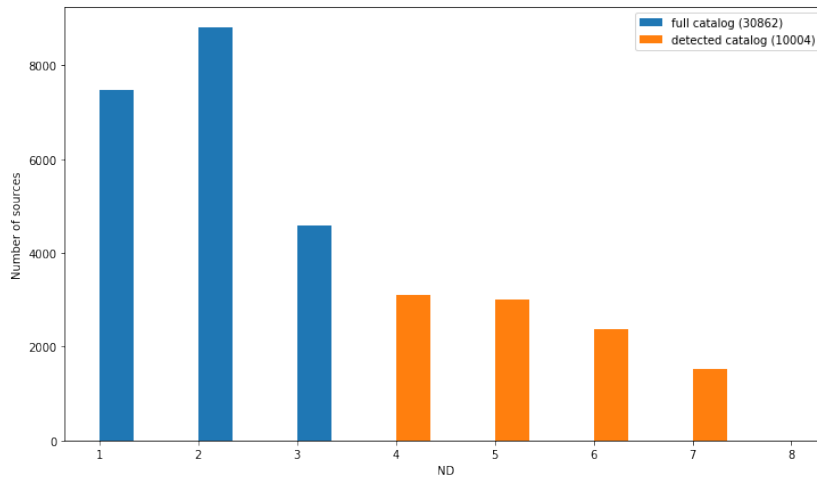


Figure 3.10: Histogram with the number of detections.

I decided to select sources with at least 4 detections, $ND \geq 4$, getting a selection of 10004 objects, about 1/3 of the original **merged catalogue**. They are visualized in orange on the Figure 3.10. I'd like to underline that I found ~ 2000 sources with 7 detections.

In Figure 3.11 are plotted in blue sky position of selected sources and in gray the entire **merged catalogue**. Sources with $ND \geq 4$ are in the central part of the **merged catalogue**, where all stacked images overlap.

The resulting catalogue, that I'll call **detected catalogue**, is composed by 10004 sources with $ND \geq 4$. This sample contains sources with enough number of detections for each source to be subject to a valid statistical analysis. I used this catalogue to test my method for variability identification.

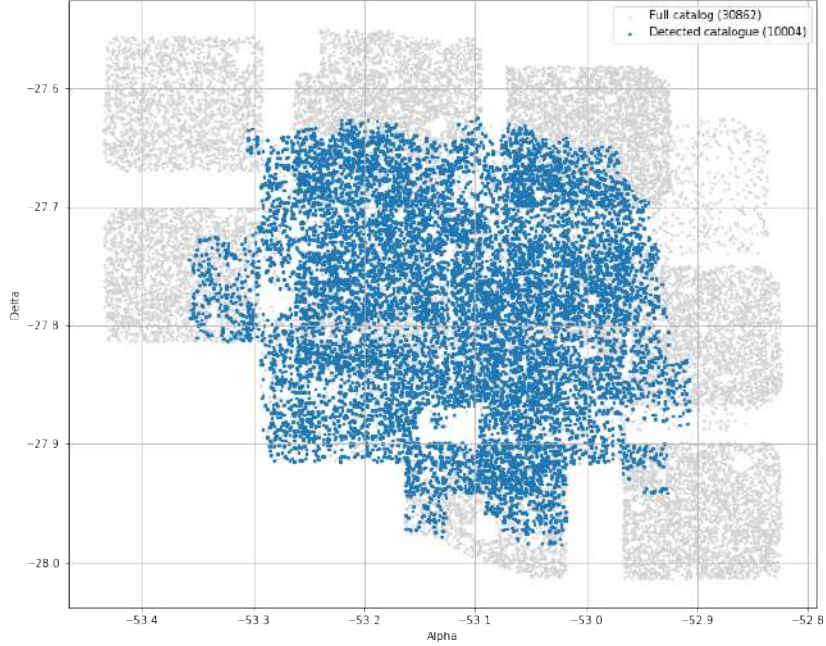


Figure 3.11: Sources with $ND \geq 4$ in blue over the full catalogue in gray

3.3 MAD calculation and variables selection

With the results of the analysis carried out so far I could use the **detected catalogue** for variability research.

Following the technique proposed by [Pouliasis et al. 2019](#), I decided to use the magnitude **Median Absolute Deviation (MAD)**, as variability estimator. The MAD may be used similarly to how one would use the deviation for the average. It is, indeed, a more robust parameter than the standard deviation (σ) for outliers.

For a univariate data set X_1, X_2, \dots, X_n , the MAD is defined as the median of the absolute deviations from the data's median:

$$MAD = median(|X_i - median(X)|) \quad (3.1)$$

In order to use the MAD as a consistent estimate with the standard deviation σ , one takes:

$$\sigma = k \cdot MAD$$

where the scale factor k depends on the distribution. In case of normally

distributed data k takes the value of:

$$k = \frac{1}{\Phi^{-1}(3/4)} \simeq 1.4826$$

where Φ^{-1} indicates the inverse of cumulative distribution function.

I calculated the sources magnitude MADs, as estimator of variability, and the median source magnitude value \overline{mag} , that is the median value among source magnitudes in different epochs.

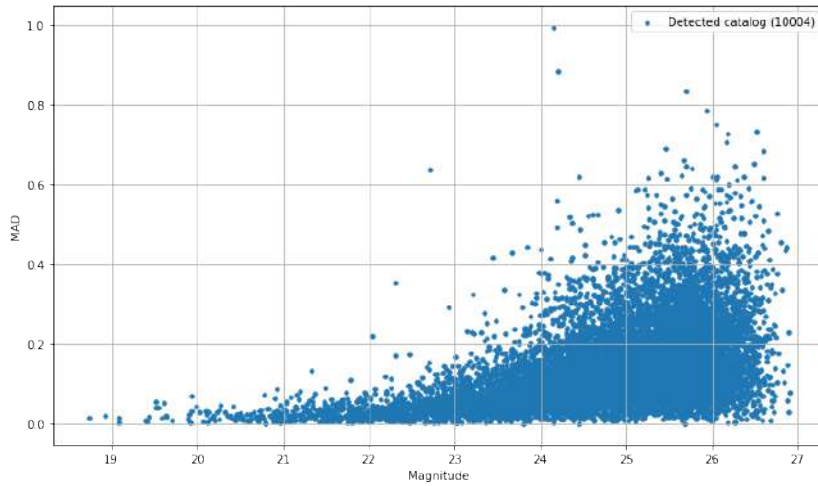


Figure 3.12: The magnitude Median Absolute Distribution (MAD) of catalogue sources as function of sources magnitude. The MAD spread increases for weak values.

MAD values as a function of median source magnitudes \overline{mag} are visualized in Figure 3.12. As expected the MAD spread increases for weaker magnitudes because of statistical noise.

The purpose was to identify sources with outlier MAD values, that are evidently detached from the distribution on Figure 3.12. These sources show a MAD value clearly larger than the median MAD values of sources with similar magnitudes.

The magnitude MADs I calculated are measures of variability and therefore these sources are “*more variable*” than other similar sources. Their variation of magnitude manifests an intrinsic source variability.

3.4 MAD* and the binning method

To analytically identify the most variable sources I had to compare their MAD values with the median MAD value of the magnitude range.

The MAD*, the MAD normalized to the median MAD value of the corresponding magnitude range, or *bin*, can be used with this purpose.

The MAD_i^* of the source i in the bin magnitude b is defined as:

$$MAD_i^* = \frac{MAD_i - (\overline{MAD})_b}{\sigma_{MAD_b}} \quad (3.2)$$

where

- MAD_i is the magnitude MAD of source i
- $(\overline{MAD})_b$ is the median value of MAD for the bin b
- σ_{MAD_b} is the bin MAD standard deviation

The MAD_i^* is the MAD spread measured in unit of σ_{MAD_b} . Therefore it is normalized for the increasing dispersion observed in the evolution with magnitude.

Following the example of [Pouliasis et al. 2019](#), I decided to identify as variables sources with a $MAD^* > 3.5$. They are sources that show a variability $3.5 \sigma_{MAD_b}$ times larger than the $(\overline{MAD})_b$, the median MAD value for the bin.

The value of MAD_i^* depends on the strategy adopted for the definition of magnitude *bins* on the distribution.

The bin strategy was thought to compare sources with similar magnitudes, i.e. the *bin* should not be very large. But, at the same time, the *bin* needs to contain enough number of sources to be statistically valid.

I studied two different binning approaches:

1. fix the **number of sources for each bin** (`obj_num`) to preserve the statistical validity of bin MAD* values
2. fix the **magnitude step** (`mag_step`) to collect together sources within the same magnitude range

Both of those methods have a free parameter, `obj_num` and `mag_step`, that have to be tuned.

I tested some values to study the effects on identification of variables. For each bin I calculated the median MAD value, $(\overline{MAD})_b$, and the median magnitude value of the bin range.

The evaluation of different binning strategies has to consider two parameters: which is the range of magnitude covered by each bin and how many sources are contained. The MAD^* , that I used for variability selection, depends on the $(\overline{MAD})_b$, that is another parameter that has to be considered for a quality estimate of the binning strategy.

I'd like to clearly select sources with an high MAD value along the entire magnitude catalogue range showed in Figure 3.12.

To evaluate the quality of a binning strategy I visualized those information with 2 graphs collected in a single figure (i.e. Figure 3.13):

- on the *top*: the histogram of the number of sources for each bin
- on the *bottom*: the MAD distribution. Green lines are the MAD median value for the bin $(\overline{MAD})_b$. Sources identified as variables, i.e. with a $MAD^* > 3.5$, are orange and blue.

A vertical red dashed line identifies the instrumental magnitude limit and separate variables of “complete sample” in orange, from the ones out of this limit in blue. I used these figures to identify strengths and fragilities of the binning strategy adopted.

3.4.1 `obj_num`: variable bin with fixed number of sources

As first method I created bins containing the same number of objects. In this way I ensured the presence of an enough number of sources to detect variability.

I sorted the catalogue for magnitude and collected together a `obj_num` number of objects. I have tested different `obj_num` values between 50 and 150. I present in this section the results for the two extreme values.

The smaller value tested was `obj_num`= 50 , i.e. 50 sources for *bin*, that identifies 69 variable sources. The corresponding graphs are in Figure 3.13.

The histogram on top shows how the number of *bin* highly increases for larger magnitudes up to make not more possible to graphically resolve single bars for $mag > 23$. It is an effect of this *bin* selection method applied on the catalogue magnitude distribution (Figure 3.9).

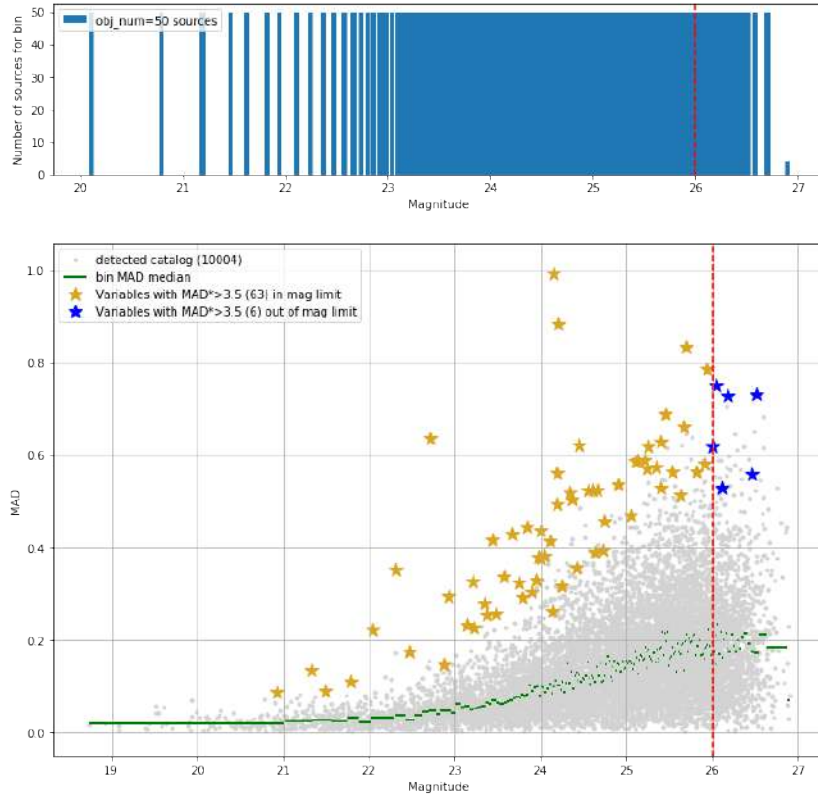


Figure 3.13: Number of sources and MAD distribution for `obj_num= 50`

The *bin* magnitude size, that is the length of green lines on the bottom plot, decreases for weaker sources. The $(\overline{MAD})_b$, the ordinate value of green line, shows instead a large scatter for $\text{mag} > 24$. This scatter is due to the small *bin* size and the consequent inclusion of fraction of MAD distribution noise.

Considering that the MAD_i^* is calculated comparing the source MAD with the $(\overline{MAD})_b$, this fluctuation affects the identification of variable sources. Indeed, because of the scatter, a source can be identified as variable or not according to whether it's included in a *bin* or in the next one. It makes this method highly inefficient for $\text{mag} > 24$, where variable (gold and blue points) and not variable (gray points) sources are clearly mixed.

Increasing the number of objects for each *bin* up to `obj_num=150`, 77 sources are identified as variables.

The result, visualized in Figure 3.14, shows a $(\overline{MAD})_b$ more stable and continue, even if it still has some scatter. The mismatching variable sources identification is not completely resolved, but it's not so extreme as in the

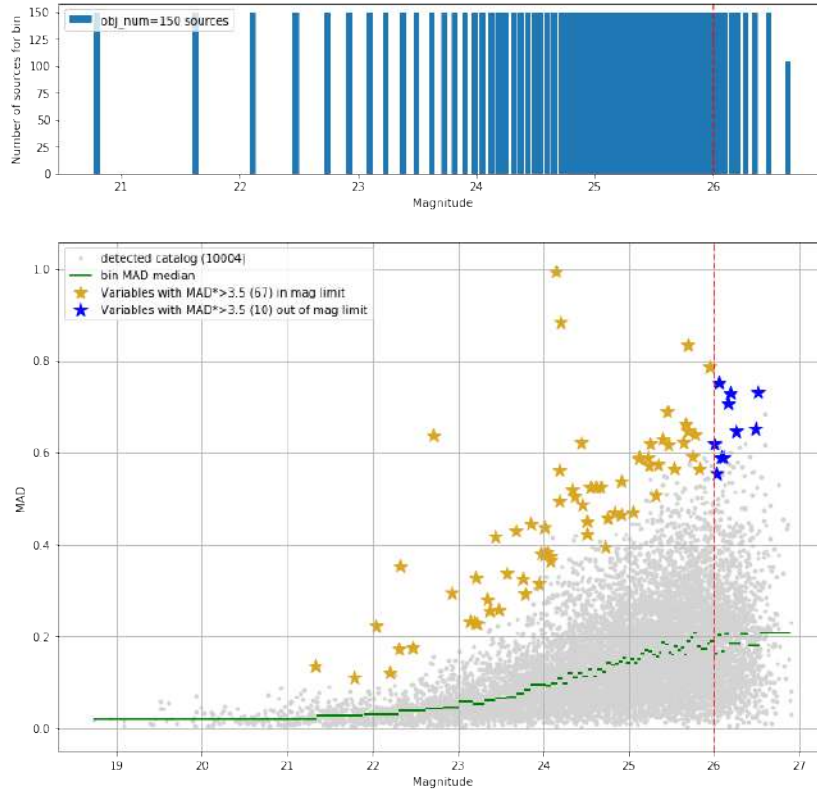


Figure 3.14: Number of sources and MAD distribution for `obj_num = 150`

previous case.

This large number of objects per *bin* introduces a problem in the magnitude range of brighter sources. The first *bin* covers 3 magnitudes, from 18.5 up to 21.5 mag, corresponding to a factor 10 in flux, a too wide value.

3.4.2 `mag_step`: fixed bin and variable source selected number

The other binning strategy is to collect sources in the same magnitude range. In that way it is avoided the problem raised in the previous selection, ensuring the comparison in a reasonable range of magnitudes.

The size of this step, `mag_step`, is the parameter that I can vary. I have tested different `mag_step` values between 0.3 and 0.5. I present in this section the results for the two extreme values.

With a value of `mag_step = 0.3 mag`, 80 sources are identified as variables. In this approach the number of sources of each *bin*, reported in the his-

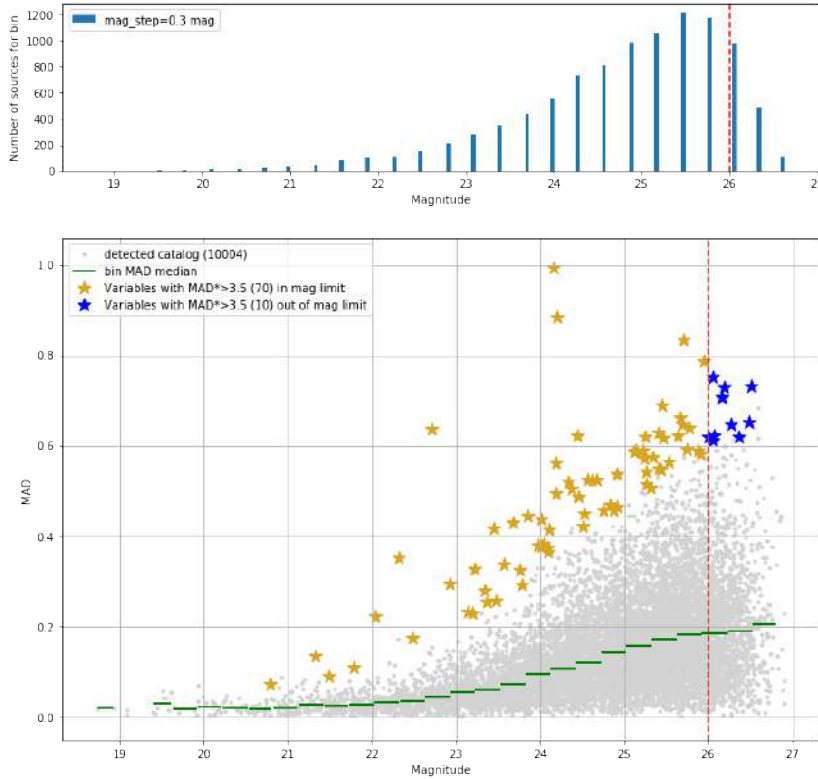


Figure 3.15: Number of sources and MAD distribution for `mag_step` = 0.3 mag

togram on top of Figure 3.15, varies a lot from few sources for the first *bin* up to ~ 1200 sources around 25.5 mag.

The advantage of this binning method is the stabilization of $(\overline{MAD})_b$, the green line, in a more continuous trend that increases with depth. It ensures a better separation between variable and not variable sources.

The main problem here is related to the poor number of sources for bright magnitude *bins*. While for the fainter part of the distribution there are enough sources to identify variability, the first *bins* collected just few sources, even less than 10, not enough for a statistical evaluation. The green line shows fluctuations in the bright part that can affect the variability identification.

Increasing the size of the magnitude step up to `mag_step` = 0.5 mag, 71 sources are identified as variables. The result is showed in Figure 3.16.

Each *bin* contains more sources and the green line, i.e. the $(\overline{MAD})_b$, is more stable also in the bright part of distribution.

The enlargement of magnitude range to this size increases also the sep-

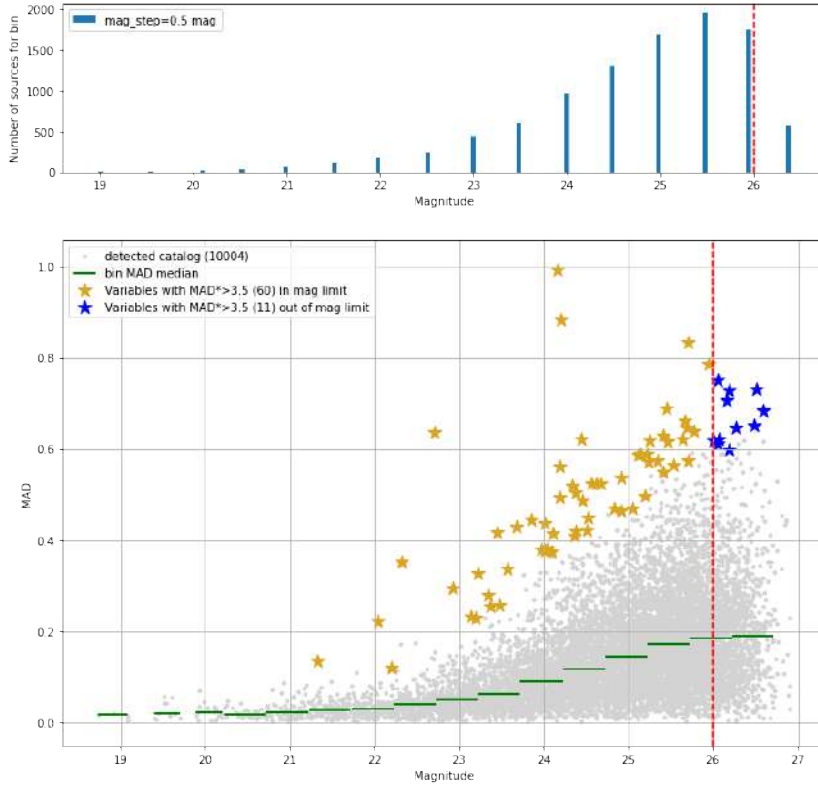


Figure 3.16: Number of sources and MAD distribution for $mag_step = 0.5$ mag

variation of median MAD value $(\overline{MAD})_b$ between *bins* in the weaker part of distribution. Because of the green line is the base value for the MAD^* calculation, its steps create an inhomogeneous identification of variable sources.

3.4.3 obj_num and mag_step: a mixed method

To minimize fragilities of the previous methods I created a binning strategy mixing both approaches: each bin has to contain at least an `obj_num` number of objects and should cover a minimum range of `mag_step` magnitudes.

The purpose is to ensure a stable value of $(\overline{MAD})_b$ over the entire magnitude distribution.

I tested different sets of parameters to optimize the selection. The purpose was to obtain a $(\overline{MAD})_b$ as continuous as possible, avoiding the scatter for bright bins and the step shape for the fainter ones. Based on visual inspection the best combination I found, reported in Figure 3.17, is `obj_num = 50` sources

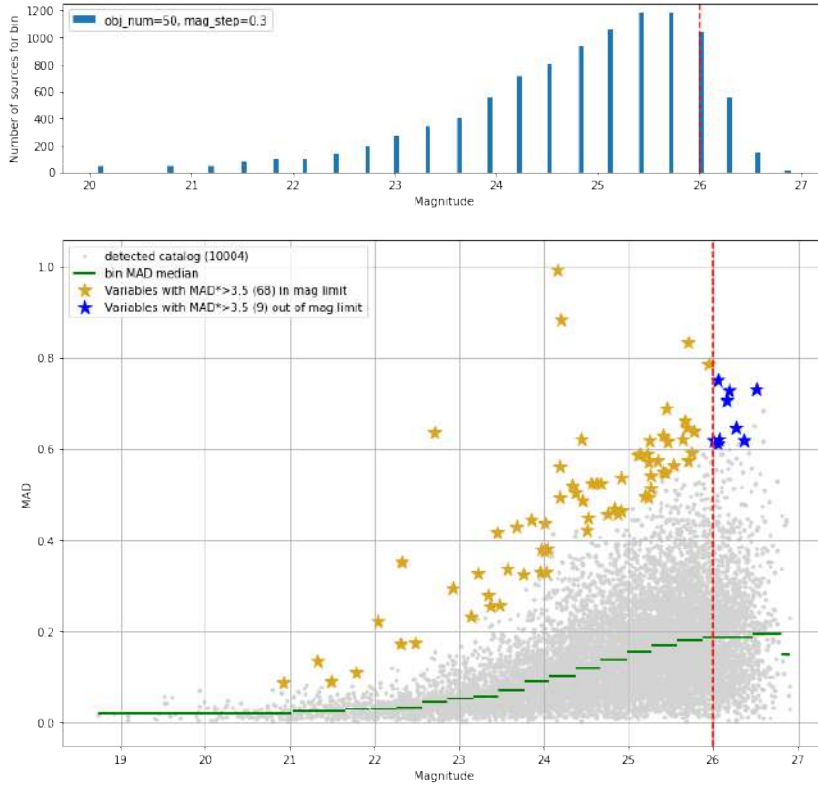


Figure 3.17: Number of sources and MAD distribution for `obj_num = 50` and `mag_step = 0.3 mag`

per bin and `mag_step = 0.3 mag` minimum.

The brightest steps of the green line, the $(\overline{MAD})_b$, are larger than 0.3 mag because of the `obj_num` limit imposed. So that the trend is more stable and does not show eliminated fluctuations. In the deeper part, instead, *bin* sizes are determined by the `mag_step` limit and collect up to ~ 1200 objects. Even in that part the trend does not show fluctuations and the variable identification is more reliable.

I detected 68 variables in the complete catalogue section, i.e. within the instrumental limit of 26 mag, and other 9 beyond this limit. The brightest variable found has a magnitude slightly < 21 mag. The distribution of variables identified is compliant with the catalogue magnitude distribution, with a higher number of variables found in the range from 25 mag to 26 mag.

3.5 Catalogue of variables

The catalogue of variables identified by the last method is the best solution found up to now.

The variables selection is still not perfect: there are some sources not identified as variable between 24 and 25 mag due to the step shape of $(\overline{MAD})_b$ line. But, with this kind of method, this is the most reasonable selection.

Some variable identifications can depend on the binning method adopted, as illustrated in the previous section. Despite this, a very large fraction of the variables in the catalogue were identified by all the methods and parameters adopted for binning.

Therefore I decided to use it as proposed detected catalogue of variable.

A more refined approach, based on a parametrization of $(\overline{MAD})_b$ line with a power law function and therefore an analytic selection, is currently under study. It introduces a dependency of fitted function from the bin selection adopted that has to be analysed.

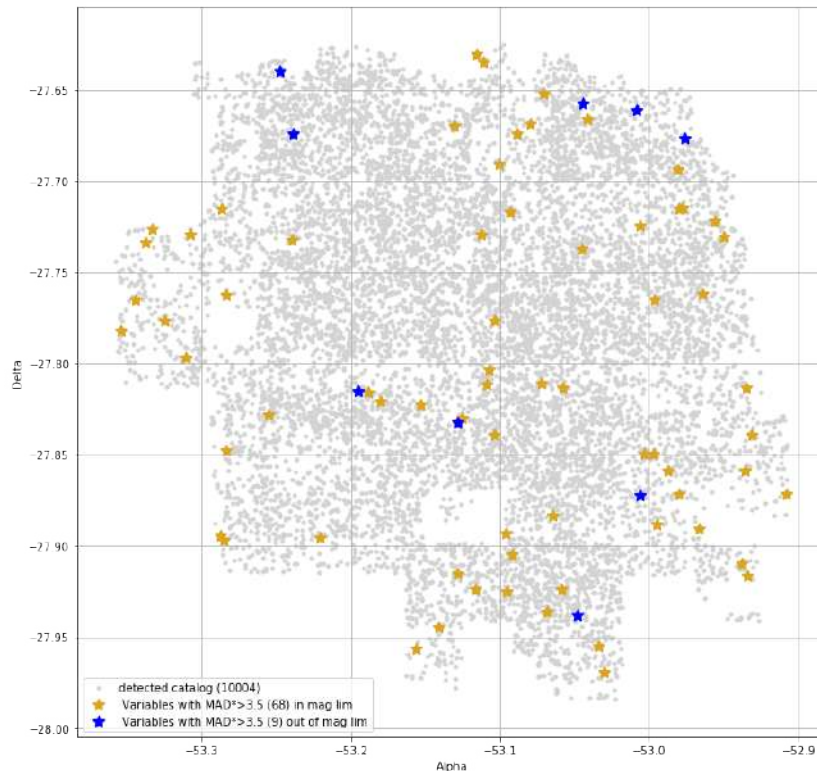


Figure 3.18: Variables on sky

The variables sources position on the sky are reported on Figure 3.18, over the entire detected catalogue. In gold variables with magnitude up to the instrumental limit and in blue the deeper ones.

The final variables catalogue here selected is presented in the Appendix B at Table B.1 sorted in order of magnitudes.

Chapter 4

Result analysis

The purpose of this work is to select AGNs in Chandra Deep Field South (CDF-S) region through their variability. I used available observations of CDF-S from VIMOS/VLT instrument in R-band to create a set of images, one per each year, with a time coverage from 2008 to 2015.

I extracted a catalogue from each image and compared them to create the *detected catalogue* that contains source magnitudes measured in different years. A detailed explanation of the image analysis, data reduction and catalogue production has been provided in Chapter 2.

Then I implemented a method to identify sources that show high variability, using the statistical parameter Median Absolute Deviation (MAD) as indicator, as explained in Chapter 3.

The resulting variability selected catalogue is composed by 77 sources. The analysis of this sample is described here.

First, in Section 4.1, I examined light-curves and images cut-outs of each single sources to identify features not revealed and evaluate reliability of each measure. This process allowed me to clean the sample discarding spurious identifications.

Then I compared the catalogue with variability-selected AGNs from literature to evaluate the completeness (Section 4.2), with X-Ray catalogues (Section 4.3) and finally with some multi-wavelength catalogues (Section 4.4) to derive a redshift value for sources.

The source characterization resulting from those cross-match is analysed in Section 4.5 and some interesting results are here reported. Some example sources, selected by this work, are individually described in Section 4.6.

4.1 Visual inspection

With the method previously described I identified a sample of 77 variable candidate AGNs. The selection is based on the comparison of the source magnitudes measured in different years, and, up to this moment I relied on *SExtractor*¹ capability to evaluate background, extract fluxes and separate blended sources correctly.

At this point I needed to visually check each source extraction to verify the presence of features on images not identified by *SExtractor*. With this aim I extracted a cut-out of 45×45 pixel around sources coordinates. Then I created a Multiple Extensions FITS (MEF) file for each variable collecting cut-outs together.

From those files I could easily visualize the source light-curve and compare each measured magnitude, and its error, with images. To optimize this comparison process I also created a Graphic Interchange Format (GIF) file for each source. In each frame is highlighted the magnitude values on the light-curve and the corresponding cut-out is visualized on the right.

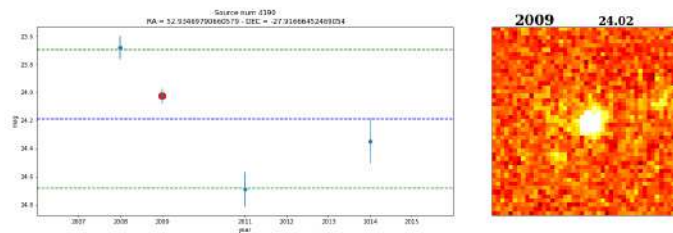


Figure 4.1: Second frame of GIF file for variable 4190. The red dot highlights the magnitude measure (2009 in that case) on the light-curve and the corresponding cut-out is visualized on the right, with the year and the magnitude value on top.

A single GIF frame for source 4190 is shown in Figure 4.1 as example.

This tool was very useful to find 4 sources with measures affected by defects. I discarded those not reliable detections from final catalogue.

Sources 9581 and 10839 have their lowest magnitude measures affected by features not correctly flagged in the original images. In case of 14495 and 18784 are clearly visible 2 blended sources, not separated by *SExtractor*.

¹*SExtractor* is the software used for the catalogue extraction. <https://www.astromatic.net/software/sextractor>

From light-curves I could identify a SN candidate. The faint source 8680 shows a very fast increase of 3 magnitudes between the year 2013 and 2014, as visible in Figure 4.2.

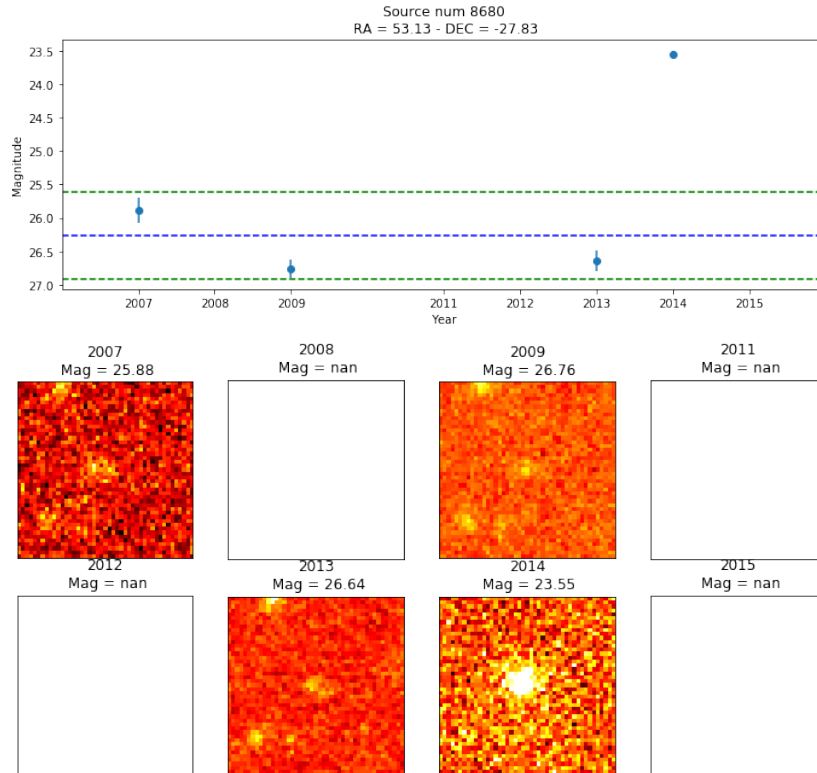


Figure 4.2: Light-curve and cut-outs for source 8680 that I proposed as SN candidate

Its light-curve shows a rise in luminosity from 26.63 mag in 2013 to 23.55 mag in the 2014.

This event should be alternatively interpreted as a Tidal Distruption event (TDE): a quiescent black hole that is suddenly activated by the fall of matter. To understand better the nature of this transient it is necessary to analyse next epochs observations.

Unfortunately the source was not detected in 2015: such a measure is important to check the presence of the rapid decrease expected for SN. Looking at the 2015 image, on 8680 coordinate it's visible a very faint source, not detected at all by *SExtractor*. It seems to support the SN thesis, because the TDE has usually a longer dimming time.

I maintained the 8680 source in the catalogue as SN candidate. The analysis of the epochs after 2015 is out of the purpose of this thesis and will be subject of future analysis.

4.2 Comparison with variability-selected AGNs from literature

I selected a final catalogue of 73 sources, discarding the 4 spurious identifications previously described.

As first I did a cross-match with variability-selected AGNs from literature. The purpose of this comparison is the evaluation of **completeness** of the selected catalogue: I'd like to understand how many known variable AGNs, included in my VIMOS R-band **detected catalogue**, have been identified by the selection method implemented.

4.2.1 Reference Catalogues

The Chandra Deep Field South (CDF-S) is one of the most observed fields on the sky. Deep pointings and repeated observations were dedicated to this field to investigate on its population. The large variety of instruments adopted (Chandra, XMM, GALEX, HST, Spitzer, Hershel, VLA), and the consequent wide range of wavelength coverage, give the opportunity to characterize sources and compare results.

Thanks to the availability of a large number of observations in the last decades several works were dedicated to the research of variable objects, including AGNs.

With the purpose of validating the method I implemented, I used some catalogues of variability-selected AGNs in CDF-S as reference to compare my results.

I choose catalogues proposed by four independent papers to explore a variety of methods and instruments.

Two of them, [Trevese et al. 2008](#) and [Falocco et al. 2015](#), used dedicated observations from wide field instruments from ground, while both [Villforth, Koekemoer, and Grogin 2010](#) and [Pouliasis et al. 2019](#) used HST/ACS observations, but on different time-scales. Moreover several statistical methods for the detection of variability were explored in these works.

This bundle of catalogues provides the possibility to evaluate my selection method and identify which kind of sources it is able to detect.

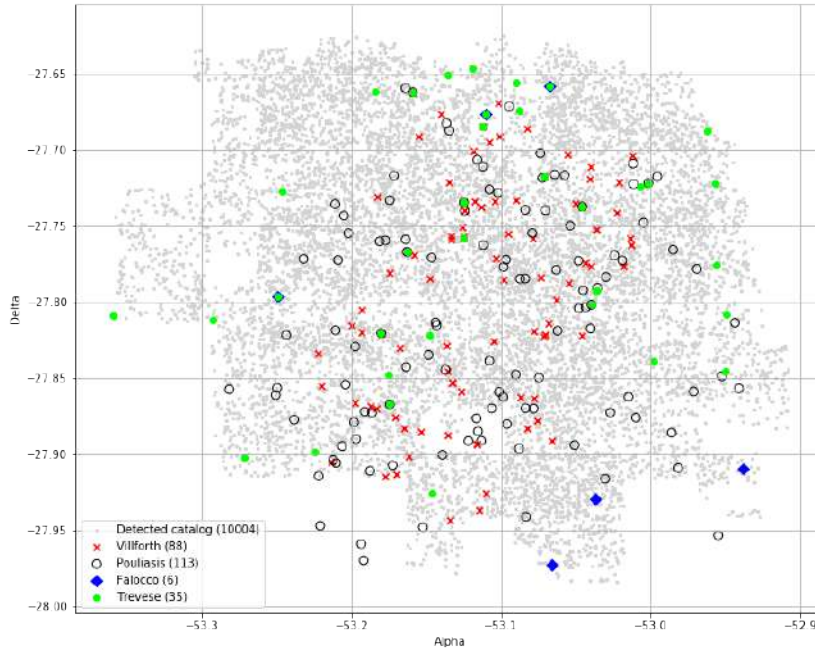


Figure 4.3: The four reference variability-selected AGNs catalogues on the sky over the complete object catalogue produced in this work (in gray)

Next I'll briefly present these four reference catalogues to introduce their data and statistical method adopted.

Trevese et al. (2008)

The catalogue proposed in [Trevese et al. 2008](#) is the oldest I used for comparison. Authors analysed images from the framework of the Southern intermediate Redshift ESO Supernova Search (STRESS), a long term project designed to measure the evolution in time of the rate of all types of Supernova (SN) events.

The Supernovae search is based on comparison of images of selected sky field obtained at different epoch. Those images were acquired with the 2.2 m ESO/MPI telescope at ESO, La Silla (Chile), equipped with the Wide Field Imager (WFI), using deep V band observations.

For the purpose of SNe research AGNs were considered as contaminant and therefore were excluded. The authors of [Trevese et al. 2008](#) were, instead, interested in those objects and optimized the catalogue extraction for AGN detections down to $V \sim 24$ mag.

They selected 8 images of field named AXAF (after the name of X-ray

AXAF satellite subsequently called Chandra) in date range from November 1999 up to December 2001. Aperture photometry was performed for each object on all images at the same (α, δ) position for various apertures. They used a normalised rms variability $\sigma^* > 3$ to identify variable candidates and characterized those sources with X-ray counterparts. On the base of X-ray to Optical ratio (X/O) they could separate AGN from starburst galaxies.

The catalogue proposed is composed by 132 variability-selected AGN, with the available X-ray fluxes (or at least an upper limit value) and, possibly, a classification based on literature.

A spectroscopic follow-up of this catalogue was described in the paper [Boutsia et al. 2009](#). They could break down the sample in AGNs, starburst galaxies and low-ionization narrow-line emission objects and provide a spectroscopic redshift. A large fraction of the observed candidates were identified as Broad Line Active Galactic Nucleus (BLAGN), confirming the efficiency of variability in detecting quasars.

The STRESS sample covers a larger area than the VIMOS observations, as appears from the green points on Figure 4.3. For my comparison I used only the 35 sources that overlap with the **detected catalogue** and that were spectroscopically confirmed as AGN by the follow up.

Villforth et al. (2010)

A second very interesting work on this topic was presented in [Villforth, Koekemoer, and Grogin 2010](#) with a corresponding spectroscopic analysis in [Villforth, Sarajedini, and Koekemoer 2012](#). As previously described the luminosity function of high-redshift AGNs significantly differs from the one at low redshift. The purpose of these papers was to detect Low Luminosity Active Galactic Nucleus (LLAGN) and contribute to the study on the evolution of AGN with the redshift.

Authors tested different statistical method for the detection of variability in sparsely sample data. They applied those methods to the Great Observatories Origins Deep Survey (GOODS) North and South fields proposing a catalogue of 139 variability-selected AGNs.

Original data have been taken with the Advanced Camera for Survey (ACS) Wide Field Channel (WFC) aboard HST in the F850LP band, that corresponds to z -band. They have selected objects from the five epoch z -band catalogue [Giavalisco et al. 2004](#) with a signal to noise of 20 or greater. The field was observed in 2003 with a 40-50 days of separation among exposures, therefore this sample can track variability on time-scale within a single year.

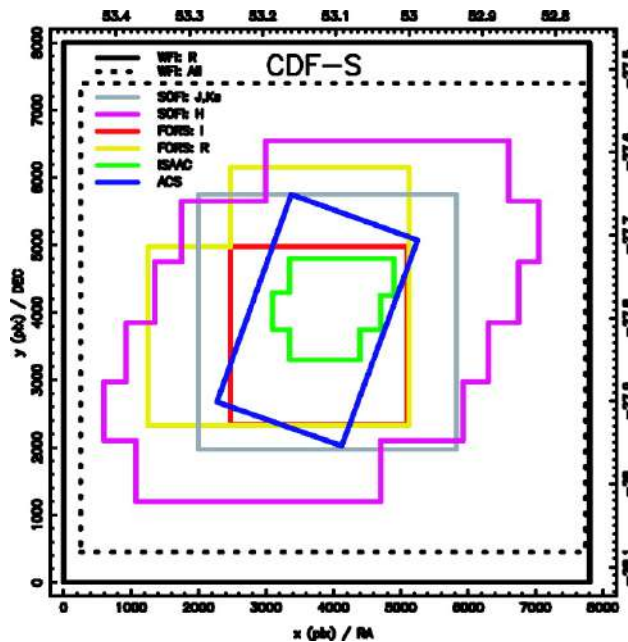


Figure 4.4: This image, extracted from [Giavalisco et al. 2004](#) paper, shows pointings overlap in the same field. This work is based on CDF-S pointing (black lines) while the GOODS South field is in blue.

In their works authors tested and discussed three methods based on χ^2 -, F - and C -statistics on a sparsely sampled data, comparing results and selection biases.

The χ^2 -statistics is used under the null hypothesis that the object is not variable. It associates each flux measurement with its estimated error. It is therefore well suited if one expects the errors to show significant deviations among epochs.

The F - and C -statistics compare the expected to the observed variances (var) and standard deviation (σ) respectively. The two statistics are not mathematically equivalent, as demonstrated in the paper, because the standard deviation is not a linear operator. Both of those methods do not relate each measurement to its error and they should therefore only be used when similar errors are expected for each epoch.

All three statistical methods tested show equal powers for mock data, however, in case in which the error determination is incorrect, χ^2 shows the highest power, followed by C and F . Authors proposed a robust statistical method that makes possible to control the number of false positives and correct for possible problems in error determination.

They released a catalogue of 139 AGN candidates in the two fields (GOODS North and South). I selected the 88 sources of GOODS South field that overlaps the CDF-S I'm studying and that were spectroscopically confirmed as AGN by the follow up.

This catalogue is visualized in red on Figure 4.3: sources positions track the GOODS (that is the ACS) coverage over the CDF-S field.

Falocco et al. (2015)

The catalogue proposed by [Falocco et al. 2015](#) is based on data in r -band from the SUDARE-VOICE survey, performed with the VLT Survey Telescope (VST). The data are centred on the Chandra Deep Field South covering an area of 2 deg^2 in the u , g , r and i bands. The field was observed regularly for about five months in all filters with different cadences collecting 27 epochs in the CDF-S.

Authors choose to use the r band due to the better temporal sampling. To build a robust catalogue of variable but persistent sources, authors selected only the sources with detections in at least six epochs and a magnitude $r < 23$ mag.

The variability-selection method adopted is the one presented in [Trevese et al. 2008](#) adapted to work on VST data. The catalogue produced contains 175 sources on a very wide area around the CDF-S. Due to its greater sky area compared with the other this catalogue is not reported on Figure 4.3.

Limiting to the area covered by the `detected catalogue`, the sample proposed by [Falocco et al. 2015](#) have been reduced to 6 sources.

Pouliasis et al. (2019)

The last catalogue I considered for the comparison was proposed in [Pouliasis et al. 2019](#).

They studied z -band Hubble Space Telescope (HST) images acquired with WFC ACS camera in the F850LP filter, the same of the GOODS field previously described. In that case authors analysed all publicly available images of GOODS field, collecting in total 437 individual images spanning up to 10 years in some regions. They calibrated the astrometry and the photometry of each single image and selected only sources with a signal-to-noise ratio $(S/N) > 5$ obtaining a catalogue limit of 25.7 mag, very close to the limit of my catalogue.

The wide time range and the large number of images available permitted the authors to have a large database where to look for variable sources. The

adopted method uses the Median Absolute Deviation (MAD) as robust measure of variability and inspired the selection method described in this thesis.

The catalogue proposed is composed by 113 variability-selected AGNs candidates, some of them with X-ray counterparts. This catalogue is visualized with black circles on Figure 4.3 and cover a similar area to the one proposed in [Villforth, Koekemoer, and Grogin 2010](#).

This catalogue has some common aspects with my case, as the statistical method adopted, the long time coverage and the magnitude limit reached. On the other side it is observed from a space instrument in a different photometric band and with a more complex temporal sampling.

4.2.2 Cross-match results

I was interested in evaluating the quality of this selection method. I'd like to check how many AGNs, detected by other works (Trevese, Villforth, Falocco and Pouliaisis), have been selected by this analysis.

My statistical identification method was applied on the catalogue presented in Section 3.2.2, the `detected catalogue`. It was derived from VI-MOS observations selecting only objects that have been detected, at least, in 4 epochs. It is composed by 10004 sources.

I made a cross-match within $< 0.5''^2$ among the `detected catalogue` and all reference catalogues to know how many AGNs from literature are contained.

Then I verified which one my procedure was able to detect.

Paper	# sources	# cross-match detected	# cross-match variables
Trevese et al. 2008	35	14	4
Villforth, Koekemoer, and Grogin 2010	88	38	2
Falocco et al. 2015	6	2	1
Pouliaisis et al. 2019	113	34	2

Table 4.1: Reference catalogues with number of sources, number of cross match found with `detected catalogue` and number of variables found in common

The result of this operation is presented in Table 4.1: for each reference catalogue the second column reports the number of sources that were selected

²as previously explained $0.5''$ corresponds at about 1/2 typical measured seeing and 2.5 times the pixel scale

for comparison. The cross-match with the `detected catalogue` identified the number of sources in the third column while the last one reports the source in common with the variable sample I identified in the catalogue. The last one is therefore the absolute number of known variable AGN, of each specific reference catalogue, that my method was able to identify.

The overlap between the `detected catalogue` and reference catalogues is in the range between 30% to 40% for all cases, independently on the instrument and the band observed.

I expected a larger overlap, especially with the Falocco catalogue thanks to the similar band observed, while only 2 of the 6 sources have a counterpart the `detected catalogue`. I verified that the 4 missing sources were detected in most of the single epoch catalogues, but, due to some instrument or sky contaminations, they were not clearly detected in the 4 epochs needed to be included in the `detected catalogue`.

To have a measure of the efficiency of the variability method implemented in the research of (known) variable AGNs on the sample, I defined the **success rate (SR)** as the ratio between the number of variables my method was able to detected and the total number of variables in common:

$$\text{success rate} = \frac{\# \text{ var detected}}{\# \text{ var total}}$$

For ground observed datasets I have a poor statistic to use the *success rate*. In case of Trevese only 4 of the 14 known AGNs were identified in the sample, while, in case of Falocco I found only 1 of the 2 common sources.

In the case of surveys observed from space, despite the larger number of common sources in the `detected catalogue` (38 and 34), only two of them, the same two for both catalogues, were identified by my method on VIMOS data. These values corresponds to *success rates* of 5% ÷ 6% for Villforth and Pouliaisis catalogues.

This result suggests a strong dependencies of variability methods from characteristics of data set used, as discussed and investigated in the next paragraph.

The cross-match result is visualized on sky in Figure 4.5.

Only 5 sources were found in common among the 73 variability-selected candidate AGNs presented in this work and the catalogues from literature³.

³Should be useful remember that, in case of Trevese and Villforth catalogues, I selected only variable sources spectroscopically confirmed as AGNs by the relative follows up.

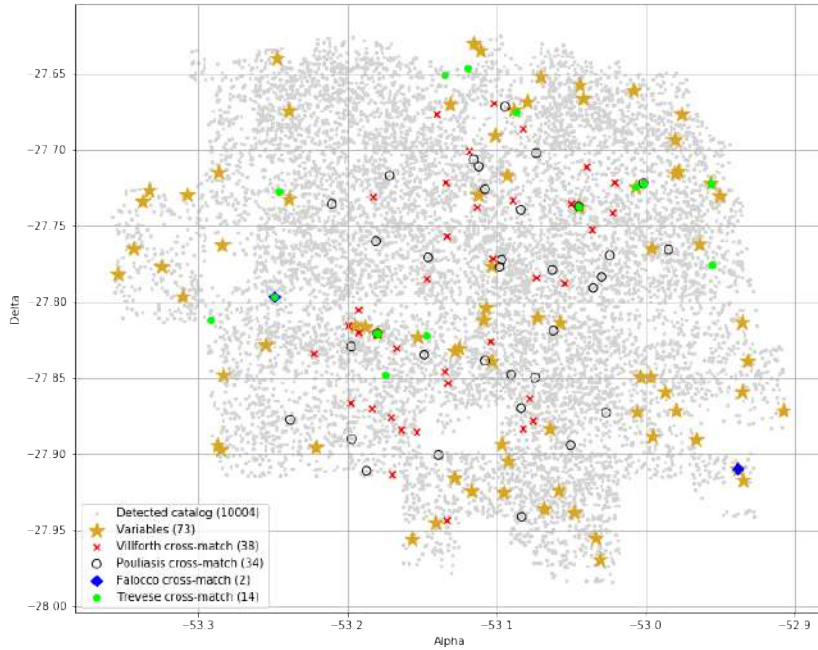


Figure 4.5: Cross match result among full catalogue and reference catalogues. Variables detected in this work are represented as gold stars

Two of them were found in 3 reference catalogues, while other two have a cross match just with the Tievese catalogue. The last one is in common with Falocco et al. .

Comparison result analysis

The previous analysis pointed out a different behaviour between ground and space observed surveys. In the first case my method found a considerable fraction of known AGNs, while it could identify only the $\sim 5\%$ of space observed variables.

It suggests a strong dependency of variability method from the characteristics of dataset used.

To better understand this point it's useful to visualize MAD and MAD* values I measured for known AGNs and their position on the distributions.

In Figure 4.6 are presented MAD (on top) and MAD* (on bottom) distributions for my sample. Sources in common with other catalogues are identified with different symbols.

Could be useful underline two points:

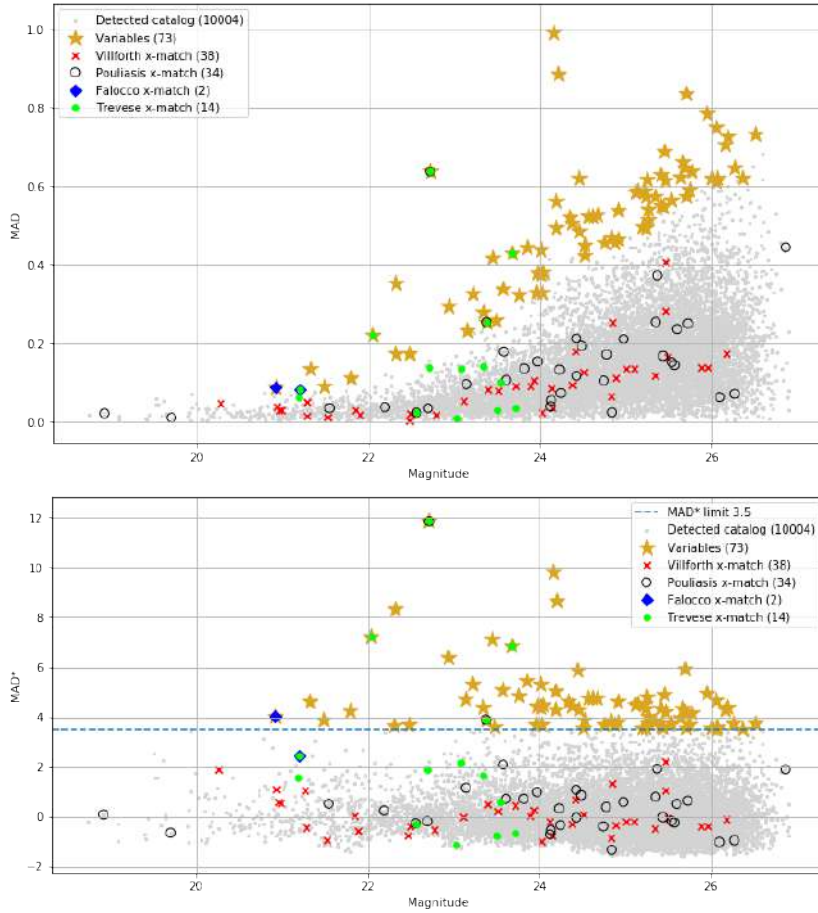


Figure 4.6: Median Absolute Deviation (MAD) (top) and MAD* (bottom) distributions for my sample and for the other four catalogues analysed here. Green dots are for [Trevese et al. 2008](#), red cross for [Villforth, Koekemoer, and Grogin 2010](#), blue squares for [Falocco et al. 2015](#) and black circles for [Pouliaxis et al. 2019](#).

1. MAD values here reported are the variations of magnitude I measured from my sample
2. MAD* values partially depend on the strategy and parameter set adopted for the binning, as described in Section 3.4

On the graphs it's easily possible to identify the two distributions for ground and space based observations.

Green and blue points represent known AGNs from **ground based** observations, Trevese and Falocco respectively. Their samples are limited in depth, < 23 mag for Falocco and < 24 mag for Trevese, but most of their

AGNs show a MAD value relatively high. That explains the greater number of identifications despite the smaller number of AGNs in the **detected catalogue**.

In case of Falocco catalogue (blue points) both of the common sources are bright and have a high MAD value. My method identified one of the two, while the other, despite its MAD value is quite high, has a MAD* not sufficient to be identified. This can be partially due to the binning strategy, because the MAD* depends on the variability shown by the other sources in the same *bin*.

The Trevese catalogue (green points) has 14 AGNs in the **detected catalogue** with one in common with Falocco sample, the not identified one. Four AGNs show a very high variability that my method was able to identify. Other five sources of this sample have a high MAD value and a MAD* value close to 2, below the threshold adopted in this work. The other five, finally, show a very low variability. This is more evident looking at the MAD* values that are close to zero.

The comparison with ground based catalogues underlined some similarity among our samples. Large part of the common AGNs shows a high magnitude variability according to my measures and I was able to identify a good fraction of them.

The small variability of some sources can be originated by different time-scale observed. The variability selection actually depends on the interplay between sampling and the (unpredictable) source variability. Falocco and Trevese used 5 months and 2 years long survey respectively and measured variability within this range.

I used observations over 9 years, stacking together all images per year. This process allowed me to create very deep images, reaching a catalogue completeness limit up to ~ 26 mag, but, on the other side, the stacking process flatten the variability within a single year. Therefore I observed variations on longer time scales than the other ground based catalogues.

Red crosses and black circles in Figure 4.6 represent, instead, **space based** observation catalogues, in that case Villforth and Pouliaxis respectively. Sources in common with them have a different distribution on the graph: they are deeper than the ground based catalogues and they seem not to show a very high variability on average. Have to be considered that sources were observed in *z*-band that should correspond to different rest-frame time scales.

My method was able to identify two of them thanks to their high variability also in the R -band images I used. Most of the other AGNs have, instead, a MAD value in the center of the distribution corresponding to a MAD* close to zero. It means that VIMOS data does not reveal variability on their light-curve.

One of the main reason is the different origin of measured emission. The high resolution of HST made possible to separate the host galaxy contribution from the nucleus, that is the variable part. In that way even a galaxy hosting a faint AGN can be identified through its variability.

From ground telescope the atmosphere seeing spreads the image PSF. During catalogues extractions I adopted the Kron-like automated aperture magnitude (`MAG_AUTO`), as previously explained, that makes not possible to detach nucleus from galaxy contribution. Therefore, in my measures, the luminosity variability of some AGN nuclei has been diluted by their host galaxies. This effect can be negligible for some sources, but can be destructive for host galaxy dominated cases.

Actually, if an aperture photometry is applied, by using a small aperture in the center of the sources, it should be possible to detect core variability of otherwise extended sources, as discussed in [Trevese et al. 2008](#).

It's interesting to underline how the Villforth and Pouliaxis catalogues themselves do not overlap so much, even if they are originated by the same instrument and have a partially common data set. These two catalogues have only 6 sources in common. It is clearly visible in Figure 4.6 where just few sources are identified both with the red cross and the black circle. According to Pouliaxis it is due to the different sources detection algorithm and the larger amount of data. The different time-scale observed is another factor that can give a contribution.

On the other side my detection method identified, on my data, a population of faint variables sources that were not detected at all in space based catalogues.

To summarize: the comparison with known variability-selected AGNs from literature shows how the result of those methods depends strongly from the characteristics of data set used and the time-scale available, both in extension and in frequency. Catalogues from literature show a small overlap among them. This is in agreement with the small overlap I found between my catalogue and the ones from literature.

Ground and space observations seems to select two different populations of variable sources, partially due to the possibility to detach nucleus from

galaxy contribution and to the different band observed. The sample of variables sources I selected is deep as the space based catalogues and identifies a new population of optically variable faint sources. Their characterization is therefore even more interesting.

4.2.3 Characterization of common variable sources

My method was able to identify 5 sources previously detected in other variability-selected AGN catalogues in literature.

SRC_NUM	mag	MAD	Tre	Vil	Fal	Pou
4590	20.92	0.09			x	
9795	23.37	0.25	x	x		x
18162	22.71	0.64	x	x		x
19159	23.67	0.43	x			
19316	22.04	0.22	x			

Table 4.2: Common detected variables

Table 4.2 reports the source identifier (`SRC_NUM`), the median R -band magnitude (`mag`) and the measured MAD, for the five variables in common with literature.

The brightest source 4950 is in common with Falocco et al. catalogue. Sources 9795 and 18162 are instead detected by all the other 3 catalogue, while the 19159 and 19316 are only in the Trevese catalogue.

All of the sources are X-ray emitters and an estimated redshift is provided in those works.

For more details on each source characterization I refer to the dedicated section and to the final catalogues published in Appendix B.

4.3 Comparison with X-ray catalogues

The X-Ray emission is a specific signature of AGNs and X-Ray surveys were often used to select AGNs. For that reason it's very interesting to compare my results with them.

I cross matched the `detected` catalogue and variables I identified with 6 X-Ray catalogues from CDF-S and the Extended Chandra Deep Field South (ECDF-S).

Four of them are produced from Advanced CCD Imaging Spectrometer camera (ACIS) aboard the *Chandra X-Ray Observatory*. They are obtained combining different sets of pointings since 1999 up to 2016, reaching increasing depths.

The [250 ks] *Chandra* catalogue was extracted from the 9 pointing of the ECDF-S observed during the 2004. The catalogue, provided in the paper [Xue et al. 2016](#), presents 1009 sources up to a sensitivity limit in the broad band $F_{0.5-8keV} > 3.5 \times 10^{-16}$ ergs cm⁻² s⁻¹. Limiting the catalogue to the area covered by VIMOS observations it was reduced to 411 sources.

As previously described *Chandra* observed the CDF-S in the year 1999 for the first time, but several pointings were progressively dedicated to this field along the next years up to 2016. The other 3 *Chandra* catalogues are obtained combining observations accumulated in the years.

The [2 Ms] was released by [Luo et al. 2008](#) using the first 23 observations in 1999 - 2000. It contains 578 sources, up to sensitivity limit $F_{0.5-8keV} > 7.1 \times 10^{-17}$ ergs cm⁻² s⁻¹.

A [4 Ms] catalogue was published by [Xue et al. 2011](#) collecting together the 54 observations between 1999 and 2010. It identified 776 sources with a sensitivity limit $F_{0.5-8keV} > 3.2 \times 10^{-17}$ ergs cm⁻² s⁻¹.

The last [7 Ms] catalogue was obtained combining all the 102 observations in 1999 - 2016 period. It was released in the paper [Luo et al. 2017](#) with 1053 sources found over the sensitivity limit of $F_{0.5-8keV} > 1.9 \times 10^{-17}$ ergs cm⁻² s⁻¹.

The other two catalogues are indeed extracted from a *XMM-Newton* telescope deep survey. The CDF-S was initially observed by this instrument in the years 2001-2002 with an exposure of 541 ks. It was observed again in the years 2008-2009 to reach the nominal exposure of 3.45 Ms. The paper [Ranalli et al. 2013](#) provides two catalogues in 2 – 10 keV and 5 – 10 keV bands: they contains 339 and 137 sources respectively, with flux limits $F_{2-10keV} > 6.6 \times 10^{-16}$ ergs cm⁻² s⁻¹ and $F_{5-10keV} > 9.5 \times 10^{-16}$ ergs cm⁻² s⁻¹.

The full list of the observations of each catalogue and the detailed description of the data reduction can be found in the aforementioned papers. All published catalogues include additional information on the sources, such as X-Ray properties, multi-wavelength counterparts and redshifts.

For each X-Ray catalogue the Table 4.3 presents the exptime, the number of sources detected and the cross match both with the `detected` catalogue and with the selected variables.

A total number of 342 counterparts were found on X-ray catalogues. The Figure 4.7 reports the MAD distribution as function of magnitudes where

Paper	Exptime	# sources	# match detected	# match variables
Xue et al. 2016	250 Ks	411	77	7
Luo et al. 2008	2 Ms	578	143	5
Xue et al. 2011	4 Ms	776	222	7
Luo et al. 2017	7 Ms	1053	296	6
Ranalli et al. 2013 (2-10 keV)	3.45 Ms	339	50	5
Ranalli et al. 2013 (5-10 keV)	3.45 Ms	137	24	3

Table 4.3: X-Ray catalogues with exposure time, the number of sources and the number of cross-match with the detected catalogue and the variable one

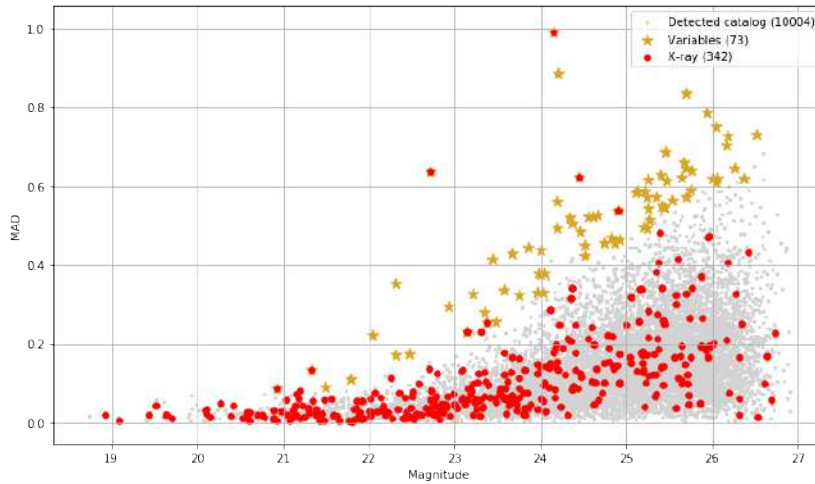


Figure 4.7: MAD as function of magnitudes with X-Ray sources highlighted

X-Ray sources were highlighted with red dots.

Only few of the variable sources seems to be X-ray emitters, while large fraction of the X-ray sources does not show a variability in the in my sample, meaning in VIMOS *R*-band with the single year epoch sampling. This last aspect should be underlined because I understood how the variability selections could depend on the different characteristic of the dataset, as the band and the sampling strategy.

As clear from Table 4.3 and from Figure 4.7, I found a small overlap between sources selected with the variability method and the X-Ray signature. Including all contributions I found an X-Ray counterpart only for 10 variables. Large part of them are in common among X-Ray catalogues and include the 5 variables found in literature and presented in the previous sec-

tion.

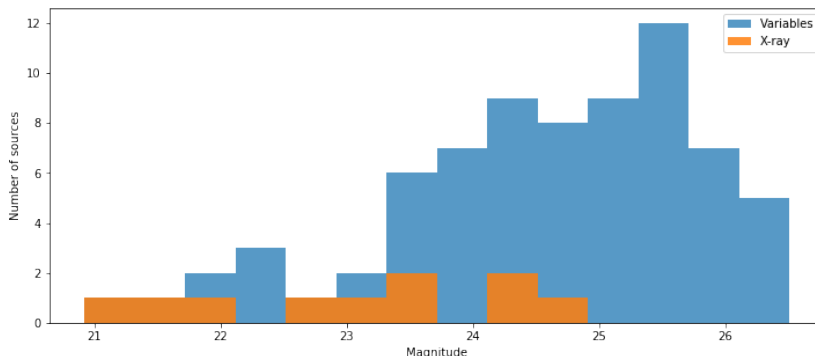


Figure 4.8: Magnitude distribution with X-Ray sources

In Figure 4.8 I present the catalogue magnitude distribution where are identified X-Ray emitters sources. They are distributed in the brightest part up to 25 mag. For the faintest population I did not find any X-Ray counterpart.

This result indicates that AGN selected through variability possess preferentially fainter optical magnitudes than the AGN population selected by X-ray emission, in agreement with the conclusions of [Baldassare, Geha, and Greene 2018](#) and [Pouliasis 2020](#).

4.4 Comparison with multi-wavelength survey catalogues

To better characterize the identified variable sources I used 2 multi-wavelength surveys of this field.

The **Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS)** project catalogue ([Guo et al. 2013](#)) was my first choice. Its observations include some of the deepest images in the visible and near-IR ever achieved over a wide area and have been complemented with the best auxiliary photometry available in the mid-IR with *Spitzer Space Telescope* and in the ultraviolet with ground-based observations.

The catalogue is based on source detection in the Wide Field Camera 3 (WFC3) F160W mosaic, which includes the data from CANDELS deep and

wide observations, as well as previous ERS⁴ and HUDF09⁵ programs.

The catalogue contains 34,930 sources in an area of 173 arcmin².

In addition to the WFC3 bands, the catalogue also includes data from UV (U band from both CTIO/MOSAIC and VLT/VIMOS), optical (*HST*/ACS F435W, F606W, F775W, F814W, and F850LP), and infrared (*HST*/WFC3 F098M, VLT/ISAAC Ks, VLT/HAWK-I Ks, and *Spitzer*/IRAC 3.6, 4.5, 5.8, 8.0 μ m) observations.

In [Dahlen et al. 2013](#) was provided a good photometric redshift for CANDELS sources. In this work 11 independent photometric methods, with different template spectral energy distributions and priors, were used to examine the properties of photometric redshift. Combining results from multiple code, authors were able to produce accurate photometric redshift and errors estimates.

For this work I adopted the CANDELS catalogue release presented in [Santini et al. 2015](#), where 10 different teams computed the stellar masses for the survey sources. They used the same photometry and same redshift within their preferred fitting code, assumptions, priors and parameter grids. Reasonable stellar masses, obtained comparing different results, are provided in the final catalogue.

The second multi-wavelength catalogue I adopted is the **Multiwavelength Survey by Yale-Chile (MUSYC)** presented in [Gawiser et al. 2006](#). They used images from MOSAIC II camera on the Blanco 4m telescope at Cerro Tololo Inter-American Observatory (CTIO), finding 84410 sources in the BVR-detected catalogue.

An high quality 32-band photometric redshift for those sources were provided in [Cardamone et al. 2010](#) work. They presented a deep 18-band optical medium-band photometry from the *Subaru Telescope* in the $\sim 30' \times 30'$ Extended Chandra Deep Field South (ECDF-S). Then authors combined the *Subaru* imaging with existing *UvVRIzJHK* and *Spitzer* images to create a uniform catalogue.

Photometric redshifts are determined using the EAZY code (see [Brammer, van Dokkum, and Coppi 2008](#)) and compared with the ~ 2000 spectroscopic redshifts available in this field.

The cross-match result between catalogue of variables and multi-wavelength catalogues is reported in Table 4.4. Combining both contributions I found an optical and IR counterpart for all but 2 of the 73 sources.

⁴WFC3 Early Release Science Data

⁵Hubble Ultra Deep Field 2009 (HUDF09)

Catalogue	# sources	# variables
CANDELS	34930	26
MUSYC	84410	71

Table 4.4: Common multi-wavelength

4.5 Variables catalogue characterization

To study and characterize selected variable sources I made a cross-match with literature, X-ray and multi-wavelength catalogues, as described in the previous sections.

Only 5 sources were previously detected as variables in literature, just 10 have an X-Ray counterpart, while almost all variables, except 2, were detected in one of the multi-wavelength catalogues.

All of those catalogues provide a **redshift** estimation, and many sources were detected in more than one catalogue. So I associated a redshift value to each source according to this reliability order: (1) Literature, (2) X-Ray catalogues, (3) CANDELS catalogue, (4) MUSYC spectroscopic redshift and (5) MUSYC photometric redshift.

In Appendix B.2 is reported the catalogue of variables with references for all cross-matches and the associated redshifts.

The knowledge of source redshift allowed me to analyse the distance distribution of selected sample.

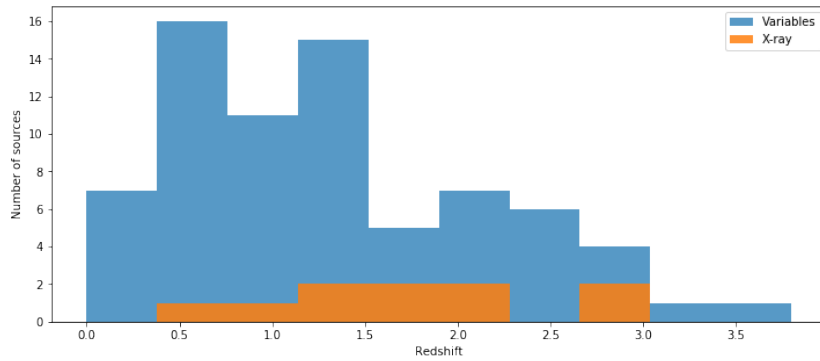


Figure 4.9: Redshift distribution with X-Ray sources

Source redshift distribution is visualized in Figure 4.9. The peak of the distribution is localized between $z = 0.5$ and $z = 1.5$ and the farthest source is at a distance of $z = 3.80$. The 10 sources with an X-Ray counterpart are

distributed between 0.5 and 3.0.

The X-Ray emission is one of the main features used to search for high redshift AGNs. With a variability method I selected, from VIMOS *R*-band images, a population of variable sources, AGN candidates, that seem to have a not detectable X-Ray emission.

To better understand which sources show an X-ray emission, in Figure 4.10 is shown the magnitude distribution with respect to the redshift, where X-ray emitters are highlighted.

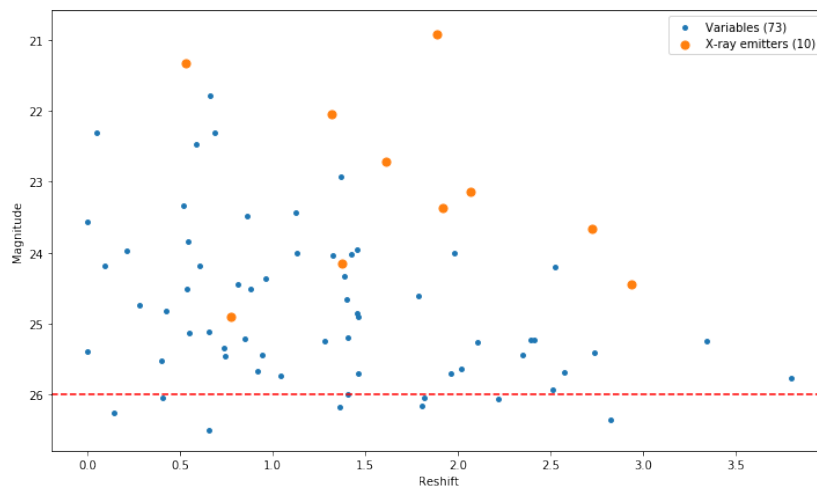


Figure 4.10: Redshift vs Magnitude

In Figure 4.10 orange dots represent X-Ray detected AGNs and the red line indicate the catalogue completeness limit at 26 mag. Only the brightest sources, for redshift range, are X-Ray emitters. It means that X-Ray AGN selection methods are able to identify just the brightest optical sources, while they lose the fainter ones.

It is important to underline that for sources not previously identified should be needed a spectroscopic confirmation of their nature. I will consider them as *candidate* or *bona-fide* AGN, but these results have to be confirmed after a spectroscopic analysis.

The variability method identified a population of optical faint candidate AGNs up to $z = 4$. The most distant ones, the 15936 at $z = 3.34$ and the 15548 at $z = 3.80$, seem not to show any X-Ray emission and they were not previously identified as bona-fide AGNs.

Then, according to results I obtained in this work, X-ray surveys seems to identify just the *tip of the iceberg* of a more numerous candidate AGN population fainter in the optical band. This population can be selected using optical variability methods, as the one presented in this work.

Variable sources were detected even beyond the catalogue completeness limit due to the instrument, and high redshift variables detected are very close to this limit. So I expect that using deeper images should be possible to find even sources at higher distance.

Finally, considering that the drop-out for R -band images is at $z \sim 5$, to search for farther sources it's needed to use IR bands. Variability methods applied to IR surveys are, therefore, very promising to search for a population of faint AGNs at Reionization Era redshifts.

I expect that the next generation of instruments for surveys, both from ground (i.e. *Vera Rubin Observatory* and *Zwicky Transient Facility*) and from space (i.e. *Euclid*) will be optimal to test the variability methods to identify AGNs at high redshift, not detectable using X-ray surveys.

4.6 Some examples of variables from this work

Here I present details for some of the 73 variables I selected in the CDF-S.

4.6.1 Some bright known variable AGNs: the 9795 and the 18162

The first two sources I present are two bright AGNs, previously identified as variables by [Trevese et al. 2008](#), [Villforth, Koekemoer, and Grogin 2010](#) and [Pouliasis et al. 2019](#).

Source 9795 was detected 5 times in my sample, from 2009 to 2015, except 2013, as shown in Figure 4.11. It has a VIMOS R -band median magnitude of 23.37 mag. My method estimated a MAD=0.253 and a MAD*=3.89, identifying it as a variable source.

The light-curve shows a steep decrease in luminosity, of ~ 0.7 mag, between 2009 and 2011, with a subsequent growth during the next years. My purpose in future is to calibrate images for epochs where I found problems in retrieving calibration data, including the 2010. A future availability of a measure for the epoch 2010 will probably better tracks the light-curve shape.

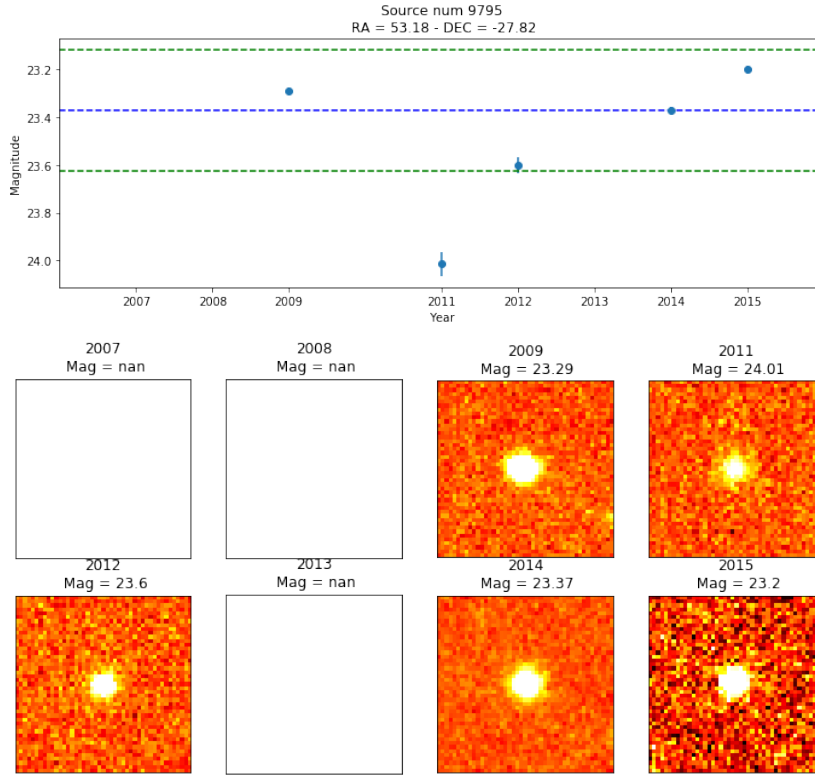


Figure 4.11: Light-curve and images cut-outs for source 9795

This source was identified by all the three variability-selected catalogues previously listed as a QSO with a spectroscopic redshift of $z = 1.92$. It corresponds to ID=96 for [Trevese et al. 2008](#) and ID=17370 for [Pouliasis et al. 2019](#) and it was identified as COMBO J0332433-274914 by [Villforth, Sarajedini, and Koekemoer 2012](#).

The 9795 is a bright X-ray emitter, with a flux measured by *Chandra* of $F_{0.5-8keV} = 9.28 \times 10^{-15}$ ergs cm^{-2} s^{-1} . The multi-wavelength survey counterparts are the ID=9889 in CANDELS and the ID=38426 in MUSYC.

The 18162 is another variable source in common with literature catalogues.

It was detected 6 times, since the 2007 up to the 2015, and its variability is nicely tracked by the light-curve on top of Figure 4.12. The source luminosity oscillates around the median magnitude of 22.71 mag, with a larger measured $\Delta\text{mag} \sim 1.2$ mag between 2008 and 2015.

I measured a MAD=0.638 that corresponds to a very high MAD*=11.86.

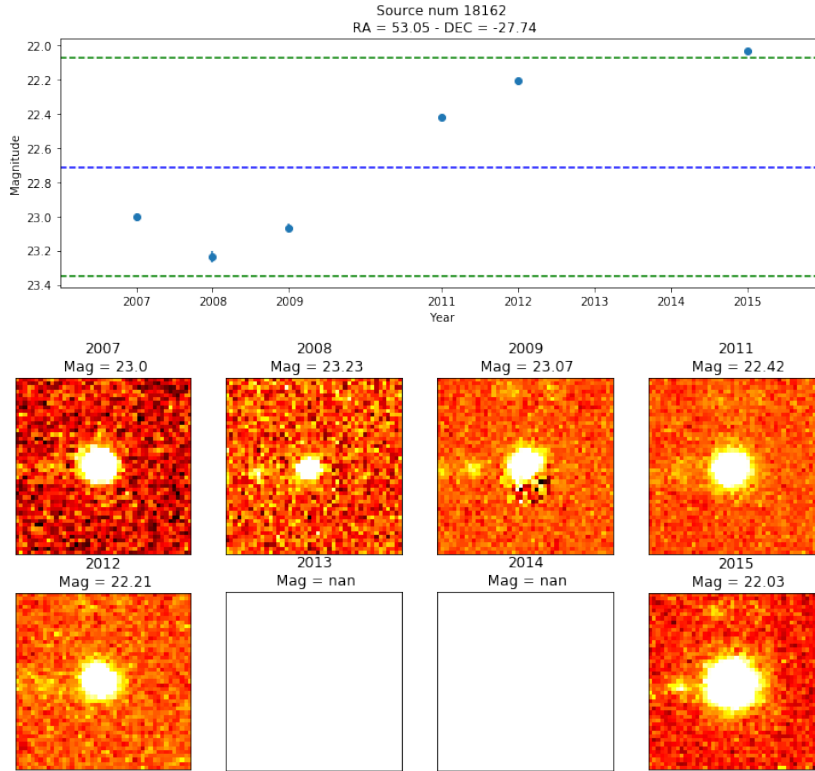


Figure 4.12: Light-curve and images cut-outs for source 18162

This is the highest value I found, as shown in Figure 4.6, where the source 18162 strongly detaches from the distribution.

In the literature this sources was identified as a bright *Seyfert 1* galaxy with an X-ray flux measured by *Chandra* of $F_{0.5-8keV} = 1.95 \times 10^{-14}$ ergs $\text{cm}^{-2} \text{s}^{-1}$ and a spectroscopic redshift of $z = 1.61$. In my reference catalogues it is the ID=84 for [Trevese et al. 2008](#) and ID=5322 for [Pouliasis et al. 2019](#).

This source was one of the objects observed in the [Villforth, Sarajedini, and Koekemoer 2012](#) spectroscopic follow-up, identified as J033210.91-274414.9. In Figure 4.13 the multi-wavelength SED proposed by authors, overlaid several of the best-fitting SEDs, redshifted to the spectroscopic redshift of 1.613. The best-fitting SED was Type 1 QSO, followed by Type 2 QSO and Mrk 231.

This object is extremely radio-loud, with a value $\log(R) \sim 3.5$ and a radio spectrum slope of $\alpha = 0.5$, and shows a power law optical-NIR spectrum. For those reasons different works proposed it as a blazar, a highly beamed and extremely variable AGN.

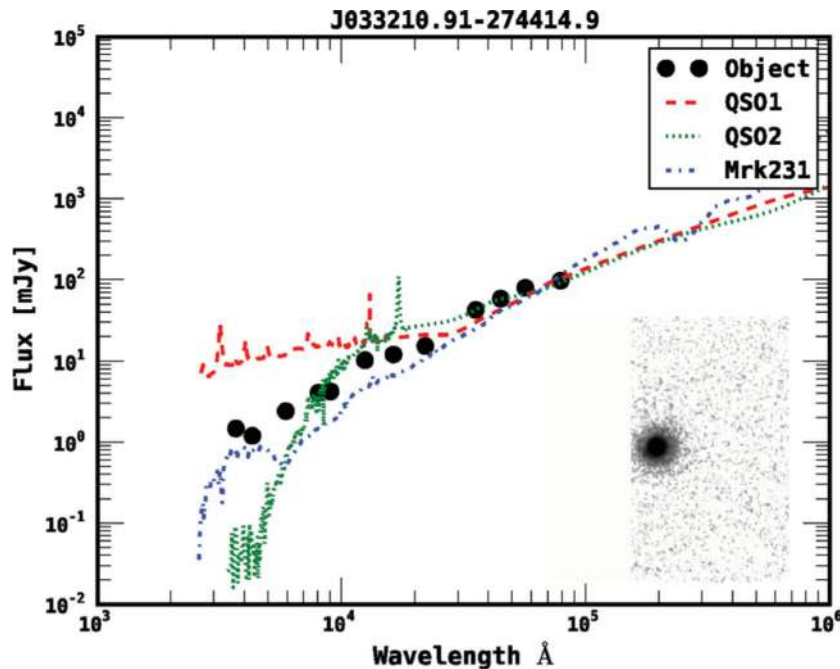


Figure 4.13: Image extracted from [Villforth, Sarajedini, and Koekemoer 2012](#) Figure 10: Multi-wavelength SED for source J033210.91-274414.9 (corresponding to 18162 in my catalogue), overlaid several of the best-fitting SEDs.

This is in good agreement with what I found on my data.

The multi-wavelength survey counterparts for source 18162 are the ID=20447 in CANDELS and the ID=52141 in MUSYC.

4.6.2 High redshift variables: the 15548 and the 15936

The other two sources, that I will briefly present, are two high redshift variables I identified in my sample. Both of them were not detected in X-ray surveys and corresponds to the two higher redshift sources in Figure 4.10.

Source 15548 was detected in five epochs with a median magnitude in R -band of 25.76 mag. I measures a MAD=0.639 and a corresponding MAD*=4.15. For these sources should be useful remember that I selected only detections with a signal-to-noise larger than 5 ($S/N > 5$).

The light-curve on top of Figure 4.14 reports a scatter of ~ 1 mag among different epochs and a high variability. The cut-outs on the bottom of the figure show a faint source with clear variability.

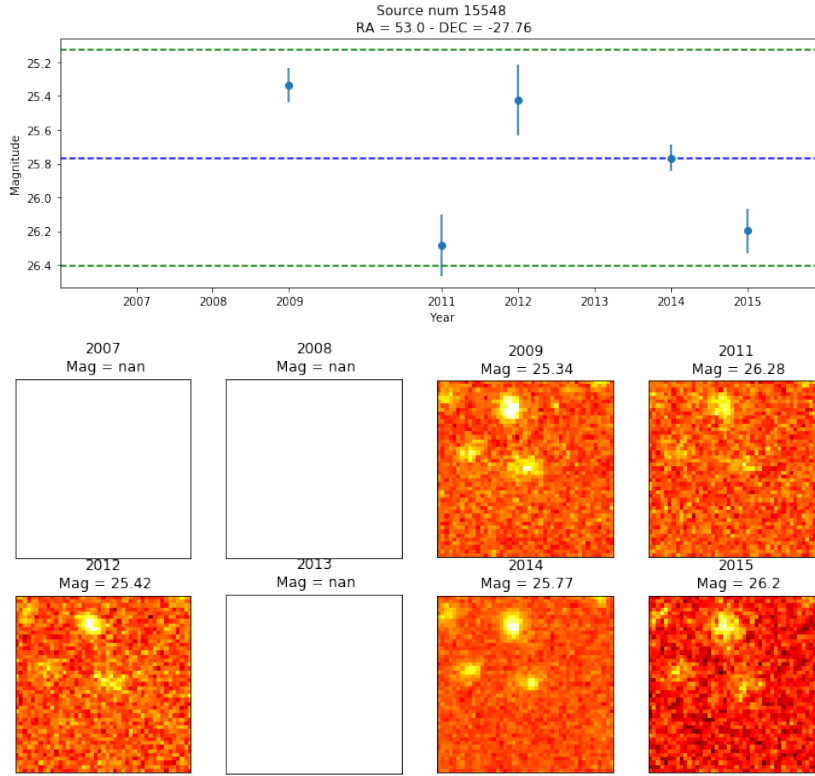


Figure 4.14: Light-curve and images cut-outs for source 15548 at $z = 3.797$

This source has a counterpart only in the MUSYC survey, corresponding to the ID=47354. Its photometric redshift, determined with EAZY code (Brammer, van Dokkum, and Coppi 2008), is $z = 3.797$. This is the farther variable I found.

The last source presented here is the 15936 detected in five epochs from 2007 to 2012. The median magnitude I measured is 25.25 mag, with fluctuations up to ~ 0.9 mag among epochs, as shown on Figure 4.15. The variability identification was triggered by the value of $MAD=0.618$ and a $MAD^*=4.83$.

Even the 15936 source was detected only in the MUSYC survey (ID=47912) that proposed a photometric redshift of $z = 3.343$.

These last two high redshift sources, proposed here as AGN thanks to their variability, were not detected from X-ray surveys. They are therefore two perfect examples of the faint candidate AGN population I proposed to

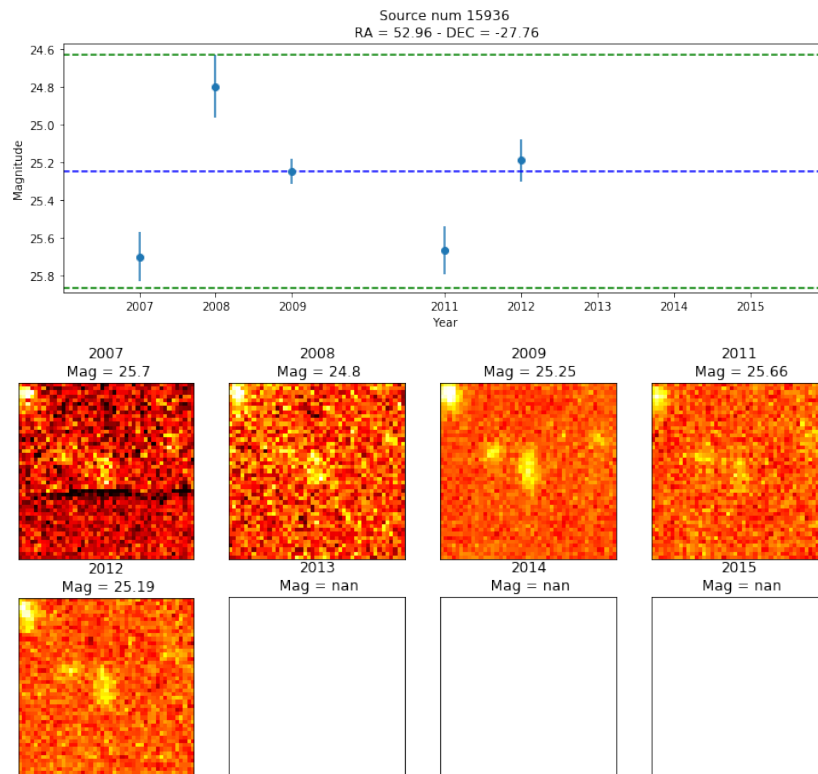


Figure 4.15: Light-curve and images cut-outs for source 15936 at $z = 3.343$

search with the variability method implemented.

Chapter 5

Summary & Conclusions

This work selected a sample of candidate Active Galactic Nuclei (AGNs) in Chandra Deep Field South (CDF-S) region through their variability using deep R-band images from the VIMOS instrument at the VLT telescope.

I present here a little overview on the entire thesis, starting from the science questions that triggered this work, up to the final results that I obtained and a view on the future applications for the method described here.

5.1 A brief summary

This work was triggered by a problem related to the Reionization Era.

According to the current Λ CDM cosmological model, after the nucleosynthetic yields of the Big Bang, the cosmic gas “recombines” and became neutral around $z \approx 1100$, when the Universe was 400,000 years old. The formation of the first PopIII stars in minihalos at $z \sim 20 - 30$ (Trenti and Stiavelli 2009) and then the first galaxies at $z \sim 10$ (Barkana and Loeb 2001; Bromm and Yoshida 2011) started the emission of the UV background (UVB) that ionized the Intergalactic Medium (IGM).

This phase, called *Reionization Era*, defines the moment at which structure formation affected every baryon in the Universe. As shown in Section 1.1.1 measures on CMB from Planck Collaboration et al. 2020b present the picture of a sudden Reionization that started a $z \sim 8 - 10$. The resonant absorption in spectra of distant quasars (e.g. Fan et al. 2006; Becker et al. 2015; McGreer, Mesinger, and D’Odorico 2015) describes an IGM completely reionized at $z \sim 5 - 6$.

It is generally agreed that the IGM is kept ionized by the integrated UV

emission from Active Galactic Nuclei (AGNs) and Star Forming Galaxies (SFG), but there is still no consensus on the relative contribution of these sources as function of epoch. Indeed, according to currently estimated luminosity function (LF), AGNs could mainly contribute to the UV-background around their activity peak at $z \approx 2.5$ (Haardt and Madau 2012).

The declining population of optical bright AGNs for $z > 3$ requires a steep increase of UVB contribution emitted by SFGs for higher redshifts. In particular for $z > 5$ SFGs can ionize the Universe assuming a global averaged absolute escape fraction into the IGM, \bar{f}_{esc} , that exceeds 20% (Haardt and Madau 2012; Bouwens et al. 2012).

Independent measures of a low SFG escape fraction (Boutsia et al. 2011; Grazian et al. 2017; Tanvir et al. 2019) put some stringent constraints on evolution of f_{esc} and disputes the view of a Reionization phase dominated by SFGs.

An exotic reionization model, initially proposed by Arons and McCray 1970, suggested the AGNs as potentially responsible for reionization. Recently different works (Giallongo et al. 2015; Giallongo et al. 2019) claimed to have found a significant population of faint AGNs at $4 < z < 6.5$. This result, combined with the measure of an high f_{esc} for AGN (Grazian et al. 2018), reassess a scenario in which quasars may actually dominate the cosmic reionization process at all epochs.

Different techniques were developed to identify AGNs as the optical color selection, the spectroscopy and X-Ray/optical flux. All of them show some limitations, especially at high redshifts, where just some bright AGNs were detected thanks to deep X-Ray surveys.

This work proposes a method to search for a population of faint AGNs through their variability.

The variability is a feature that can be efficiently used to select AGNs, as demonstrated by Baldassare, Geha, and Greene 2018. Indeed authors show that variability selected objects usually are positioned on color diagrams, as the BTP, far from the typical AGNs area.

This method is based on a set of observations of the same field in different epochs. In case of Chandra Deep Field South (CDF-S), the field I choose for my investigation, different independent works found samples of variable AGNs. Trevese et al. 2008 and Falocco et al. 2015 used ground based CDF-S surveys to identify bright AGNs up to 24 mag. The Great Observatories Origins Deep Survey (GOODS) project provided to Villforth, Koekemoer, and Grogin 2010 and Poulisis et al. 2019 a set of deep Hub-

ble/ACS multi epoch observations. In their works authors presented two catalogues of fainter AGNs (≤ 26 mag) identified thanks to their variability.

Due to the photometric dropout, the redshift that can be reached by a variability method is strictly related to the selected band. To investigate the reionization epoch it is, thus, necessary to use near-IR surveys. These surveys are still not available, but different projects, as the *Vera Rubin Observatory* and the *Euclid* telescope, are in the final phases.

Then the idea was to adopt as deep as possible available data to test a robust variability AGN selection method in a very well studied field, as the CDF-S. This field was indeed observed, since the 1999, by large number of instruments in different wavelengths and epochs.

In this work I have used the deep imaging database of the CDF-S that is available in the ESO Science Archive. In particular, I have chosen to analyze the *R*-band observations carried out with the VIMOS instrument at VLT, since these observations covers an extended time interval of 14 years, from 2002 to 2015. The *R*-band allows us to select variable AGNs up to $z \sim 4$, where the contribution of SFG to the production of ionizing photons starts to rapidly increase, according to mainstream models of reionization driven by galaxies (e.g. [Finkelstein et al. 2019](#)).

I downloaded all raw science observations and the corresponding calibration frames available. Then I built a pipeline, described in Chapter 2, to produce a stacked image by summing up all observations for each year. Due to the lack of some calibration frames I was able to produce the stacked image just for 8 epochs, from 2007 to 2015, except 2010.

After an accurate photometric calibration, I extracted a catalogue from each of the 8 stacked images and cross-matched them to produce a merged catalogue (see Section 3.1). This catalogue contains, for each source, the flux measures at different epochs. I limited the sample to the sources detected at least in four epochs, obtaining the **detected catalogue**. This is composed by 10004 sources, with a apparent magnitude limit of 26 mag, but some sources were detected until ~ 27 mag, though at low completeness level.

The method I built, described in Chapter 3, is based on the statistical parameter Median Absolute Deviation (MAD) as measure of magnitude variability, as proposed by [Pouliasis et al. 2019](#).

I compared the MADs obtained from sources in the same magnitude range. To identify most variables objects, I defined the MAD* as the MAD normalized in the magnitude *bin* (see Section 3.4). Then I selected sources

that show a $MAD^* > 3.5$ as variables (see Figure 4.6 which top figure is also reported in Figure 5.1).

With this method I identified a sample of bona-fide 77 variable sources in the CDF-S reported in Appendix B.1.

5.2 Results & Conclusions

I analysed the catalogue of variables to characterize the sources selected by this method in Chapter 4.

Before proceeding to the analysis, I visually inspected all sources light-curves and images cut-outs to identify not reliable measurements. With this process I discarded 4 sources, leaving a proposed catalogue of 73 variables, including a candidate SN (see Section 4.1).

I compared my catalogue with the other 4 variability selected samples in the CDF-S from [Trevese et al. 2008](#), [Falocco et al. 2015](#), [Villforth, Koekemoer, and Grogin 2010](#) and [Pouliasis et al. 2019](#).

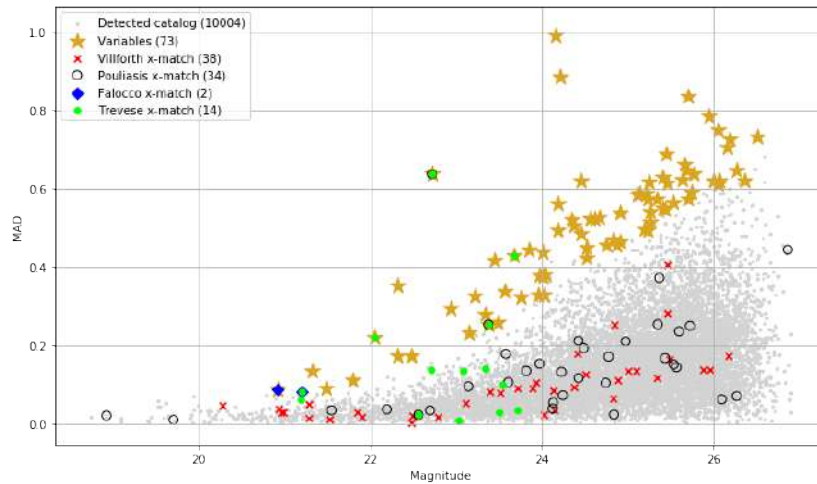


Figure 5.1: Median Absolute Deviation (MAD) distribution for of my sample compared with reference catalogues. Gold stars are variable sources I identified in my sample, green dots for [Trevese et al. 2008](#), red cross for [Villforth, Koekemoer, and Grogin 2010](#), blue squares for [Falocco et al. 2015](#) and black circles for [Pouliasis et al. 2019](#).

As shown in Figure 5.1, I found a small overlap between my catalogues

and the others. Similar results were found comparing the reference catalogues to each other. This underlines a strong dependency of variability method results from the photometric bands observed, sampling time-scale and statistical methods adopted.

Then I looked for X-Ray counterparts of my variables in all available *Chandra* and *XMM* catalogues. I included also the catalogue [7 Ms] released by [Luo et al. 2017](#) that is one of the deepest X-Ray catalogue ever produced. Only 10 of the 73 variables show an X-Ray measurable flux.

To derive redshifts I cross-matched my catalogue with two multi-wavelength surveys: Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) and Multiwavelength Survey by Yale-Chile (MUSYC). I characterized all but 2 sources.

I could then study the source magnitude distribution as function of redshift.

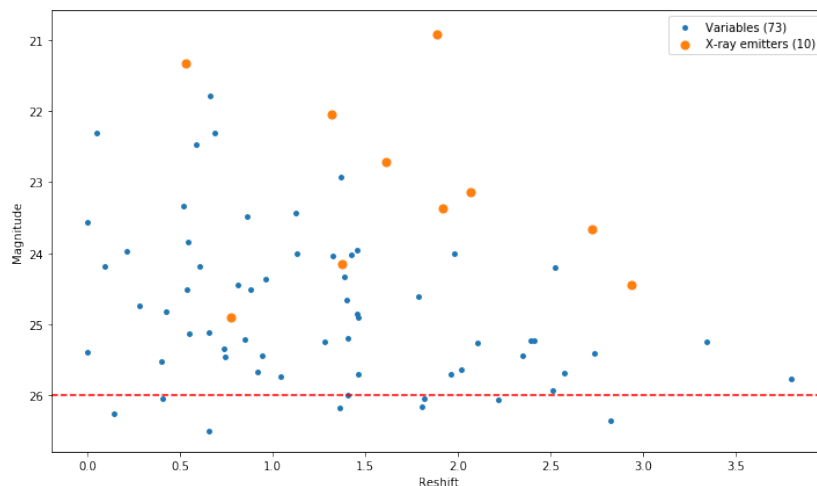


Figure 5.2: The magnitude distribution as function of redshift. Sources with X-ray counterparts are highlighted in orange

Variables with X-ray emissions, showed with orange points in Figure 5.2, are the optical brightest sources along the distribution. X-ray surveys identify just the *tip of the iceberg* of a more numerous AGN population fainter in the optical band.

This result indicates that AGN selected through variability possess preferentially fainter optical magnitudes than the AGN population selected by X-ray emission, in agreement with the conclusions of [Baldassare, Geha, and](#)

Greene 2018 and Pouliaxis 2020.

To summarize the results: the optical variability is able to identify a large number of candidate low-luminosity AGN, especially at high redshifts, including those that remain undetectable even at the deeper observations in the X-rays.

Moreover variability selection strongly depends on the characteristics of the method implemented and data adopted. Therefore results obtained with a single instrument/dataset cannot be considered exhaustive for variable objects selection.

5.3 What about the future

Variability is one of the *hot topics* for current and future surveys.

Time-domain surveys, as the *Zwicky Transient Facility (ZTF)* that had first light at Palomar Observatory in 2017, are excellent tools to find young supernovae and search for rare and exotic transients. Its photometric variability catalogue, with nearly 300 observations each year, is ideal for studies of variable stars, binaries, AGN, and asteroids.

The *Vera C. Rubin Observatory Project* is a very promising instrument for variability studies. The telescope will have an 8.4-meter primary mirror and a 3200 Megapixel digital camera to cover 3.5° field of view, with a pixel scale of $0.2''$. The Vera Rubin will observe the whole sky every 3 days in 6 (*UGRIZY*) bands.

It will be used for color selection for $z < 2.5$ AGN, but it will have also great potential with variability detection. It is expected that the efficiency of AGN selection by variability may be comparable to the color selection efficiency. Moreover the detection fraction will increase as the number of epochs increases, and the use of all six bands will improve the detection fraction even further.

The Vera Rubin temporal information will be especially useful for selecting low-luminosity AGN which would otherwise be swamped by their hosts, as well as radio-loud AGN, which have larger variability amplitudes and shorter variability timescale.

Because of the Y-band drop-out is at $z \sim 7$, its observations will be an excellent tool to investigate at reionization epochs

Finally a couple of new space telescopes will be very interesting for variability studies: the *James Webb Space Telescope (JWST)* and the *Euclid* telescope.

The first one, JWST, will provide images with a strongly improved infrared resolution and sensitivity over Hubble. Its multi epochs observed fields will be very powerful tools to detect very faint variable AGNs in the far Universe through variability.

The second one, Euclid, will observe one third of the full sky, with an excellent resolution, in visible and three infrared bands (*Y*, *J* and *H*). It will also point some fields with regular cadence over 6 years. These observations will be very useful for variability studies.

Thanks to the availability of some deep surveys I expect a very promising future for variability AGNs selection methods.

Acronyms

- ACS** Advanced Camera for Survey. 31, 85, 87, 89
- AGN** Active Galactic Nucleus. 17, 33, 57, 85–87, 89–91, 96
- AGNs** Active Galactic Nuclei. 3, 109, 110
- BLAGN** Broad Line Active Galactic Nucleus. 87
- BPT** Baldwin-Phillips-Terlevich. 27
- CANDELS** Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. 4, 30, 31, 99, 100, 104, 106, 113, 126
- CDF-S** Chandra Deep Field South. 1, 3, 4, 27, 29, 31, 33, 34, 36, 57, 82, 85, 88, 89, 96, 97, 103, 109–112, 141
- CMB** cosmic microwave background. 5, 6
- CTIO** Cerro Tololo Inter-American Observatory. 100
- ECDF-S** Extended Chandra Deep Field South. 96, 97, 100
- ESO** European Southern Observatory. 33
- FWHM** Full Width at Half Maximum. 45
- GIF** Graphic Interchange Format. 83, 141
- GOODS** Great Observatories Origins Deep Survey. 30, 31, 87–89, 110, 141
- HST** Hubble Space Telescope. 30, 85, 87, 89
- HUDF09** Hubble Ultra Deep Field 2009. 100
- IGM** Intergalactic Medium. 3, 109

JWST James Webb Space Telescope. 115

kSZ kinetic Sunyaev-Zeldovich. 11

LAEs Ly α emitters. 13

LBT Large Binocular Telescope. 19

LF luminosity function. 13, 110

LLAGN Low Luminosity Active Galactic Nucleus. 87

LLSs Lyman-limit systems. 15

MAD Median Absolute Deviation. 3, 82, 90, 93, 94, 111, 112, 142

MB MasterBias. 36

MEF Multiple Extensions FITS. 34–36, 83

MF MasterFlat. 36

MUSYC Multiwavelength Survey by Yale-Chile. 4, 100, 104, 106, 107, 113, 126

NSA NASA-Sloan Atlas. 28

S/N signal-to-noise ratio. 55, 89

SFG Star Forming Galaxies. 3, 17, 19, 20, 110, 111

SMBH super-massive black hole. 26

SN Supernova. 34, 84, 86

STRESS Southern inTermediate Redshift ESO Supernova Search. 86, 87

TDE Tidal Distruption event. 84

UVB UV background. 5, 22, 109

VIMOS VISIBLE Multi-Object Spectrograph. 1, 29, 31, 33, 36, 37, 85, 90

VLT Very Large Telescope. 33

VST VLT Survey Telescope. 89

WCS World Coordinate System. 41, 61

WFC Wide Field Channel. 87, 89

WFC3 Wide Field Camera 3. 31, 99, 100

WFI Wide Field Imager. 86

ZTF Zwicky Transient Facility. 114

Appendix A

Calibration

A.1 *SExtractor* configuration file

Here is presented the full configuration file used for the extraction of catalogues from single epoch images, as described in 3.1.

```
# Configuration file for SExtractor 2.19.5
#
#----- Catalog -----
CATALOG_NAME      test.cat      # name of the output catalog
CATALOG_TYPE      FITS_LDAC     # NONE,ASCII,ASCII_HEAD, ASCII_SKYCAT,
                                # ASCII_VOTABLE, FITS_1.0 or FITS_LDAC
PARAMETERS_NAME   default.param # name of the file containing catalog contents
#----- Extraction -----
DETECT_TYPE       CCD           # CCD (linear) or PHOTO (with gamma correction)
DETECT_MINAREA    12           # min. # of pixels above threshold
DETECT_MAXAREA    0            # max. # of pixels above threshold (0=unlimited)
THRESH_TYPE       RELATIVE      # threshold type: RELATIVE (in sigmas)
                                # or ABSOLUTE (in ADUs)
DETECT_THRESH     0.86         # <sigmas> or <threshold>,<ZP> in mag.arcsec-2
ANALYSIS_THRESH   0.86         # <sigmas> or <threshold>,<ZP> in mag.arcsec-2

FILTER            Y            # apply filter for detection (Y or N)?
FILTER_NAME       default.conv  # name of the file containing the filter
FILTER_THRESH     # Threshold[s] for retina filtering

DEBLEND_NTHRESH   32           # Number of deblending sub-thresholds
DEBLEND_MINCONT   0.005        # Minimum contrast parameter for deblending
```

```

CLEAN          Y          # Clean spurious detections? (Y or N)?
CLEAN_PARAM    1.0        # Cleaning efficiency

MASK_TYPE      CORRECT    # type of detection MASKing: can be one of
                        # NONE, BLANK or CORRECT

#----- WEIGHTing -----

WEIGHT_TYPE    MAP_WEIGHT # type of WEIGHTing: NONE, BACKGROUND,
                        # MAP_RMS, MAP_VAR or MAP_WEIGHT
RESCALE_WEIGHTS Y        # Rescale input weights/variances (Y/N)?
WEIGHT_IMAGE   weight.fits # weight-map filename
WEIGHT_GAIN    Y          # modulate gain (E/ADU) with weights? (Y/N)
WEIGHT_THRESH  #          # weight threshold[s] for bad pixels

#----- FLAGging -----

FLAG_IMAGE     flag.fits  # filename for an input FLAG-image
FLAG_TYPE      OR         # flag pixel combination: OR, AND, MIN, MAX
                        # or MOST

#----- Photometry -----

PHOT_APERTURES 5          # MAG_APER aperture diameter(s) in pixels
PHOT_AUTOPARAMS 2.5, 3.5 # MAG_AUTO parameters: <Kron_fact>,<min_radius>
PHOT_PETROPARAMS 2.0, 3.5 # MAG_PETRO parameters: <Petrosian_fact>,
                        # <min_radius>
PHOT_AUTOAPERS 0.0,0.0   # <estimation>,<measurement> minimum apertures
                        # for MAG_AUTO and MAG_PETRO
PHOT_FLUXFRAC  0.5       # flux fraction[s] used for FLUX_RADIUS

SATUR_LEVEL    50000.0    # level (in ADUs) at which arises saturation
SATUR_KEY      SATURATE   # keyword for saturation level (in ADUs)

MAG_ZEROPOINT  0.0       # magnitude zero-point
MAG_GAMMA      4.0       # gamma of emulsion (for photographic scans)
GAIN           0.0       # detector gain in e-/ADU
GAIN_KEY       GAIN      # keyword for detector gain in e-/ADU
PIXEL_SCALE    0.205     # size of pixel in arcsec (0=use FITS WCS info)

#----- Star/Galaxy Separation -----

SEEING_FWHM    0.81      # stellar FWHM in arcsec
STARNNW_NAME   default.nnw # Neural-Network_Weight table filename

#----- Background -----

BACK_TYPE      AUTO      # AUTO or MANUAL
BACK_VALUE     0.0       # Default background value in MANUAL mode

```

```

BACK_SIZE      64          # Background mesh: <size> or <width>,<height>
BACK_FILTERSIZE 3          # Background filter: <size> or <width>,<height>

BACKPHOTO_TYPE LOCAL      # can be GLOBAL or LOCAL
BACKPHOTO_THICK 24        # thickness of the background LOCAL annulus
BACK_FILTTHRESH 0.0      # Threshold above which the background-
                          # map filter operates

#----- Check Image -----

CHECKIMAGE_TYPE APERTURES # can be NONE, BACKGROUND, BACKGROUND_RMS,
                          # MINIBACKGROUND, MINIBACK_RMS, -BACKGROUND,
                          # FILTERED, OBJECTS, -OBJECTS, SEGMENTATION,
                          # or APERTURES
CHECKIMAGE_NAME check.fits # Filename for the check-image

#----- Memory (change with caution!) -----

MEMORY_OBJSTACK 3000      # number of objects in stack
MEMORY_PIXSTACK 300000    # number of pixels in stack
MEMORY_BUFSIZE  1024      # number of lines in buffer

#----- ASSOCIation -----

ASSOC_NAME      sky.list  # name of the ASCII file to ASSOCIate
ASSOC_DATA      2,3,4     # columns of the data to replicate (0=all)
ASSOC_PARAMS    2,3,4     # columns of xpos,ypos[,mag]
ASSOCCOORD_TYPE PIXEL     # ASSOC coordinates: PIXEL or WORLD
ASSOC_RADIUS    2.0       # cross-matching radius (pixels)
ASSOC_TYPE      NEAREST   # ASSOCIation method: FIRST, NEAREST, MEAN,
                          # MAG_MEAN, SUM, MAG_SUM, MIN or MAX
ASSOCSELEC_TYPE MATCHED   # ASSOC selection type: ALL, MATCHED or -MATCHED

#----- Miscellaneous -----

VERBOSE_TYPE    NORMAL    # can be QUIET, NORMAL or FULL
HEADER_SUFFIX   .head     # Filename extension for additional headers
WRITE_XML       N         # Write XML file (Y/N)?
XML_NAME        sex.xml   # Filename for XML output
XSL_URL         file:///usr/share/sextractor/sextractor.xsl
                          # Filename for XSL style-sheet
NTHREADS       1         # 1 single thread

FITS_UNSIGNED   N         # Treat FITS integer values as unsigned (Y/N)?
INTERP_MAXXLG   16        # Max. lag along X for 0-weight interpolation
INTERP_MAXYLG   16        # Max. lag along Y for 0-weight interpolation
INTERP_TYPE     NONE      # Interpolation type: NONE, VAR_ONLY or ALL

#----- Experimental Stuff -----

```

```
PSF_NAME      default.psf    # File containing the PSF model
PSF_NMAX      1          # Max.number of PSFs fitted simultaneously
PATTERN_TYPE  RINGS-HARMONIC # can RINGS-QUADPOLE, RINGS-OCTOPOLE,
               # RINGS-HARMONICS or GAUSS-LAGUERRE
SOM_NAME      default.som  # File containing Self-Organizing Map weights
```

Appendix B

Catalogues

B.1 Catalogue of Variables: main measures

Here is presented the full variable catalogue selected with `obj_num = 0.5` and `mag_step = 0.3 mag` described in section 3.5.

This catalogue is composed by 77 sources, with 68 in the complete sample of the catalogue and 9 out of the instrumental limit. The catalogue in Table B.1 is sorted per `SRC_NUM`, the internal reference number.

Columns are as follow:

`SRC_NUM` identifier number from detected catalogue

α & δ right ascension & declination of source

`CL_STAR` the `CLASS_STAR` value

`mag` median source magnitude

`N_D` number of detection

`MAD` Median Absolute Deviation value

`MAD*` Median Absolute Deviation normalized value

SRC_NUM	α	δ	CL_STAR	mag	N_D	MAD	MAD*
1669	53.031	-27.969	0.88	24.66	5	0.53	4.8
2258	53.157	-27.956	0.90	24.20	4	0.88	8.7
2377	53.034	-27.955	0.13	24.34	5	0.52	4.6
2824	53.141	-27.945	0.96	23.95	4	0.33	3.7
3103	53.048	-27.938	0.50	26.51	5	0.73	3.7
3219	53.069	-27.936	0.41	25.94	4	0.79	5.0
3748	53.117	-27.924	0.90	21.78	4	0.11	4.3
3831	53.096	-27.925	0.64	24.91	4	0.54	4.6
3866	53.059	-27.924	0.25	24.91	5	0.46	3.8
4143	53.128	-27.915	0.12	22.31	5	0.17	3.6
4190	52.935	-27.917	0.96	24.19	4	0.49	4.3
4590	52.938	-27.910	0.98	20.92	4	0.09	4.0
4850	53.092	-27.905	0.51	25.69	4	0.65	4.2
5292	53.285	-27.897	0.67	24.86	5	0.46	3.7
5309	53.221	-27.896	0.72	25.23	5	0.57	4.4
5365	53.288	-27.894	0.76	24.51	5	0.42	3.6
5391	53.097	-27.893	0.77	24.37	5	0.51	4.5
5447	52.966	-27.891	0.69	24.45	5	0.62	5.9
5549	52.995	-27.888	0.68	25.13	5	0.59	4.5
5663	53.065	-27.883	0.65	25.63	5	0.62	4.0
6045	53.006	-27.873	0.59	26.06	4	0.62	3.6
6069	52.908	-27.872	0.70	24.04	4	0.38	4.5
6081	52.980	-27.872	0.68	25.40	5	0.63	4.4
6731	52.987	-27.859	0.74	25.44	5	0.55	3.6
6745	52.936	-27.859	0.59	25.34	4	0.57	3.8
7377	53.003	-27.849	0.44	25.47	5	0.62	4.2
7388	52.997	-27.849	0.57	25.23	4	0.49	3.5
7512	53.284	-27.848	0.69	24.46	5	0.49	4.3
8187	53.104	-27.839	0.91	25.25	6	0.54	4.0
8292	52.931	-27.839	0.57	22.47	4	0.17	3.7
8680	53.129	-27.832	0.41	26.26	4	0.65	3.8
9032	53.125	-27.830	0.51	23.97	6	0.38	4.5
9581	53.255	-27.828	0.04	21.48	5	0.09	3.9
9588	53.153	-27.823	0.10	25.11	4	0.59	4.5
9795	53.180	-27.821	0.98	23.37	5	0.25	3.9
10174	53.188	-27.816	0.32	25.70	5	0.84	5.9
10210	53.195	-27.815	0.61	26.36	4	0.62	3.5
10440	53.058	-27.813	0.00	25.67	4	0.66	4.3
10509	52.935	-27.813	0.97	24.03	4	0.33	3.7

10661	53.109	-27.812	0.35	23.85	6	0.45	5.5
10839	53.073	-27.811	0.06	23.21	6	0.33	5.3
11546	53.108	-27.804	0.38	25.22	5	0.59	4.5
12225	53.311	-27.797	0.94	24.01	5	0.38	4.4
13845	53.354	-27.782	0.96	23.48	4	0.26	3.6
14390	53.325	-27.776	0.50	22.93	4	0.29	6.4
14495	53.103	-27.777	0.07	23.75	5	0.32	4.9
15548	52.997	-27.765	0.22	25.77	5	0.64	4.2
15577	53.344	-27.765	0.94	24.19	4	0.56	5.1
15822	53.284	-27.762	0.73	24.52	4	0.45	3.9
15936	52.964	-27.762	0.08	25.25	5	0.62	4.8
18162	53.045	-27.737	0.98	22.71	6	0.64	11.9
18402	53.338	-27.734	0.77	24.61	4	0.52	4.7
18492	53.240	-27.733	0.64	25.70	6	0.57	3.6
18684	52.950	-27.731	0.53	25.45	4	0.69	4.9
18784	53.308	-27.729	0.93	24.55	5	0.52	4.7
18855	53.113	-27.730	0.01	22.31	4	0.35	8.4
18994	53.333	-27.727	0.96	23.34	4	0.28	4.4
19159	53.007	-27.724	0.94	23.67	5	0.43	6.9
19316	52.956	-27.722	0.97	22.04	5	0.22	7.2
19606	53.093	-27.717	0.73	24.83	5	0.47	3.9
19715	52.980	-27.715	0.67	25.26	6	0.51	3.7
19784	53.287	-27.715	0.96	23.14	6	0.23	4.7
19787	52.978	-27.714	0.54	25.19	4	0.50	3.6
21214	52.981	-27.694	0.24	25.41	5	0.55	3.6
21400	53.101	-27.691	0.98	21.32	4	0.14	4.6
22326	52.976	-27.677	0.61	26.00	4	0.62	3.6
22479	53.239	-27.674	0.48	26.18	4	0.73	4.4
22548	53.089	-27.674	0.84	23.44	6	0.42	7.1
22848	53.131	-27.670	0.52	24.75	5	0.46	3.7
22954	53.080	-27.668	0.96	23.57	4	0.34	5.1
23082	53.042	-27.666	0.49	25.74	5	0.59	3.7
23531	53.009	-27.661	0.68	26.05	5	0.61	3.5
23888	53.045	-27.657	0.63	26.17	5	0.71	4.3
24324	53.071	-27.652	0.52	25.53	6	0.56	3.7
25368	53.248	-27.640	0.54	26.05	5	0.75	4.7
25881	53.111	-27.635	0.95	24.01	4	0.44	5.3
26228	53.116	-27.630	0.97	24.15	4	0.99	9.9

Table B.1: Variables catalogue

B.2 Catalogue of Variables: cross-match results and characterization

Here is presented the result of the cross match among my variables catalogue and some reference catalogues as presented in Chapter 4.

This catalogue presents 73 sources after the removal of 4 not valid detections.

Columns are as follow:

`SRC_NUM` identifier number from *detected catalogue*

`mag` source median magnitude

`var` flag for cross-match with AGNs variables from literature:

T from paper [Trevese et al. 2008](#)

V from paper [Villforth, Koekemoer, and Grogin 2010](#)

F from paper [Falocco et al. 2015](#)

P from paper [Pouliasis et al. 2019](#)

X flag for X-Ray catalogue cross-match:

1 from [250 Ks] [Xue et al. 2016](#) (Chandra telescope)

2 from [2 Ms] [Luo et al. 2008](#) (Chandra telescope)

3 from [4 Ms] [Xue et al. 2011](#) (Chandra telescope)

4 from [7 Ms] [Luo et al. 2017](#) (Chandra telescope)

5 from [3.45 Ms] [Ranalli et al. 2013](#) (XMM telescope)

`IDCANDELS` CANDELS catalogue identifier

`IDMUSYC` MUSYC catalogue identifier

`z` best reshift

`ztype` redshift origin:

1 from literature

2 from X-Ray catalogue

3 from CANDELS

4 from MUSYC (spectroscopic)

5 from MUSYC (photometric)

SRC_NUM	mag	var	X	ID ^{CANDELS}	ID ^{MUSYC}	z	z ^{type}
1669	24.66				14712	1.403	5
2258	24.20				16739	2.526	4
2377	24.34				17005	1.387	5
2824	23.95			130	18625	1.459	3
3103	26.51				19553	0.656	5
3219	25.94			349	19974	2.512	3
3748	21.78			872	22875	0.665	3
3831	24.91		3	742	21869	0.778	3
3866	24.91				22113	1.46	5
4143	22.31			1227	23941	0.6874	3
4190	24.19				23364	0.606	5
4590	20.92	F	1,2,3,4,5		24812	1.887	2
4850	25.69				24941	2.576	5
5292	24.86				26189	1.455	5
5309	25.23			2189	26339	2.414	3
5365	24.51				26592	0.879	5
5391	24.37			2401	26811	0.966	3
5447	24.45		1,2,3,4,5		27179	2.94	2
5549	25.13				27364	0.549	5
5663	25.63			3147		2.019	3
6045	26.06				29829	2.218	5
6069	24.04				30154	1.326	5
6081	25.40					0.0	0
6731	25.44				32053	0.946	5
6745	25.34				32138	0.741	5
7377	25.46				33515	0.746	5
7388	25.23				33624	2.393	4
7512	24.46				33738	0.811	5
8187	25.25			7680	35257	1.283	3
8292	22.47				35457	0.588	5
8680	26.26			8404	36147	0.143	3
9032	23.97			8726	36910	0.2143	3
9588	25.11			9467	37972	0.66	3
9795	23.37	T,V,P	1,2,3,4,5	9889	38426	1.92	1
10174	25.70			10300	38915	1.463	3
10210	26.36			10342	39028	2.822	3
10440	25.67			10601	39384	0.918	3
10509	24.03				39542	1.428	5
10661	23.85			10797	39724	0.543	3

11546	25.22			11690	40864	0.85	3
12225	24.01				42081	1.133	5
13845	23.48				44631	0.861	5
14390	22.93				45532	1.372	5
15548	25.77				47354	3.797	5
15577	24.19				47489	0.095	5
15822	24.52				47791	0.536	5
15936	25.25				47912	3.343	5
18162	22.71	T,V,P	1,2,3,4,5	20447	52141	1.613	1
18402	24.61				52562	1.788	5
18492	25.70				52709	1.96	5
18684	25.45				53155	2.35	5
18855	22.31			21308	53818	0.048	3
18994	23.34				53854	0.517	5
19159	23.67	T	4,5	21937	54369	2.726	1
19316	22.04	T	4		54781	1.318	1
19606	24.83			22690	55382	0.425	3
19715	25.26				55696	2.105	4
19784	23.14		1,2,3,4,5		55835	2.07	2
19787	25.19				55860	1.409	5
21214	25.41				59303	2.739	4
21400	21.32		1,2,3,4	25273	60501	0.533	2
22326	26.00				62036	1.405	5
22479	26.18				62483	1.365	5
22548	23.44				62833	1.128	4
22848	24.75			24062	63147	0.283	3
22954	23.57					0.0	0
23082	25.74				63845	1.047	5
23531	26.05				64499	1.82	5
23888	26.17				64970	1.805	5
24324	25.53				65853	0.402	5
25368	26.05				67635	0.406	5
25881	24.01				68818	1.981	5
26228	24.15		1		69592	1.378	2

Table B.2: Catalogue of variables

Bibliography

- Gunn, James E. and Bruce A. Peterson (Nov. 1965). “On the Density of Neutral Hydrogen in Intergalactic Space.” In: *ApJ* 142, pp. 1633–1636. DOI: 10.1086/148444.
- Markarian, B. E. (Jan. 1967). “Galaxies with an ultraviolet continuum.” In: *Astrofizika* 3, pp. 55–68.
- Arons, J. and R. McCray (Jan. 1970). “Photo-Ionization of Intergalactic Hydrogen by Quasars”. In: *Astrophys. Lett.* 5, p. 123.
- Sunyaev, R. A. and Ya. B. Zeldovich (Feb. 1980). “The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement.” In: *MNRAS* 190, pp. 413–420. DOI: 10.1093/mnras/190.3.413.
- Baldwin, J. A., M. M. Phillips, and R. Terlevich (Feb. 1981). “Classification parameters for the emission-line spectra of extragalactic objects.” In: *PASP* 93, pp. 5–19. DOI: 10.1086/130766.
- Zamorani, G. (Dec. 1985). “A comparison between optically and X-ray selected quasars.” In: *ApJ* 299, pp. 814–820. DOI: 10.1086/163747.
- Trevese, Dario et al. (July 1989). “A Survey of Faint Variable Objects in SA 57”. In: *AJ* 98, p. 108. DOI: 10.1086/115129.
- Giallongo, E., D. Trevese, and F. Vagnetti (Aug. 1991). “Optical Variability of Quasars: Statistics and Cosmological Properties”. In: *ApJ* 377, p. 345. DOI: 10.1086/170365.
- Hook, I. M. et al. (May 1994). “The variability of optically selected quasars.” In: *MNRAS* 268, pp. 305–320. DOI: 10.1093/mnras/268.2.305.
- Bertin, E. and S. Arnouts (June 1996). “SExtractor: Software for source extraction.” In: *A&AS* 117, pp. 393–404. DOI: 10.1051/aas:1996164.
- Cristiani, S. et al. (Feb. 1996). “The optical variability of QSOs.” In: *A&A* 306, p. 395. arXiv: astro-ph/9506140 [astro-ph].
- Barkana, R. and A. Loeb (July 2001). “In the beginning: the first sources of light and the reionization of the universe”. In: *Phys. Rep.* 349.2, pp. 125–238. DOI: 10.1016/S0370-1573(01)00019-9. arXiv: astro-ph/0010468 [astro-ph].

- Giacconi, Riccardo et al. (Apr. 2002). “Chandra Deep Field South: The 1 Ms Catalog”. In: *ApJS* 139.2, pp. 369–410. DOI: 10.1086/338927.
- Trèvese, Dario and Fausto Vagnetti (Jan. 2002). “Quasar Spectral Slope Variability in the Optical Band”. In: *ApJ* 564.2, pp. 624–630. DOI: 10.1086/324541. arXiv: astro-ph/0110075 [astro-ph].
- Giavalisco, M. et al. (Jan. 2004). “The Great Observatories Origins Deep Survey: Initial Results from Optical and Near-Infrared Imaging”. In: *ApJ* 600.2, pp. L93–L98. DOI: 10.1086/379232. arXiv: astro-ph/0309105 [astro-ph].
- Fan, Xiaohui et al. (July 2006). “Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars”. In: *AJ* 132.1, pp. 117–136. DOI: 10.1086/504836. arXiv: astro-ph/0512082 [astro-ph].
- Furlanetto, Steven R., Matias Zaldarriaga, and Lars Hernquist (Jan. 2006). “The effects of reionization on $\text{Ly}\alpha$ galaxy surveys”. In: *MNRAS* 365.3, pp. 1012–1020. DOI: 10.1111/j.1365-2966.2005.09785.x. arXiv: astro-ph/0507266 [astro-ph].
- Gawiser, Eric et al. (Jan. 2006). “The Multiwavelength Survey by Yale-Chile (MUSYC): Survey Design and Deep Public UBVRIz’ Images and Catalogs of the Extended Hubble Deep Field-South”. In: *ApJS* 162.1, pp. 1–19. DOI: 10.1086/497644. arXiv: astro-ph/0509202 [astro-ph].
- Kewley, Lisa J. et al. (Nov. 2006). “The host galaxies and classification of active galactic nuclei”. In: *MNRAS* 372.3, pp. 961–976. DOI: 10.1111/j.1365-2966.2006.10859.x. arXiv: astro-ph/0605681 [astro-ph].
- De Santis, C. et al. (Feb. 2007). “ConvPhot: A profile-matching algorithm for precision photometry”. In: *New A* 12.4, pp. 271–288. DOI: 10.1016/j.newast.2006.10.004. arXiv: astro-ph/0701232 [astro-ph].
- McQuinn, Matthew et al. (Oct. 2007). “Studying reionization with $\text{Ly}\alpha$ emitters”. In: *MNRAS* 381.1, pp. 75–96. DOI: 10.1111/j.1365-2966.2007.12085.x. arXiv: 0704.2239 [astro-ph].
- Brammer, Gabriel B., Pieter G. van Dokkum, and Paolo Coppi (Oct. 2008). “EAZY: A Fast, Public Photometric Redshift Code”. In: *ApJ* 686.2, pp. 1503–1513. DOI: 10.1086/591786. arXiv: 0807.1533 [astro-ph].
- Luo, B. et al. (Nov. 2008). “The Chandra Deep Field-South Survey: 2 Ms Source Catalogs”. In: *ApJS* 179.1, pp. 19–36. DOI: 10.1086/591248. arXiv: 0806.3968 [astro-ph].
- Trevese, D. et al. (2008). “Variability-selected active galactic nuclei from supernova search in the Chandra deep field south”. In: *A&A* 488.1, pp. 73–81. DOI: 10.1051/0004-6361:200809884. arXiv: 0805.2948 [astro-ph].
- Boutsia, K. et al. (Apr. 2009). “Spectroscopic follow-up of variability-selected active galactic nuclei in the Chandra Deep Field South”. In: *A&A* 497.1,

- pp. 81–95. DOI: 10.1051/0004-6361/200811092. arXiv: 0902.0933 [astro-ph.GA].
- Nonino, M. et al. (Aug. 2009). “Deep U Band and R Imaging of GOODS-South: Observations, Data Reduction and First Results”. In: *ApJS* 183.2, pp. 244–260. DOI: 10.1088/0067-0049/183/2/244. arXiv: 0906.4250 [astro-ph.CO].
- Trenti, Michele and Massimo Stiavelli (Apr. 2009). “Formation Rates of Population III Stars and Chemical Enrichment of Halos during the Reionization Era”. In: *ApJ* 694.2, pp. 879–892. DOI: 10.1088/0004-637X/694/2/879. arXiv: 0901.0711 [astro-ph.CO].
- Cardamone, Carolin N. et al. (Aug. 2010). “The Multiwavelength Survey by Yale-Chile (MUSYC): Deep Medium-band Optical Imaging and High-quality 32-band Photometric Redshifts in the ECDF-S”. In: *ApJS* 189.2, pp. 270–285. DOI: 10.1088/0067-0049/189/2/270. arXiv: 1008.2974 [astro-ph.CO].
- Villforth, Carolin, Anton M. Koekemoer, and Norman A. Grogin (Nov. 2010). “A New Extensive Catalog of Optically Variable Active Galactic Nuclei in the GOODS Fields and a New Statistical Approach to Variability Selection”. In: *ApJ* 723.1, pp. 737–754. DOI: 10.1088/0004-637X/723/1/737. arXiv: 1008.3384 [astro-ph.CO].
- Bolton, J. S. et al. (Sept. 2011). “How neutral is the intergalactic medium surrounding the redshift $z = 7.085$ quasar ULAS J1120+0641?” In: *MNRAS* 416.1, pp. L70–L74. DOI: 10.1111/j.1745-3933.2011.01100.x. arXiv: 1106.6089 [astro-ph.CO].
- Boutsia, K. et al. (July 2011). “A Low Escape Fraction of Ionizing Photons of $L_{\text{Ly}\alpha} \sim L^*$ Lyman Break Galaxies at $z = 3.3$ ”. In: *ApJ* 736.1, 41, p. 41. DOI: 10.1088/0004-637X/736/1/41. arXiv: 1104.5237 [astro-ph.GA].
- Bromm, Volker and Naoki Yoshida (Sept. 2011). “The First Galaxies”. In: *ARA&A* 49.1, pp. 373–407. DOI: 10.1146/annurev-astro-081710-102608. arXiv: 1102.4638 [astro-ph.CO].
- Mortlock, Daniel J. et al. (June 2011). “A luminous quasar at a redshift of $z = 7.085$ ”. In: *Nature* 474.7353, pp. 616–619. DOI: 10.1038/nature10159. arXiv: 1106.6088 [astro-ph.CO].
- Xue, Y. Q. et al. (July 2011). “The Chandra Deep Field-South Survey: 4 Ms Source Catalogs”. In: *ApJS* 195.1, 10, p. 10. DOI: 10.1088/0067-0049/195/1/10. arXiv: 1105.5643 [astro-ph.CO].
- Bouwens, R. J. et al. (June 2012). “Lower-luminosity Galaxies Could Reionize the Universe: Very Steep Faint-end Slopes to the UV Luminosity Functions at $z \sim 5-8$ from the HUDF09 WFC3/IR Observations”. In: *ApJ* 752.1, L5, p. L5. DOI: 10.1088/2041-8205/752/1/L5. arXiv: 1105.2038 [astro-ph.CO].

- Fiore, F. et al. (Jan. 2012). “Faint high-redshift AGN in the Chandra deep field south: the evolution of the AGN luminosity function and black hole demography”. In: *A&A* 537, A16, A16. DOI: 10.1051/0004-6361/201117581. arXiv: 1109.2888 [astro-ph.CO].
- Haardt, Francesco and Piero Madau (Feb. 2012). “Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background”. In: *ApJ* 746.2, 125, p. 125. DOI: 10.1088/0004-637X/746/2/125. arXiv: 1105.2039 [astro-ph.CO].
- Villforth, Carolin, Vicki Sarajedini, and Anton Koekemoer (2012). “The spectral energy distributions, host galaxies and environments of variability-selected active galactic nuclei in GOODS-South”. In: *MNRAS* 426.1, pp. 360–376. DOI: 10.1111/j.1365-2966.2012.21732.x. arXiv: 1207.4478 [astro-ph.CO].
- van Dokkum, Pieter G., J. Bloom, and Malte Tewes (July 2012). *L.A. Cosmic: Laplacian Cosmic Ray Identification*. ascl: 1207.005.
- Dahlen, Tomas et al. (Oct. 2013). “A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation”. In: *ApJ* 775.2, 93, p. 93. DOI: 10.1088/0004-637X/775/2/93. arXiv: 1308.5353 [astro-ph.CO].
- Guo, Yicheng et al. (Aug. 2013). “CANDELS Multi-wavelength Catalogs: Source Detection and Photometry in the GOODS-South Field”. In: *ApJS* 207.2, 24, p. 24. DOI: 10.1088/0067-0049/207/2/24. arXiv: 1308.4405 [astro-ph.CO].
- Kewley, Lisa J. et al. (Sept. 2013). “The Cosmic BPT Diagram: Confronting Theory with Observations”. In: *ApJ* 774.1, L10, p. L10. DOI: 10.1088/2041-8205/774/1/L10. arXiv: 1307.0514 [astro-ph.CO].
- Ranalli, P. et al. (July 2013). “The XMM deep survey in the CDF-S. III. Point source catalogue and number counts in the hard X-rays”. In: *A&A* 555, A42, A42. DOI: 10.1051/0004-6361/201321211. arXiv: 1304.5717 [astro-ph.HE].
- Chornock, R. et al. (May 2014). “GRB 140515A at $z=6.33$: Constraints on the End of Reionization From a Gamma-ray Burst in a Low Hydrogen Column Density Environment”. In: *arXiv e-prints*, arXiv:1405.7400, arXiv:1405.7400. arXiv: 1405.7400 [astro-ph.CO].
- Becker, George D. et al. (Mar. 2015). “Evidence of patchy hydrogen reionization from an extreme Ly α trough below redshift six”. In: *MNRAS* 447.4, pp. 3402–3419. DOI: 10.1093/mnras/stu2646. arXiv: 1407.4850 [astro-ph.CO].
- Bouwens, R. J. et al. (Oct. 2015). “Reionization After Planck: The Derived Growth of the Cosmic Ionizing Emissivity Now Matches the Growth of the Galaxy UV Luminosity Density”. In: *ApJ* 811.2, 140, p. 140. DOI: 10.1088/0004-637X/811/2/140. arXiv: 1503.08228 [astro-ph.CO].

- Falocco, S. et al. (July 2015). “SUDARE-VOICE variability-selection of active galaxies in the Chandra Deep Field South and the SERVS/SWIRE region”. In: *A&A* 579, A115, A115. DOI: 10.1051/0004-6361/201425111. arXiv: 1505.02668 [astro-ph.GA].
- Finkelstein, Steven L. et al. (Sept. 2015). “The Evolution of the Galaxy Rest-frame Ultraviolet Luminosity Function over the First Two Billion Years”. In: *ApJ* 810.1, 71, p. 71. DOI: 10.1088/0004-637X/810/1/71. arXiv: 1410.5439 [astro-ph.GA].
- Giallongo, E. et al. (June 2015). “Faint AGNs at $z \lesssim 4$ in the CANDELS GOODS-S field: looking for contributors to the reionization of the Universe”. In: *A&A* 578, A83, A83. DOI: 10.1051/0004-6361/201425334. arXiv: 1502.02562 [astro-ph.CO].
- McGreer, Ian D., Andrei Mesinger, and Valentina D’Odorico (Feb. 2015). “Model-independent evidence in favour of an end to reionization by $z \approx 6$ ”. In: *MNRAS* 447.1, pp. 499–505. DOI: 10.1093/mnras/stu2449. arXiv: 1411.5375 [astro-ph.CO].
- Santini, P. et al. (2015). “Stellar Masses from the CANDELS Survey: The GOODS-South and UDS Fields”. In: *ApJ* 801.2, 97, p. 97. DOI: 10.1088/0004-637X/801/2/97. arXiv: 1412.5180 [astro-ph.GA].
- Grazian, A. et al. (Jan. 2016). “The Lyman continuum escape fraction of galaxies at $z = 3.3$ in the VUDS-LBC/COSMOS field”. In: *A&A* 585, A48, A48. DOI: 10.1051/0004-6361/201526396. arXiv: 1509.01101 [astro-ph.GA].
- Guaita, L. et al. (Mar. 2016). “Limits on the LyC signal from $z \sim 3$ sources with secure redshift and HST coverage in the E-CDFS field”. In: *A&A* 587, A133, A133. DOI: 10.1051/0004-6361/201527597. arXiv: 1601.03057 [astro-ph.GA].
- Planck Collaboration et al. (Dec. 2016). “Planck intermediate results. XLVII. Planck constraints on reionization history”. In: *A&A* 596, A108, A108. DOI: 10.1051/0004-6361/201628897. arXiv: 1605.03507 [astro-ph.CO].
- Xue, Y. Q. et al. (June 2016). “The 2 Ms Chandra Deep Field-North Survey and the 250 ks Extended Chandra Deep Field-South Survey: Improved Point-source Catalogs”. In: *ApJS* 224.2, 15, p. 15. DOI: 10.3847/0067-0049/224/2/15. arXiv: 1602.06299 [astro-ph.GA].
- Grazian, A. et al. (June 2017). “Lyman continuum escape fraction of faint galaxies at $z \sim 3.3$ in the CANDELS/GOODS-North, EGS, and COSMOS fields with LBC”. In: *A&A* 602, A18, A18. DOI: 10.1051/0004-6361/201730447. arXiv: 1703.00354 [astro-ph.GA].
- Greig, Bradley et al. (Apr. 2017). “Are we witnessing the epoch of reionisation at $z = 7.1$ from the spectrum of J1120+0641?” In: *MNRAS* 466.4,

- pp. 4239–4249. DOI: 10.1093/mnras/stw3351. arXiv: 1606.00441 [astro-ph.CO].
- Luo, B. et al. (Jan. 2017). “The Chandra Deep Field-South Survey: 7 Ms Source Catalogs”. In: *ApJS* 228.1, 2, p. 2. DOI: 10.3847/1538-4365/228/1/2. arXiv: 1611.03501 [astro-ph.GA].
- Madau, Piero (Dec. 2017). “Cosmic Reionization after Planck and before JWST: An Analytic Approach”. In: *ApJ* 851.1, 50, p. 50. DOI: 10.3847/1538-4357/aa9715. arXiv: 1710.07636 [astro-ph.GA].
- Ota, Kazuaki et al. (July 2017). “A New Constraint on Reionization from the Evolution of the Ly α Luminosity Function at $z \sim 6$ -7 Probed by a Deep Census of $z = 7.0$ Ly α Emitter Candidates to $0.3L^*$ ”. In: *ApJ* 844.1, 85, p. 85. DOI: 10.3847/1538-4357/aa7a0a. arXiv: 1703.02501 [astro-ph.GA].
- Bañados, Eduardo et al. (Jan. 2018). “An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5”. In: *Nature* 553.7689, pp. 473–476. DOI: 10.1038/nature25180. arXiv: 1712.01860 [astro-ph.GA].
- Baldassare, Vivienne F., Marla Geha, and Jenny Greene (Dec. 2018). “Identifying AGNs in Low-mass Galaxies via Long-term Optical Variability”. In: *ApJ* 868.2, 152, p. 152. DOI: 10.3847/1538-4357/aae6cf. arXiv: 1808.09578 [astro-ph.GA].
- Bosman, Sarah E. I. et al. (Sept. 2018). “New constraints on Lyman- α opacity with a sample of 62 quasars at $z \lesssim 5.7$ ”. In: *MNRAS* 479.1, pp. 1055–1076. DOI: 10.1093/mnras/sty1344. arXiv: 1802.08177 [astro-ph.GA].
- Boutsia, K. et al. (Dec. 2018). “A High Space Density of L* Active Galactic Nuclei at $z \sim 4$ in the COSMOS Field”. In: *ApJ* 869.1, 20, p. 20. DOI: 10.3847/1538-4357/aae6c7. arXiv: 1811.06404 [astro-ph.GA].
- Davies, Frederick B. et al. (Sept. 2018). “Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at $z \lesssim 7$ ”. In: *ApJ* 864.2, 142, p. 142. DOI: 10.3847/1538-4357/aad6dc. arXiv: 1802.06066 [astro-ph.CO].
- Eilers, Anna-Christina, Frederick B. Davies, and Joseph F. Hennawi (Sept. 2018). “The Opacity of the Intergalactic Medium Measured along Quasar Sightlines at $z \sim 6$ ”. In: *ApJ* 864.1, 53, p. 53. DOI: 10.3847/1538-4357/aad4fd. arXiv: 1807.04229 [astro-ph.GA].
- Grazian, A. et al. (2018). “The contribution of faint AGNs to the ionizing background at $z \sim 4$ ”. In: *A&A* 613, A44, A44. DOI: 10.1051/0004-6361/201732385. arXiv: 1802.01953 [astro-ph.GA].
- Mason, Charlotte A. et al. (Mar. 2018). “The Universe Is Reionizing at $z \sim 7$: Bayesian Inference of the IGM Neutral Fraction Using Ly α Emission from Galaxies”. In: *ApJ* 856.1, 2, p. 2. DOI: 10.3847/1538-4357/aab0a7. arXiv: 1709.05356 [astro-ph.CO].

- Millea, Marius and François Bouchet (Sept. 2018). “Cosmic microwave background constraints in light of priors over reionization histories”. In: *A&A* 617, A96, A96. DOI: 10.1051/0004-6361/201833288. arXiv: 1804.08476 [astro-ph.CO].
- Sánchez-Sáez, P. et al. (Sept. 2018). “The QUEST-La Silla AGN Variability Survey: Connection between AGN Variability and Black Hole Physical Properties”. In: *ApJ* 864.1, 87, p. 87. DOI: 10.3847/1538-4357/aad7f9. arXiv: 1808.00967 [astro-ph.GA].
- Eilers, Anna-Christina et al. (Aug. 2019). “Anomaly in the Opacity of the Post-reionization Intergalactic Medium in the Ly α and Ly β Forest”. In: *ApJ* 881.1, 23, p. 23. DOI: 10.3847/1538-4357/ab2b3f. arXiv: 1906.05874 [astro-ph.GA].
- Finkelstein, Steven L. et al. (July 2019). “Conditions for Reionizing the Universe with a Low Galaxy Ionizing Photon Escape Fraction”. In: *ApJ* 879.1, 36, p. 36. DOI: 10.3847/1538-4357/ab1ea8. arXiv: 1902.02792 [astro-ph.CO].
- Giallongo, E. et al. (Oct. 2019). “Space Densities and Emissivities of Active Galactic Nuclei at $z \lesssim 4$ ”. In: *ApJ* 884.1, 19, p. 19. DOI: 10.3847/1538-4357/ab39e1.
- Kulkarni, Girish et al. (May 2019). “Large Ly α opacity fluctuations and low CMB τ in models of late reionization with large islands of neutral hydrogen extending to $z \lesssim 5.5$ ”. In: *MNRAS* 485.1, pp. L24–L28. DOI: 10.1093/mnrasl/slz025. arXiv: 1809.06374 [astro-ph.CO].
- Pouliasis, E. et al. (2019). “Robust identification of active galactic nuclei through HST optical variability in GOODS-S: comparison with the X-ray and mid-IR-selected samples”. In: *MNRAS* 487.3, pp. 4285–4304. DOI: 10.1093/mnras/stz1483. arXiv: 1905.11995 [astro-ph.GA].
- Schindler, Jan-Torge et al. (July 2019a). “The Extremely Luminous Quasar Survey in the Pan-STARRS 1 Footprint (PS-ELQS)”. In: *ApJS* 243.1, 5, p. 5. DOI: 10.3847/1538-4365/ab20d0. arXiv: 1905.04069 [astro-ph.GA].
- Schindler, Jan-Torge et al. (Feb. 2019b). “The Extremely Luminous Quasar Survey in the Sloan Digital Sky Survey Footprint. III. The South Galactic Cap Sample and the Quasar Luminosity Function at Cosmic Noon”. In: *ApJ* 871.2, 258, p. 258. DOI: 10.3847/1538-4357/aaf86c. arXiv: 1812.04639 [astro-ph.GA].
- Tanvir, N. R. et al. (Mar. 2019). “The fraction of ionizing radiation from massive stars that escapes to the intergalactic medium”. In: *MNRAS* 483.4, pp. 5380–5408. DOI: 10.1093/mnras/sty3460. arXiv: 1805.07318 [astro-ph.GA].
- Meiksin, Avery (Feb. 2020). “The influence of metagalactic ultraviolet background fluctuations on the high-redshift Ly α forest”. In: *MNRAS* 491.4,

- pp. 4884–4893. DOI: 10 . 1093 / mnras / stz3395. arXiv: 1912 . 00828 [astro-ph.CO].
- Planck Collaboration et al. (Sept. 2020a). “Planck 2018 results. I. Overview and the cosmological legacy of Planck”. In: A&A 641, A1, A1. DOI: 10 . 1051/0004-6361/201833880. arXiv: 1807.06205 [astro-ph.CO].
- Planck Collaboration et al. (Sept. 2020b). “Planck 2018 results. VI. Cosmological parameters”. In: A&A 641, A6, A6. DOI: 10 . 1051 / 0004 - 6361/201833910. arXiv: 1807.06209 [astro-ph.CO].
- Pouliasis, Ektoras (Feb. 2020). “Identification of Active Galactic Nuclei through different selection techniques”. PhD thesis. IAASARS, National Observatory of Athens.

List of Tables

2.1	List of parameters submitted to Eso Archive Facilities to query for VIMOS <i>R</i> -band pointing 15' around the center of CDF-S .	34
2.2	Number of observation for each year, the exposure time range of the images and coverage period during the year. Asterisks indicates years I was able to calibrate and use for my analysis.	35
3.1	Zero-point values with number of cross-matched sources and σ value	61
3.2	List of catalogues with effective area covered in deg ² , catalogue magnitude limit, number of sources detected and number of cross match with the 2009 catalogue	66
4.1	Reference catalogues with number of sources, number of cross match found with detected catalogue and number of variables found in common	90
4.2	Common detected variables	96
4.3	X-Ray catalogues with exposure time, the number of sources and the number of cross-match with the detected catalogue and the variable one	98
4.4	Common multi-wavelength	101
B.1	Variables catalogue	125
B.2	Catalogue of variables	128

List of Figures

1.1	Artistic representation of Universe history. The figure illustrates deepness reached by Hubble Deep fields and the limits of Reionization era. <i>Illustration: Nik Spencer; Source: NASA/WMAP Science Team; R. Ellis (Caltech)</i>	6
1.2	From Fan et al. 2006 Figure 5. Evolution of the optical depth with combined Ly α , Ly β and Ly γ results.	8
1.3	From Becker et al. 2015 Figure 1. Spectra of $z \sim 6$ quasars analysed in the work.	9
1.4	From Becker et al. 2015 Figure 6. Ly α τ_{eff} measurements. Black circles and arrows (lower limits) are from Fan et al. 2006 . Blue squares, error bars, and arrows are from Becker et al. 2015 . The two blue arrows at $z \simeq 5.63$ and 5.80 with $\tau_{eff} \gtrsim 7$ are from the line of sight towards ULAS J0148+0600.	10
1.5	From Planck Collaboration et al. 2020b Figure 44 (Top): marginalized constraints on the optical depth to reionization from lowE alone, assuming different models of reionization and different priors over the model parameters. Only reionization parameters are varied here. The solid lines use a flat prior on τ , while the dashed line uses a flat prior on the knot amplitudes.	12
1.6	From Planck Collaboration et al. 2020a Figure 36 : Free electron fraction, $x_e(z)$, constrained from Planck using the FlexKnots method of Planck Collaboration et al. 2020b and Millea and Bouchet 2018 , along with astrophysical constraints, as tabulated in Bouwens et al. 2015 , updated to include Greig et al. 2017 , Bañados et al. 2018 , Mason et al. 2018 , and Davies et al. 2018 . The shaded bands are middle 68th and 95th percentiles	13

1.7	From Ota et al. 2017 Figure 20 (Left). Cosmic reionization history (neutral fraction x_{HI} as a function of redshift) constrained by their $z = 7$ LAE survey and previous studies using various probes. The blue and light blue shaded regions show the 68% and 95% allowed intervals of reionization history constrained by the analysis of the pre-2016 Planck CMB observations data by Planck Collaboration et al. 2016	14
1.8	From Madau 2017 Figure 1 (Left). Curves represent the average neutral fraction $(1 - Q)$ of an IGM that recombines according two different models of t_{rec} and the contribution of LLSs. The data points represents constraints on the ionization state of the IGM from a set of independent measures: i.e. Gunn-Peterson optical depth, Ly α forest opacity, dark pixel statistics and all the other available at paper time.	16
1.9	From Haardt and Madau 2012 Figure 7 (Right): comoving galaxy emissivity (in units of 10^{23} erg s $^{-1}$ Mpc $^{-3}$ Hz $^{-1}$) of 1 ryd photons escaping into the IGM (blue dashed line). The solid red line shows the best-fit quasar emissivity for comparison, while the dot-dashed line shows the total quasars + galaxies emissivity.	18
1.10	From Grazian et al. 2017 Figure 4: Stack of the 69 galaxies with $3.27 \leq z_{spec} \leq 3.40$ in the R (<i>left</i>) and U bands (<i>right</i>) for the COSMOS, EGS, and CANDELS/GOODS- North fields. The size of each image is 16 arcsec by 14 arcsec.	20
1.11	From Grazian et al. 2017 Figure 3: Measured values and the upper limits at 1σ for the LyC relative escape fraction of galaxies at $z \sim 3.3$ in the COSMOS (blue, green), EGS (gray), and CANDELS/GOODS-North fields (red). Small descending arrows show the upper limits of f_{esc}^{rel} associated with individual galaxies. The large arrows (cyan, purple, magenta, orange, gold, dark green) represent the limits to f_{esc}^{rel} derived by stacking the $z \sim 3.3$ galaxies in different luminosity intervals.	21
1.12	From Giallongo et al. 2019 Figure 4: UV 1450 Å AGN luminosity functions in two redshift intervals. Different symbols represent $1/V_{max}$ data points from different surveys as explained in the figure box. Dashed and continuous curves are LF best fits obtained using different models and sub-samples.	23
1.13	From Giallongo et al. 2019 Figure 3: Dimensional X/H as a function of the H -band magnitude for some AGN candidates. Straight lines represent the adopted X-Ray flux limits of the three fields at the 40% completeness level adopted in their work	25

1.14	From Baldassare, Geha, and Greene 2018 Figure 8: BPT diagram showing the positions of the variability-selected AGNs. The black lines show the traditional BPT diagram separation lines (Baldwin, Phillips, and Terlevich 1981 ; Kewley et al. 2006), and the green dashed line shows the star forming main sequence Kewley et al. 2013	28
1.15	The CDF-S 1Ms image (Giacconi et al. 2002) located in the Southern Hemisphere constellation of Fornax. Most of the objects seen in the Chandra Deep Field-South are active galaxies and quasars powered by massive black holes. The intensities of the X-rays in this image are depicted, from lowest to highest energies, by red, yellow, and blue respectively	30
1.16	Bandpass transmission curves for VIMOS filters V , R and I for the four quadrants. The transmission value includes contribution from filters, CCD and telescope optics.	32
2.1	Representation of Data calibration pipeline for VIMOS instrument	39
2.2	The first detector of a raw Sky Flat of 2009	43
2.3	One of the MasterFlats created for year 2009	44
2.4	Raw pointing observed in 2009, August 18 th	46
2.5	Top: Stacked image for the year 2009. Bottom: corresponding weight map	56
3.1	Nonino stacked image in R -band of VIMOS instrument	59
3.2	Magnitude difference between reference and 2009 catalogues for a <code>MAG_ZEROPOINT = 27.20</code> mag. Green lines are at 3σ from median value	60
3.3	<i>Left image:</i> Weight map for year 2009. <i>Right image:</i> Catalogues of year 2009 divided according image weight. In green the full catalogue, in red deep areas and in blue the most shallow areas	62
3.4	Ratio of sources selected as stars with three thresholds (<code>CLASS_STAR > 0.5</code> , <code>0.7</code> and <code>0.9</code>) for full, deep and faint catalogues	64
3.5	Ratio of sources selected as stars (<code>CLASS_STAR > 0.9</code>) for full and deep catalogue. In blue are underlined values for <code>SEEING_FWHM = 1.3''</code> and in gray a reference value of <code>SEEING_FWHM = 0.9''</code>	65
3.6	In blue are represented sources identified as stars (<code>CLASS_STAR > 0.9</code>) over the full catalogue (in gray). On the <i>left image</i> for <code>SEEING_FWHM = 0.9''</code> while on the <i>right</i> the one with <code>SEEING_FWHM = 1.3''</code>	65

3.7	Sources magnitude difference between 2009 and 2011 catalogues	67
3.8	Magnitude distributions for the 8 catalogues. The red dashed line indicates the instrumental magnitude limit for year 2009 used as reference.	69
3.9	Merged catalogue source magnitude distribution and instrumental limit at 26 mag	69
3.10	Histogram with the number of detections.	70
3.11	Sources with $ND \geq 4$ in blue over the full catalogue in gray	71
3.12	The magnitude Median Absolute Distribution (MAD) of catalogue sources as function of sources magnitude. The MAD spread increases for weak values.	72
3.13	Number of sources and MAD distribution for <code>obj_num=50</code>	75
3.14	Number of sources and MAD distribution for <code>obj_num = 150</code>	76
3.15	Number of sources and MAD distribution for <code>mag_step = 0.3 mag</code>	77
3.16	Number of sources and MAD distribution for <code>mag_step = 0.5 mag</code>	78
3.17	Number of sources and MAD distribution for <code>obj_num = 50</code> and <code>mag_step = 0.3 mag</code>	79
3.18	Variables on sky	80
4.1	Second frame of GIF file for variable 4190. The red dot highlights the magnitude measure (2009 in that case) on the light-curve and the corresponding cut-out is visualized on the right, with the year and the magnitude value on top.	83
4.2	Light-curve and cut-outs for source 8680 that I proposed as SN candidate	84
4.3	The four reference variability-selected AGNs catalogues on the sky over the complete object catalogue produced in this work (in gray)	86
4.4	This image, extracted from Giavalisco et al. 2004 paper, shows pointings overlap in the same field. This work is based on CDF-S pointing (black lines) while the GOODS South field is in blue.	88
4.5	Cross match result among full catalogue and reference catalogues. Variables detected in this work are represented as gold stars	92

4.6	Median Absolute Deviation (MAD) (top) and MAD* (bottom) distributions for my sample and for the other four catalogues analysed here. Green dots are for Trevese et al. 2008 , red cross for Villforth, Koekemoer, and Grogin 2010 , blue squares for Falocco et al. 2015 and black circles for Pouliasis et al. 2019	93
4.7	MAD as function of magnitudes with X-Ray sources highlighted	98
4.8	Magnitude distribution with X-Ray sources	99
4.9	Redshift distribution with X-Ray sources	101
4.10	Redshift vs Magnitude	102
4.11	Light-curve and images cut-outs for source 9795	104
4.12	Light-curve and images cut-outs for source 18162	105
4.13	Image extracted from Villforth, Sarajedini, and Koekemoer 2012 Figure 10: Multi-wavelength SED for source J033210.91-274414.9 (corresponding to 18162 in my catalogue), overlaid several of the best-fitting SEDs.	106
4.14	Light-curve and images cut-outs for source 15548 at $z = 3.797$	107
4.15	Light-curve and images cut-outs for source 15936 at $z = 3.343$	108
5.1	Median Absolute Deviation (MAD) distribution for of my sample compared with reference catalogues. Gold stars are variable sources I identified in my sample, green dots for Trevese et al. 2008 , red cross for Villforth, Koekemoer, and Grogin 2010 , blue squares for Falocco et al. 2015 and black circles for Pouliasis et al. 2019	112
5.2	The magnitude distribution as function of redshift. Sources with X-ray counterparts are highlighted in orange	113

Listings

2.1	Example of pipeline configuration file: main tag containing MasterFlat creation step parameters	37
2.2	Example of <i>main tag</i> for trail masking	48
2.3	Example of configuration tag for masking below a threshold	49
2.4	Example of configuration tag for sources extraction	50