Leibniz-Institut für Astrophysik Potsdam

Optical surveys of AGN and their host galaxies

Dissertation
zur Erlangung des akademischen Grades
“doctor rerum naturalium”
(Dr. rer. nat.)
in der Wissenschaftsdisziplin “Astrophysik”

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

Daria Dubinovska

Potsdam
4 Oktober 2012

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2013/6473/
URN urn:nbn:de:kobv:517-opus-64739
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-64739

Acknowledgments

These nearly 5 years of my life, which I spent working on my thesis at AIP gave me a priceless experience of working and becoming friends with wonderful people. If somebody would have told me 5 years ago, when I was defending my Master thesis in Kyiv, how my life is going to change in the next few years, I probably would not believe it. I have been to places where I have never imagined myself going, I learned couple of languages (and I'm sure I will learn at least one more), I met in person great people, whose names I saw in the schoolbooks back when I was a kid. Thus and so I would like to say a great Thank You to Lutz Wisotzki, my supervisor, who gave me such a chance. I'm very thankful to my family, who, of course, always supported me, and to my dear brother, who at the moment of a great concern told me that if I want to study Astronomy I should not listen to anybody and just do it. I want to say thank you to Vasyl Mykolayovych Ivchenko, the head of the astronomy and space-physics department of the KNU, who supported me when I applied for PhD positions, and to all people at the department who was passionately and patiently teaching me during 5 years. I want to say thanks to all my friends here at AIP and those who already went somewhere else, to Asmus, Gabi, Isabelle, Adriana, Ivan and Regina, Marija, Tanya, Oli and Katja, Mary, Carol, Andreas, Daniel, Berndt and Sebastian, Andreas K., Jakob, Cecilia and Sebastian, Manee, Aldo, Rob and Franziska, Ole, Valentina, Karleyne, Ada, Pierre, Axel, Uli, Corrado, Barry and many others, whom I probably forgot here now, who made my life here a pleasure. To my friends Olya and Yulia, for helping and supporting me during last 4 years. A special great thank you to Natasha, to whom I owe the greatest moment of feeling of being an astronomer while observing at La Silla. And finally, to Fran. I don't have enough words to tell you how grateful I am for all your help and support to me to finish this thesis, without you I would give up.

Contents

Abstract

This thesis rests on two large Active Galactic Nuclei (AGNs) surveys. The first survey deals with galaxies that host low-level AGNs (LLAGN) and aims at identifying such galaxies by quantifying their variability. While numerous studies have shown that AGNs can be variable at all wavelengths, the nature of the variability is still not well understood. Studying the properties of LLAGNs may help to understand better galaxy evolution, and how AGNs transit between active and inactive states.

In this thesis, we develop a method to extract variability properties of AGNs. Using multi-epoch deep photometric observations, we subtract the contribution of the host galaxy at each epoch to extract variability and estimate AGN accretion rates. This pipeline will be a powerful tool in connection with future deep surveys such as PANSTARS.

The second study in this thesis describes a survey of X-ray selected AGN hosts at redshifts $z > 1.5$ and compares them to quiescent galaxies. This survey aims at studying environments, sizes and morphologies of star-forming high-redshift AGN hosts in the COSMOS Survey at the epoch of peak AGN activity.

Between redshifts $1.5 < z < 3.8$, the COSMOS HST/ACS imaging probes the UV regime, where separating the AGN flux from its host galaxy is very challenging. Nevertheless, we successfully derived the structural properties of 249 AGN hosts using two-dimensional surface-brightness profile fitting with the GALFIT package. This is the largest sample of AGN hosts at redshift $z > 1.5$ to date. We analyzed the evolution of structural parameters of AGN and non-AGN host galaxies with redshift, and compared their disturbance rates to identify the more probable AGN triggering mechanism in the $43.5 < \log_{10} L_X < 45$ luminosity range. We also conducted mock AGN and quiescent galaxies observations to determine errors and corrections for the derived parameters.

We find that the size – absolute magnitude relations of AGN hosts and non-AGN galaxies are very similar, with estimated mean sizes in both samples decreasing by $\sim$ 50% between redshifts $z = 1.5$ and $z = 3.5$. Morphological classification of both active and quiescent galaxies shows that the majority of the AGN host galaxies are disc-dominated, with disturbance rates that are significantly lower than among the non-AGN galaxies. Such a finding suggests that Major Mergers are probably not responsible for triggering AGN accretion in most of these galaxies. Other secular mechanisms should therefore be responsible.

Zusammenfassung

Diese Arbeit basiert auf zwei grossen Durchmusterungen aktiver Galaxienkerne (engl., *active galactic nuclei, AGN*). Die erste Durchmusterungen hat zum Ziel, Muttergalaxien leuchtschwacher AGNs (sog. *low-level AGN* oder LLAGN) zu identifizieren und die Variabilität der AGN zu quantifizieren. Während zahlreiche Studien gezeigt haben, dass AGN bei allen Wellenlängen variabel sein können, ist die Natur der Varibilität noch immer nicht verstanden. Das Studium der Eigenschaften von LLAGN trägt zum Verständnis bei, wie sich Galaxien entwickeln und wie AGN zwischen aktiven und inaktiven Zuständen wechseln.

In dieser Arbeit entwickeln wir eine Methode, um die Eigenschaften der AGN-Variabilität zu messen. Mittels tiefer, photometrischer Beobachtungen zu mehreren Zeitpunkten subtrahieren wir den Beitrag einer Muttergalaxie, um die AGN-Variabilität zu bestimmen und Akkretionsraten des AGN abzuschätzen. Diese Methode wird im Zusammenhang mit zukünftigen, tiefen Durchmusterungen wie z.B. PANSTARRS ein mächtiges Werkzeug darstellen.

Die zweite grosse Durchmusterung behandelt nach ihrer Roentgen-Helligkeit ausgewählte AGN-Muttergalaxien mit einer Rotverschiebung von $z > 1.5$, um sie mit inaktiven Galaxien zu vergleichen. Ziel ist, bei sternbildenden, hochrotverschobenen AGN-Muttergalaxien aus der COSMOS-Durchmusterung die Umgebung, Grösse und Morphologie zum Zeitpunkt der stärksten AGN Aktivität zu studieren.

Die COSMOS-HST/ACS-Photometrie untersucht den ultravioletten Spektralbereich von Muttergalaxien bei einer Rotverschiebung von $1.5 < z < 3.8$, wo es sehr schwierig ist, den Flussbeitrag des AGN von dem der Muttergalaxie zu trennen. Dennoch gelang es uns, mittels zweidimensionaler Modellierung der Flächenhelligkeitprofile Struktureigenschaften für 249 AGN-Muttergalaxien aus der COSMOS-Durchmusterung abzuleiten. Dies ist für AGN-Muttergalaxien mit einer Rotverschiebung $z > 1.5$ die bislang grösste Stichprobe. Wir haben die Entwicklung der Strukturparameter mit der Rotverschiebung von AGN-Muttergalaxien und Galaxien ohne aktiven Kern analysiert und die Häufigkeit morphologischer Besonderheiten bei diesen beiden Gruppen miteinander verglichen, um die wahrscheinlichste Ursache für AGN-Aktivität für Galaxien im Roentgen-Leuchtkraft-Bereich von $43.5 < \log_{10} L_X < 45$ zu bestimmen. Um Fehler und Korrekturfaktoren für die bestimmten Parameter zu erhalten, wurden simulierte AGN-Muttergalaxien und inaktive Galaxien untersucht.

Als Ergebnis finden wir, dass AGN-Muttergalaxien und Galaxien ohne aktiven Kern eine sehr ähnliche Relation von Galaxiengrösse zu absoluter Helligkeit aufweisen, wobei die abgeschätzte, mittlere Grösse zwischen Rotverschiebungen von $z = 1.5$ und $z = 3.5$ in beiden untersuchten Gruppen um $\sim 50\%$ abnimmt. Die morphologische Klassifikation sowohl aktiver als auch inaktiver Galaxien zeigt, dass die Mehrheit der AGN-Muttergalaxien scheibendominiert ist und signifikant weniger morphologische Anomalien aufweist als inaktive Galaxien. Dieses Resultat weist darauf hin, dass in den meisten AGN-Muttergalaxien sog., *Major Merg-*

ers, also die Verschmelzung zweier etwa gleich grosser Galaxien, wahrscheinlich nicht Akkretion auf den AGN auslösen, sondern stattdessen andere, sekuläre Mechanismen dafür verantwortlich sind.

Introduction

This Thesis consists of two parts, describing two projects I have been working on during my PhD studies. Both parts describe the results of two large surveys of active galactic nuclei (AGN) we have carried out, which are regarding properties of AGNs and their host galaxies: Part I describes the attempt to measure optical variability of X-ray selected AGNs in order to constrain a method that could be employed as a tool to search for marginally active AGNs, and Part II is dedicated to a large survey of high-redshift AGN host galaxies. This short introduction will provide a very brief outline of the discovery history, together with known facts and assumptions that are made in order to explain the current state of the art of AGN astronomy, and briefly summarizing the aspects that will be addressed later in this Thesis.

The very first optical spectrum of an active galaxy was taken back in 1908 by E. A. Fath at the Lick Observatory (Fath, 1909). He observed NGC 1068 and noted the presence of strong emission lines in the spectrum. Only in 1943 C. Seyfert realized that there is a class of galaxies whose spectra have high-excitation nuclear emission lines, whose width implied gas velocities of up to 8500 km/s (Seyfert, 1943). This class of galaxies did not receive much attention until more than a decade later in 1955, when radio emission of NGC 1068 and NGC 1275 was detected. After that, in 1959 L. Woltjer attempted to explain the nature of broad emission lines observed in Seyfert's galaxies, and estimate the sizes and masses of regions where those lines were formed (Woltjer, 1959). He pointed out the outstanding compactness of these regions, showing that ~10% of the mass of such a galaxy is compressed within its central ~100 pc. The physical nature behind such enormous amounts of energy produced in such a compact volume remained unanswered.

In the late 1950s radio surveys discovered a new class of very powerful radio sources – quasars. In 1960 A. Sandage and T. Matthews identified strong radio source 3C48 with a point-source object, a faint blue star of 16 mag. However, after its spectrum was taken, it became clear it was not a star since it showed broad emission lines. The spectrum remained unexplained until 1963 (Matthews & Sandage, 1963). In 1962 radio source 3C273, observed by C. Hazard (Hazard et al., 1963) and separately by J. Bolton (Bolton et al., 1963), via measurements of lunar occultation was identified with another optical point-source. Spectroscopic observations of this object by M. Schmidt in 1963 revealed the presence of similar broad emission lines as detected in 3C48 (Schmidt, 1963). Schmidt realized that these lines were nothing other than redshifted Hydrogen Balmer-series and Mg II λ2798 emission lines. He calculated the emitting object to be receding at an incredible speed of ~47 000 km/s equivalent to redshift $z = 0.158$. This redshift was about 10 times higher than the most distant Seyfert galaxy known at that time. The cosmological

interpretation of this redshift implied enormous luminosity of more than 100 times higher than the luminosity of Milky Way or M31. This discovery started the new era in astronomy. Many new powerful radio sources soon after the first discoveries were identified with point-like optical counterparts. Initially, they were referred as quasi-stellar radio sources, or shorter quasars, but later it became clear that not all quasars have radio emission, so the new name QSO (quasi-stellar object) was assigned to generalize both classes of radio-loud and radio-quiet objects. [1]

Since QSOs were discovered, it was clear that these objects are extraordinary. Not only were they the most distant objects known at that time, but also their remarkable luminosities of 10^{46} erg s^{-1} did not have analogs in the known nearby universe, suggesting the existence of an unknown exotic central engine to generate such powerful emission that was observed in these objects.

The first suggestion that Black Holes (BH) are involved in producing emission detected in QSOs was proposed by Zel'dovich & Novikov (1964). Nowadays, it is commonly accepted that all massive galaxies, or at least those that have a bulge component, have a Supermassive Black Hole (SMBH) in their centers (Kormendy & Richstone, 1995; Magorrian et al., 1998; Ferrarese & Ford, 2005) and that the evolution of this central SMBH and its host galaxy seems to be tightly coupled. Indeed, the existence of several scaling relations between the central SMBH and galactic bulge, such as the M_{BH}-L_{bulge}, M_{BH}-M_{bulge}, M_{BH}-σ relations (Ferrarese & Merritt, 2000; Gebhardt et al., 2000, i.e.) suggests that the growth of both components should occur at the same evolutionary stages. It has been suggested that the most active SMBH growth happens during the episodes of quasar activity (Yu & Tremaine, 2002), and if this is the case such episodes of activity should occur for all galaxies a number of times throughout their evolution. An alternative explanation suggests that the BH mass – bulge mass correlation is a result of statistical assembly of a number of Major Mergers (MM) that a galaxy has been through. However, this case does not seem very appealing: Peng (2007) shows that to reach the observed tightness of correlations between the SMBH and galactic bulge from initially uncorrelated relations, a galaxy would need to go through at least 10 MM. In the context of numerous observations and recent cosmological simulations such a high number appears highly unlikely, therefore suggesting that the correlations between BH and bulge properties do not have statistical origin. Periods of quasar activity seem to be a common stage of galactic evolution, but the mechanism which switches galaxies between inactive and active states is still under debate. Attempting to explain the observed distribution of the present day BH masses, one should look back at the epoch of peak AGN accretion, that takes place around redshift $z \sim 2$.

The current cosmological model, ΛCDM, predicts a "bottom-up" hierarchical structure formation. In such model of the Universe small galaxies form first and later they assemble into more massive ones (White & Rees, 1978). Naturally, the

[1] Short historical information is summarized from " An introduction to active galactic nuclei" by Peterson (1997)

most crucial role in the formation of a present day structures is then to be played by MM (Kauffmann et al., 1993; Cole et al., 2000; Somerville et al., 2001; Bell et al., 2007; Robaina et al., 2010). Various cosmological simulations are successful at reproducing scenarios in which AGN activity and a burst of star-formation are initiated when two or more galaxies of comparable mass merge (Springel et al., 2005; Di Matteo et al., 2005; Hopkins et al., 2006, 2005, 2008). Within a galaxy, a merger induces gravitational torques, which deprive gas of angular momentum, leading to a large-scale gas inflow to the center of a galaxy, that subsequently will feed nuclear star-formation and BH activity (Mihos & Hernquist, 1996; Hernquist, 1989; Barnes & Hernquist, 1992; Springel et al., 2005; Cox et al., 2006, 2008). These simulations support a number of observations that appeal to MM as the AGN triggering mechanism (Bahcall et al., 1997; Canalizo & Stockton, 2001; Zakamska et al., 2006; Urrutia et al., 2008; Silverman et al., 2011). Discovery of Ultraluminous Infrared Galaxies (ULIRGs) additionally had strengthened the merger-AGN connection, since 95% of these galaxies are found in an interacting state, while many of them have a detectable AGN (Sanders et al., 1988a,b; Sanders & Mirabel, 1996; Surace et al., 1998) .

However, studies of host galaxies of medium luminosity AGNs of $L_X = 10^{42} - 10^{44}$ erg s^{-1}, carried out at different redshift ranges, reveal results which contradict this theory, showing i) a significant prevalence of disc-dominated morphologies among the AGN hosts, which simulations show to be very hard to preserve during MMs and ii) a lack of severely disturbed galaxies in comparison to non-active galaxies (Dunlop et al., 2003; Sánchez et al., 2004; Grogin et al., 2005; Gabor et al., 2009; Cisternas et al., 2011; Schawinski et al., 2011, 2012; Kocevski et al., 2012), suggesting some different mechanisms of transporting gas to the galactic center, such as secular evolution and minor mergers. Kormendy & Kennicutt (2004), for example, had reviewed how secular processes in a galaxy provide transportation of gas from the outskirts of a galaxy to its center. In this case, gas feeds star-formation, that builds up a pseudo-bulge and may as well feed the central SMBH in a galaxy, turning once quiescent nuclei into an active AGN. However, it remains under debate which mechanism to which extend prevails at each luminosity range.

What happens after the AGN is triggered? As we just briefly mentioned, observations show that the star-formation and the SMBH growth are tightly correlated with each other and therefore probably evolve together. There are several ways this co-evolution can take place, as summarized by Zheng et al. (2009) from studies by Silk & Rees (1998); Kauffmann & Haehnelt (2000); Wyithe & Loeb (2003); Granato et al. (2004); Merloni et al. (2004); Haiman et al. (2004); Cattaneo et al. (2005); Fontanot et al. (2006): in the strongest case of co-evolution, bulge and BH growth occurs at the same time, i.e. whenever the star-formation starts, SMBH accretion starts as well, consuming a fraction of the gas available in the central regions of a galaxy. This scenario is unlikely due to a large number of galaxies that are found to be actively forming stars but show no signatures of nuclear activity. The second scenario suggests a time delay (or different time scales) between the onset of star-formation and the start of AGN accretion. This approach is imple-

mented in simulations by Di Matteo et al. (2005) and Hopkins et al. (2005, 2006). Simulations had shown that at the first passage during an interaction only a burst of star-formation is triggered, while the BH accretion starts later, close to the final stage just before the coalescence. The third scenario of co-evolution is the self-regulated growth of galactic components. As implied by the name of this model, the growth of one component regulates the growth of the other even though they occur in different events (Silk & Rees, 1998; Murray et al., 2005; Di Matteo et al., 2005).

The variety of the spectral properties of AGNs observed can be explained using AGN Unification Theory. According to it, the observed taxonomy of active galactic nuclei can be explained as the same central engine, observed from the different viewing angles (Antonucci & Miller, 1985; Antonucci, 1993; Urry & Padovani, 1995) as shown in Figure 1. The central engine consists of a SMBH and its accretion disk, surrounded by a large optically thick torus. In the plane, perpendicular to the accretion disc plane, there is a so-called broad line region (BLR). This region is formed by hot and fast moving gas, which is responsible for the formation of broad emission lines in the AGN spectra. Outside of the BLR is a narrow line region (NLR), which consists of gas, which forms narrow emission lines in the spectra. According to the Unification scheme, the AGN is detected as radio loud or radio quiet depending on the presence of a jet. Division into classes depends on the viewing angle and AGN luminosity. For instance, if the transition between two types of Seyfert galaxies is defined by the presence or absence of high-ionization broad emission lines, the difference between Seyferts and QSOs is defined simply by a luminosity cut at $M_B > -21.51 + 5\log h_0$, where $h_0 = 0.7$, due originally to Schmidt & Green (1983). The X-ray luminosity function of AGNs is luminosity dependent, with the peak of space density of AGNs shifting towards higher redshifts with the increasing luminosity: from $z \sim 0.7$ for local Seyfert galaxies with $L_X \sim 10^{44}$ erg s^{-1} to $z \sim 2$ for bright $L_X > 10^{44.5}$ erg s^{-1} (Ueda et al., 2003; Hasinger et al., 2005; La Franca et al., 2005; Silverman et al., 2008; Ebrero et al., 2009; Yencho et al., 2009).

Throughout this Thesis we will be operating with two X-ray selected samples of AGNs, which both have X-ray luminosities of $L_X > 10^{42}$ erg s^{-1}, therefore it is almost certain this emission has an AGN origin and can not be confused with the X-ray emission from star-forming regions. Each sample has both Type I and Type II AGNs, which determines the degree of obscuration: if the determined neutral hydrogen column densities in the nuclei exceeds the value of $N_H > 10^{22}$ cm^{-2}, which can be translated to an X-ray hardness ratio (HR) cut at $HR = -0.2$, then the nucleus is obscured and assigned to the Type II (Wang et al., 2004). This division into types can not be explained by the geometrical unification model, but instead should rather be considered as objects observed at different evolutionary stages. This hypothesis is supported by a number of X-ray surveys, which show that obscuration/absorption is more common at lower luminosities (Ueda et al., 2003; La Franca et al., 2005; Hasinger, 2008), and, possibly, at high-redshifts (La Franca et al., 2005; Hasinger, 2008; Ballantyne et al., 2006; Treister & Urry, 2006).

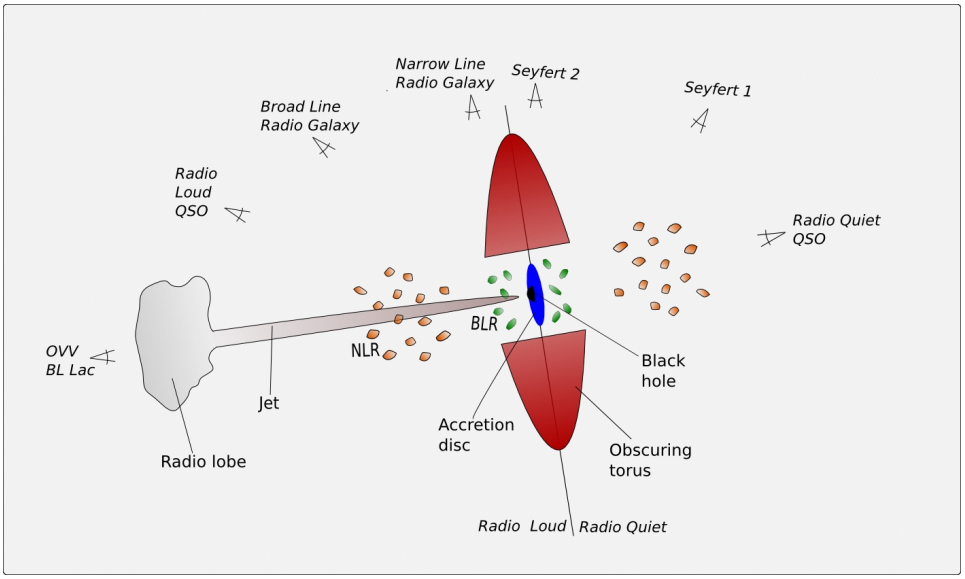

Figure 1: AGN unification scheme: how the observed type of the AGN changes depending on the viewing angle. Elaborated by Ioanna Arka.

Cycles of AGN activity can probably occur one or several times during a lifetime of a galaxy. But what happens when nuclear activity is shut down? Galaxies spend most of their lifetime in a quiescent state. It appears natural that every galaxy should have a reservoir of gas available, produced during regular cycles of stellar evolution. In that case, if we assume that all galaxies that have bulges host SMBHs, some low level accretion seems to be unavoidable (Soria et al., 2006a,b). However this accretion is hard to detect and quantify since it does not manifest itself with the signatures characteristic of active galaxies. Most probably, at this stage galaxies switch to a different accretion mode, which has lower radiative efficiency of just a few percent of the Eddington's rate, the so-called radiatively inefficient accretion flow, RIAF (Yuan, 2007). This assumption comes from the necessity to explain the absence of the big blue bump (BBB) in the SED of low-luminosity AGNs (LLAGNs). The BBB has a thermal origin, produced by the geometrically thin, optically thick accretion disc, and is characteristic of the radiatively efficient accretion mode observed in AGNs (Shakura & Sunyaev, 1973). This mechanism is currently favoured to explain the SEDs of AGNs.

There are a number of signatures that can point out a presence of weak activity in an otherwise quiescent galaxy: weak radio, UV and X-ray emission, or emission lines with ratios that are not characteristic for stellar excitation. However, all these signatures for a nearly quiescent BH will naturally be weak and very hard to detect. They can be easily confused with those of stellar origin or obscured by dust. Emission line ratios give a rather reliable result, however detecting weak and narrow emission lines characteristic to low-level activity requires long integration times and high spectral and spatial resolution. Measured emission

line ratios of such LLAGNs highlight the existence of a separate class of active nuclei, which is probably the most common class of low-level AGNs, called Low Ionization Nuclear Emission-line Regions (LINERs) (Heckman, 1980; Ho et al., 1997; Kauffmann et al., 2003; Ho, 2008). These objects occupy a distinct region in the "Baldwin, Phillips & Terlevich" (BPT) diagram (Baldwin et al., 1981), and are separated from the more active classes of Seyfert galaxies (see Ho (2008) for a review). Numerous surveys had found low-level activity to be very common in the local galaxies, with most of them being LINERs (Stauffer, 1982; Ho et al., 1996). Therefore it is probably fair to conclude that at higher redshifts galaxies which appear as quiescent may also have such low-level SMBH accretion.

Probably, one of the most effective ways to detect the presence of a LLAGN within a photometric study is variability. It is a distinctive signature of the presence of an AGN, because the emission of the integrated stellar population does not vary in time. Moreover, variability becomes more pronounced with decreasing luminosity of an AGN (Trevese et al., 1994; Wold et al., 2007). If an efficient and fast pipeline to search for such variability existed, one could apply it to the data obtained during repeated deep photometric campaigns (such as PANSATRS) and prove (or disprove) the ubiquity of low-level activity in quiescent galaxies.

Part I
Variability in the Low-level Active Galactic Nuclei

1 Introduction and motivation

There is growing evidence that SMBHs are present in all galaxies, or at least those that have a bulge component. Moreover the growth of the BH and the galactic bulge it resides in, seems to be coupled, which is evident from the existence of scaling relations between the SMBH mass and several properties of the galactic bulge, such as its mass, luminosity and stellar velocity dispersion (Kormendy & Richstone, 1995; Magorrian et al., 1998; Gebhardt et al., 2000; Ferrarese & Merritt, 2000; Tremaine et al., 2002; Marconi & Hunt, 2003; Häring & Rix, 2004). It is now commonly accepted that every galaxy went through an active phase at least once in their lifetime (Lynden-Bell, 1969; Soltan, 1982; Richstone et al., 1998). The most intensive growth of the SMBH mass probably happens during the brightest phase of such activity (Soltan, 1982; Yu & Tremaine, 2002). But these episodes of activity are short and last at most a few hundreds megayears and the remaining 99% of their lifetime galaxies spend in a quiescent state. Surprisingly, there are very few details known about this stage of galaxy evolution.

As we mentioned, every galaxy that had undergone recent star-formation or still forms stars should have a reservoir of gas, that is produced during the normal stellar life cycle. A fraction of this gas can be consumed by the central SMBH, causing some residual weak nuclear activity. There are implications that such accretion probably happens in a different accretion mode than the radiatively-efficient accretion in quasars due to the absence of the big blue bump in the SED of those weakly accreting galaxies. Such low-level nuclear activity is found to be present in most local galaxies. It can manifests itself with a presence of hard X-ray emission, radio emission or weak low-ionization narrow emission lines. Due to their weakness, these signatures, are hard to detect and they require long integration times and high spectral resolution (Maoz, 2007). But probably the most efficient way to identify not only a low-level AGN (LLAGN), but any kind of active nucleus, is to detect variability of its continuum emission. It has been shown by numerous studies that all AGNs are variable at all wavelengths, when observed for long enough (Ptak et al., 1998; Grupe et al., 2001; Stalin et al., 2004; Hovatta et al., 2008). Since the integrated stellar population does not vary in time, presence of even weak variation of continuum emission unavoidably indicates the presence of AGN accretion. The physics of such variability is still poorly understood, but the prevailing hypothesis states that some sort of accretion disc instabilities are responsible. For LLAGNs

such a method can be very useful because the variability amplitude grows with decreasing AGN luminosity, and these changes in flux are more easy to detect than other signatures of weak AGN accretion (Trevese et al., 1994; Wold et al., 2007).

1.1 AGN selection techniques

There is a variety of techniques to select AGNs. Among them probably the most used are color selection (Sandage, 1973; Croom et al., 2004, 2009), X-ray selection (Hasinger et al., 2005), radio emission (Griffith & Stern, 2010) and variability. All of these are efficient at finding AGNs that possess some common properties. Thus through evaluation of resulting samples one can establish which selection effects dominate which technique.

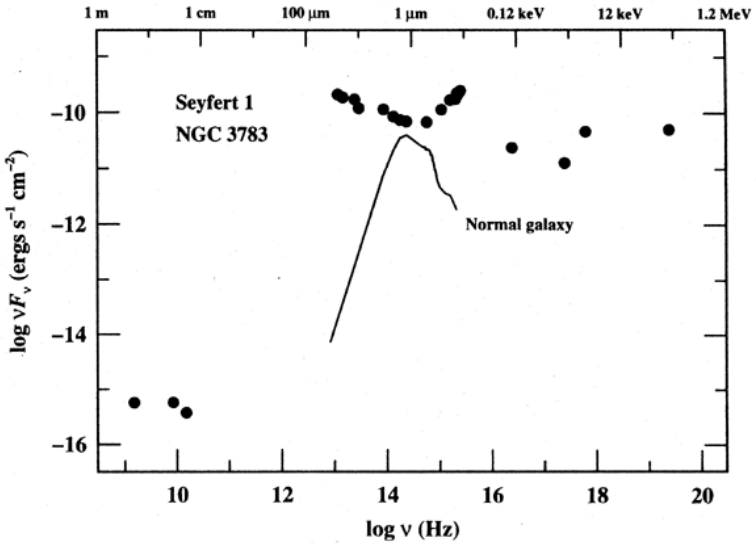

Figure 2: The comparison of SEDs of a Seyfert I and a quiescent galaxy. Figure taken from the book Peterson (1997).

The SEDs of a QSO and a normal inactive galaxy, shown in Figure 2, have a number of significant differences, which can be used to distinguish them employing minimum observational effort. This is what X-ray, radio and IR AGN selection methods are based on. Observing a galaxy in a certain spectral regime, where stellar emission makes negligible input into its SED, allows to efficiently distinguish the presence of any emission from the SMBH accretion. Figure 3 shows the detailed structure of an active galaxy's SED, with all the characteristic features marked on the plot. The characteristic minimum around 1μm is the region of an overlap between hot thermal emission of the AGN accretion disc (T~100 000 K), the BBB, and the cool thermal spectrum due to emission by warm dust (T~2000 K), often referred to as the "IR bump". In the X-ray range, some AGN SEDs show a sudden drop of emission towards higher energies, called the "soft X-ray excess",

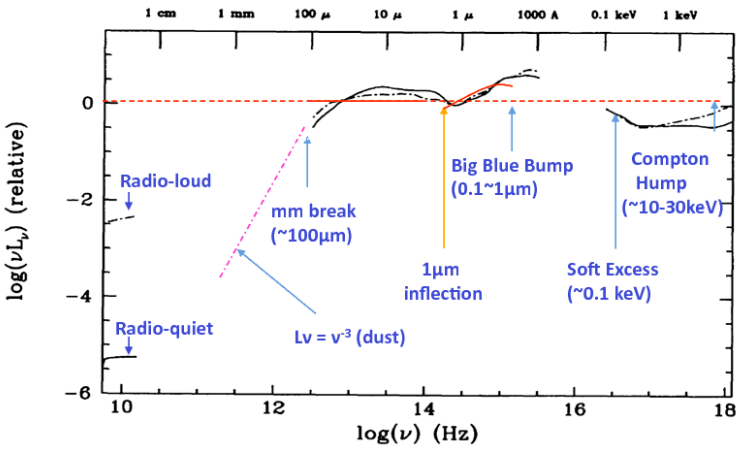

Figure 3: The SED of a QSO with all spectral features marked out. Taken from Elvis et al. (1994).

which is the high energy end of the BBB. Depending on the presence of jets, AGNs can also be detected by their radio emission, however a 90% fraction of AGNs are radio-quiet.

Another method that employs distinctions in the SEDs of active and inactive galaxies to isolate QSOs is color selection. It utilizes UV and optical bands photometry in order to construct the color diagrams. This method is efficient mainly for the point-like sources, where the nucleus emission dominates over the host galaxy emission. QSOs can be distinguished from galaxies and stars by their exceptional blue colors in all color diagrams, however the greatest separation between AGNs and stellar loci is in the *ugr* color plane, where the *u-g* color is dominated by the UV excess of quasars[1]. This selection method remains efficient until redshift z~2, where intervening absorption from the Lyα forest starts to play significant role. The NIR color selection is efficient both for the reddened QSOs and the UV bright (Maddox et al., 2012; Glikman et al., 2012) thus it is efficient at all redshifts.

In Figure 4 the color selection diagrams, constructed for the Sloan Digital Sky Survey (SDSS) QSOs selection, are shown. In the case of SDSS, QSOs were targeted spectroscopically after being isolated as outliers from the stellar locus in the *(u-g, g-r, r-i)* (denoted ugri) and *(g-r, r-i, i-z)* (denoted griz) color cubes. Figure 5 shows the loci of various types of objects in the GALEX-SDSS 2-color plane. There is some overlap between regions occupied by point sources and extended sources, as well as between stars and QSOs inside the point-source region. To confirm the quasar nature of the color-selected sources they have to be observed spectroscopically.

In the case of Seyfert galaxies or LLAGNs, the SED of those is strongly contaminated or even dominated by the emission of the host galaxy. Therefore color

[1]This does not, however, apply to the dust reddened QSOs

selection becomes much less efficient and other selection methods are used to select those.

One of the most efficient methods of selecting AGNs is, as we said before, X-ray emission (Barger et al., 2003; Hasinger et al., 2005; Brusa et al., 2010). X-rays have very high penetrative power and using a telescope with a high sensitivity like Chandra or XMM-Newton, one can detect AGNs in the luminosity ranges of 10^{42} – 10^{46} ergs s^{-1} out to high redshifts and up to obscuring neutral hydrogen column densities of $N \sim 10^{24}$ cm^{-2}.

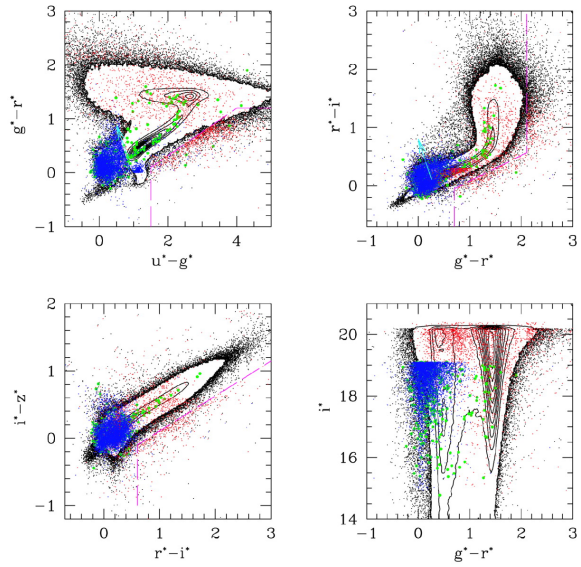

Figure 4: SDSS QSO target selection. Black points and contours belong to the stellar locus. Blue points are ugri-selected QSO candidates, red points are griz-selected quasar candidates, green points are Radio selected QSOs (FIRST survey). The light blue lines show the blue end of the ugri selection cut. The magenta lines show part of the three different high-redshift QSO inclusion regions. Taken from Richards et al. (2002).

1.2 Existing variability studies

Since variability is a common feature shared by all types of AGNs, it is used to identify active galaxies. However, the nature of variability may not only be due to nuclear activity. There are several physical processes, which may be responsible for periodical changes in continuum emission: (1) the accretion disc instabilities (Rees, 1984; Kawaguchi et al., 1998; Pereyra et al., 2006), which may lead to large-scale changes of the amount of in-falling material that can be responsible for variability of AGN emission (Hopkins et al., 2006, 2008); (2) the change in flux may indicate the onset of a nuclear starburst in a host galaxy (Aretxaga et al., 1997) however in that case variability is wavelength-dependent; (3) another cause for the continuum variability can be microlensing by an intervening mass concentration (Hawkins, 1996; Zackrisson et al., 2003); (4) finally, multiple SN explosion can

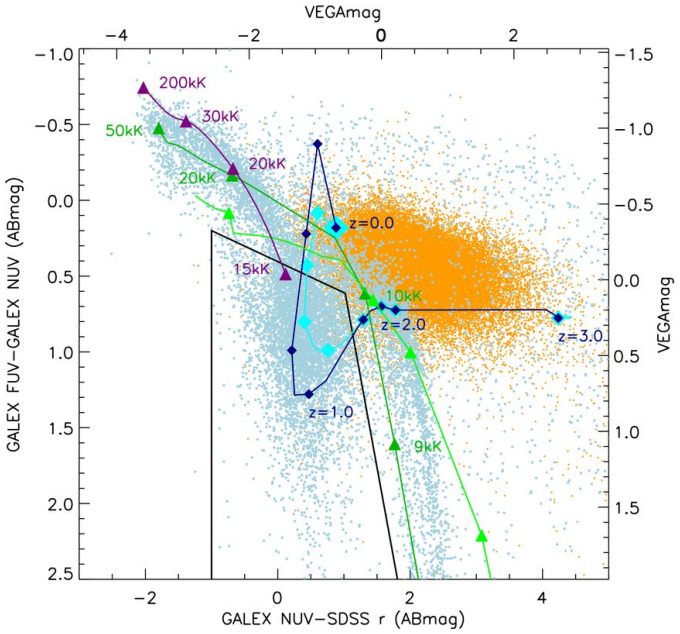

Figure 5: Objects from the combined SDSS-GALEX 2-color plane, from the catalog of Bianchi et al. (2010), showing a color selected sample of QSOs. The extended galaxies are shown in orange and occupy a distinct area separate from most of the point sources, denoted in light blue. The area occupied by point sources is split in two loci, loci of stars and QSOs, with their templates shown. Templates for stars of main sequence are dark green, supergiants in light green, and white dwarfs shown in magenta. Also two types of QSOs templates are shown: for normal (cyan) and three times (dark blue) enhanced *Lyα* emission. This sample can be contaminated by cooler white dwarfs. Plot is taken from Hutchings & Bianchi (2010).

cause variations of the continuum emission (Terlevich et al., 1992). Depending on the nature of variability, its timescales may vary from days to years with amplitudes typically $\geq$ 10% (Giveon et al., 1999; Vanden Berk et al., 2004; Rengstorf et al., 2004; Sesar et al., 2007; Bramich et al., 2008; Schmidt et al., 2010; Palanque-Delabrouille et al., 2011). In order to constrain whether variability detected in the unresolved galactic nucleus indeed originates from an AGN it commonly requires spectroscopic or X-ray (Radio/IR) follow up to confirm its nature.

In the context of the low-level AGN variability survey we were carrying out, we would like to point out two surveys, which were done in the Chandra Deep Field South (CDFS) and Chandra Deep Field North (CDFN), that aimed at creating complete samples of variability selected AGNs in those fields. Since our survey was to be carried out in the same fields we, hypothetically, could cross-check our results with the final samples from these studies to probe the completeness of these samples.

In Trevese et al. (2008) the authors report on their analysis of data collected for the STRESS Supernova Survey in CDFS, observed with the wide field channel (WFC) of the wide-field imager (WFI) on the 2.2-meter telescope at La Silla, Chile. The aim of the authors was to compile a complete catalog of variable sources in the CDFS field by performing aperture photometry of every source detected in the field, observed at a minimum during 5 out of 8 epochs of observations that were available. The final number of sources suitable for variability measurements, that passed the 5-epoch threshold requirement, consisted of 7267 objects. For each of these objects the authors constructed a low-resolution light-curve, using measurements of flux in a fixed size 4-pixels aperture (which corresponds to $r \sim 1''$). The characteristic mean magnitude m and its variance σ^2 for each of k objects then were computed as:

$$\overline{m}_k = \frac{1}{N_{epo}} \sum_{i=1}^{N_{epo}} m_k^i$$

$$\sigma_k = [\frac{1}{N_{epo}} \sum_{i=1}^{N_{epo}} (m_k - \overline{m}_k)^2]^{\frac{1}{2}},$$

where m is the aperture magnitude measurement for an object k, i denotes the epoch, and σ is the *RMS* of the measurements, used to quantify the variability. To distinguish AGN variability from the measurements' uncertainties due to noise, the authors apply a $3\Sigma_\sigma$ threshold defined as:

$$\Sigma_\sigma(V) = \langle [\sigma - s(V)]^2 \rangle^{\frac{1}{2}}$$

$$s(V) = \langle \sigma \rangle,$$

where σ(V) represents the average RMS of the V-band magnitude measurements.

The results of this survey are shown in Figure 6. The authors found significant variability, which exceeded the $3\Sigma_\sigma(V)$ detection threshold, in 132 AGN candidates, shown on the plot above the solid line. It is notable that not all of the

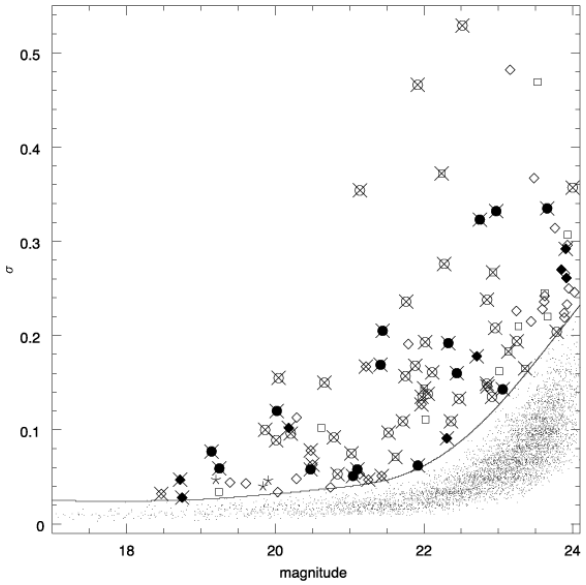

Figure 6: Variability measurements for V-band detected objects in the CDFS for the sample of Trevese et al. (2008). Objects below the solid line, which indicates the $3\Sigma_\sigma$ threshold for reliable AGN variability detection, are considered as non-variable. Above the line, variable objects are indicated as follows: filled symbols are objects with spectroscopic redshift; empty symbols are objects without spectroscopic redshift; crosses are objects with an X-ray counterpart; circles, diamonds and stars are objects classified by COMBO-17 Photometric Survey as QSOs, galaxies and stars, respectively; squares are objects outside the COMBO-17 field. Figure is taken from Trevese et al. (2008).

variable sources were previously classified as AGNs: some of them do not have a detected X-ray counterpart, and/or were previously classified as stars or quiescent galaxies in COMBO-17. To determine the nature of variability in these sources, they required a spectroscopic follow-up, which was subsequently carried out by Boutsia et al. (2009). The authors confirmed the AGN nature of the majority of the objects. Therefore, in our X-ray pre-selected sample, the selection can be subjected to a bias, missing objects that were not detected in X-rays or misclassified in COMBO-17. However, we believe the variability selected sample of Trevese et al. (2008) may be subjected to incompleteness as well, because of the method used for quantifying variability. Fixed size aperture photometry for AGNs of different luminosities would unavoidably lead to a different level of host galaxy contribution to the aperture flux with respect to the AGN flux, and therefore it would result in a lower sensitivity and larger errors for the detected variability measurements.

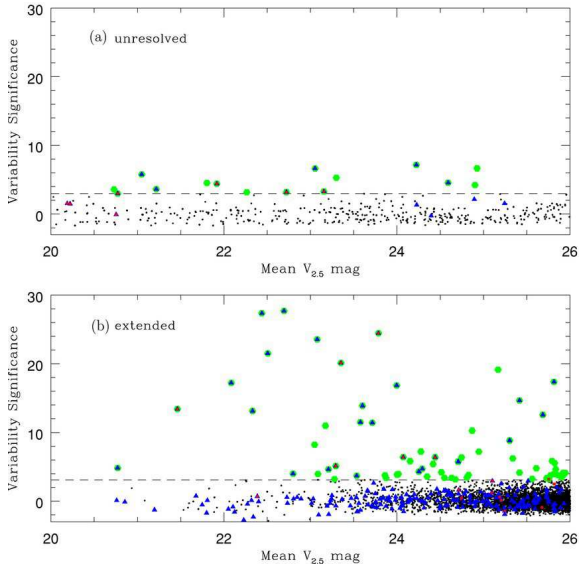

Figure 7: AGN Variability measurements as a function of the apparent V-band magnitude, measured separately for point sources and extended objects. The dashed line represents the threshold for variability significance. In green are shown detections, that have strong variability above 3.1σ threshold for extended sources and 3.0σ for unresolved point-sources. Blue triangles indicate X-ray sources from 2Ms Chandra surveys. Red squares represent Spitzer mid-IR power-law sources. Plot is taken from Sarajedini et al. (2011).

Another variability study we would like to bring the reader's attention to, is reported by Sarajedini et al. (2011). It was conducted in the GOODS North and South fields (that are part of the CDFN and CDFS correspondingly) , using data acquired with Hubble Space Telescope (HST) in the period from 2008 until 2011. This data were collected the Advanced Camera for Surveys (ACS) of HST during 5 epochs of observations 2-3 months apart from each other. They claim detection of variability in 85 sources in both fields, using the same method of aperture photometry as in Trevese et al. (2008). To compute magnitudes, they used the 5-pixel

aperture (which corresponds to angular diameter of $0.15''$), arguing that smaller aperture sizes are more sensitive to PSF instabilities. They match their variability-selected sample to the X-ray catalog of Alexander et al. (2003) and find that 51% of their variable sources have no X-ray detection. The results of this survey are shown in Figure 7.

We claim that both of these catalogs are subjected to incompleteness. For instance, the GOODS-S field, where a part of the Sarajedini et al. (2011) survey was conducted, has 42 variable AGN candidates detected. While this field encompasses $\sim \frac{1}{8}$ fraction of the CDFS field of Trevese et al. (2008), the final variability samples in these two studies have only 8 objects in common. This can happen as a result of an overestimated threshold to reject unreliable detections in both studies. Also, it has to be taken into account that the spatial resolution of ACS is ~ 8 times higher than that of WFI, which results in better sensitivity of AGN variability measurements. However, the observations with HST/ACS have a disadvantage of shorter time sampling to compare with Trevese et al. (2008) survey.

1.3 Motivation

The primary motivation of conducting a survey of variable LLAGNs is to answer the question if this state of nuclear activity indeed occurs during the transition of a galaxy from active to quiescent states. To attempt to give an answer to this question we need to explore how common are low-level active nuclei in galaxies, in which type of galaxies they occur and if these galaxies have any common characteristic features that may be linked with the presence of low-level AGN. Variability selection can be useful to detect weak AGN activity at the level of just a few percent of Eddington's accretion rate. Searching for variability in a pre-selected X-ray detected sample gives an advantage of a confidence of an AGN detection, since neither weak X-ray emission nor variability by themselves are unique to the presence of an AGN. We aimed to address these issues by conducting a deep photometric survey of variable X-ray detected sources in the CDFS and CDFN, using archival CDFS data and performing new observations of CDFN with the Large Binocular Telescope (LBT), in the USA.

To apply variability as a tool to find LLAGNs, we primarily care to detect variability in the X-ray selected galactic nuclei. The measured amplitude of variability can be converted into a lower limit of the non-stellar flux, which in turn can be converted or extrapolated into the rest frame UV flux, depending on the band in which observations were taken, and redshifts of galaxies that host these variable nuclei. Together with the X-ray luminosity, the AGN UV luminosity estimate would allow to constrain the SEDs of the variability detected AGNs and can be converted into an SMBH accretion rate. Such a method should provide a better tool to isolate LLAGNs than a weak X-ray luminosity by itself, since a single X-ray detection can not provide a reliable measurement of accretion rates.

With this survey we were aiming to develop a universal pipeline, that without major adjustments could be applied to data, obtained with different telescopes, and

return high-precision variability amplitude measurements. We wanted to develop a method with high sensitivity and precision, which would yield good estimates even for the faintest AGNs, avoiding the influence of their host galaxies. The algorithm we embodied in our pipeline, exploits high-resolution HST observations of each galaxy in the sample, to create its 2-dimensional brightness profile. Degrading the high-resolution model to the resolution of ground-based observations, we can then subtract a low-resolution model from ground-based observations of a galaxy at every epoch, and use the residual flux to quantify variability. This method has several valuable advantages that can play a final decisive role in detecting marginal variability of an AGN: (i) it is less contaminated by host galaxy emission, (ii) it allows to eliminate the error of measurements due to PSF instability by taking into account the time and/or position dependent PSF model iII) subject to availability of high-resolution data, our pipeline can be easily adjustable to perform measurements on data, observed with various ground-based telescopes (in case no high-resolution data is available, the pipeline can as well be supplied with ground-based data instead, which then will be used as a reference point).

2 Variability Samples

We planned to conduct our survey in two fields, Chandra Deep Field South and North (CDFS and CDFN). Our primary interest though would be CDFN, since we counted to obtain multi-epoch deep photometric observations of this field with the Large Binocular Camera (LBC) camera of the LBT, Arizona, USA. To get sufficient sampling to measure variability, we requested repetitive visits of the field every 6 months starting in the latter half of 2008. This field also has deep HST coverage, with data publicly available in the archive. In order to set up a pipeline to process future data and constrain the method, we collected archival observations of CDFS with the WFI. These observations have sufficient depth and time-sampling, as well as high-resolution ACS observations available from the GEMS Survey. The data were kindly provided by L. Wisotzki.

The X-ray selected samples of AGNs in both CDFS and CDFN fields were drawn from the Extended CDFS (ECDFS) Point-Source Catalog by Lehmer et al. (2005) and The CDFN Survey XIII 2 Ms Point-Source Catalog by Alexander et al. (2003). Henceforth in the text we refer mainly to the CDFS sample, unless it is stated otherwise, since we do not have sufficient data observed in the CDFN. The ECDFS is a survey of four adjacent 250 ks Chandra observations. The area covered by the survey is approximately square and has a total area of ∼0.3 deg^2 and it incorporates the existing 1 Ms CDFS survey. The ECDFS survey limits of sensitivity reach $\sim 1.1\times10^{-16}$ erg cm^{-2} s^{-1} and $\sim 6.7\times10^{-16}$ erg cm^{-2} s^{-1} for the 0.5−2.0 KeV and 2 − 8 KeV bands, respectively. This sensitivity is sufficient to detect the X-ray emission from moderate-luminosity AGNs ($L_X = 10^{43} - 10^{44}$ erg s^{-1}) at redshifts up to $z = 6$. Within the field there are 762 distinct X-ray point sources, which were carefully matched to the Galaxy Evolution from Morphologies and SEDs (GEMS) survey (Rix et al., 2004) for high-resolution imaging and COMBO-17 survey (Wolf et al., 2003) for photometric redshifts. The resulting sample includes 606 entries. We carried out the analysis and testing of the pipeline using a sub-sample of ∼300 sources, for which high-resolution postage-stamp images were available from the GEMS survey.

2.1 High-resolution data: AGN and PSF stars samples

For high-resolution imaging we used HST observations for the GEMS survey. It is a large imaging survey carried out in two HST bands (F606W filter corresponding to broad V-band and F850LP corresponding to SDSS-z-band) using the Wide Field Camera of ACS. The survey covers over 900 arcmin^2 in the ECDFS (see Figure 8, on the right), and is split into a mosaic of 9x9 ACS tiles. In the central 25% of the field, GEMS incorporates the GOODS survey (Giavalisco et al., 2004), which we did not employ in this study. Each pointing in GEMS has an integrated exposure time of ∼ 2200 seconds in each band, with limiting magnitudes of $m_{F606W} = 28.3(5\sigma)$ AB and $m_{F850LP} = 27.1(5\sigma)$ AB.

The area covered by GEMS is overlapping with one of the 4 fields of the

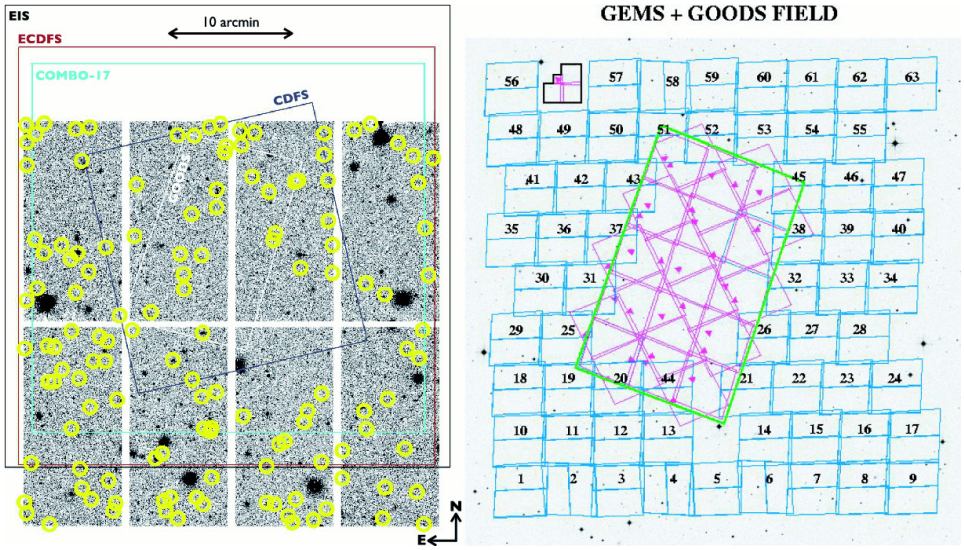

Figure 8: Left: V-band image of the CDFS, taken with the WFI, 2.2-meter telescope, La Silla, Chile. The extent of other surveys conducted in this field are shown in color, as: i) in black is the ESO Imaging Survey (EIS) (Arnouts et al., 2001); ii) in cyan is the COMBO-17 survey (Wolf et al., 2003); iii) in blue is the CDFS 1 Ms survey (Alexander et al., 2003; Giacconi et al., 2002); iv) in red is the ECDFS 250 ks survey (Lehmer et al., 2005); and v) in white is the GOODS survey (Giavalisco et al., 2004). Yellow circles mark variable AGNs, that belong to the catalog of Trevese et al. (2008), from where this plot was taken. Right: GEMS field mosaic. In the central $\frac{1}{4}$ of the field GEMS incorporates the GOODS Deep Field Survey. The entire field consists of ~90 HST pointings. Each observed tile is combined with MultiDrizzle from 3 dithered exposures, ~ 760 seconds each, observed with the F606W and F850LP filters. Plot is taken from Caldwell et al. (2008).

COMBO-17 survey (see Figure 8, on the left), which is a multi-color ~ 1 deg^2 survey, that utilizes observations carried out in 17 optical bands with WFI. It provides photometric measurements for a sample of ~50,000 galaxies and ≤1,000 quasars to R ≤24 (Vega zero-point). Multi-band photometry from the COMBO-17 Survey provides photometric redshift estimates available from low-resolution SED fitting for the entire Survey catalog.

GEMS HST data was reduced prior to this project by the members of GEMS collaboration. For the variability study, we used the F606W band images, which by the nature of stellar emission are deeper. As it is seen in Figure 9, the contamination of the host galaxy relative to the AGN flux depends on the AGN type. The intensity of the continuum of a host galaxy drops dramatically shortwards of the 4000 $\AA$ break, therefore all galaxies at redshifts $z > 0.5$ observed with the F606W will have a rather insignificant contamination from the host galaxy. On the other hand this will make the modeling of the host galaxy brightness profile more complicated and subjected to larger error. To separate the AGN and its host galaxy, we performed 2-dimensional brightness profile fitting of the high-resolution HST data using GALFIT (Peng et al., 2002, 2010). We compiled a catalog of model parameters created for each galaxy in our sample. This catalog will be used later to create models for low-resolution ground-based WFI data to extract residual flux for the variability measurements.

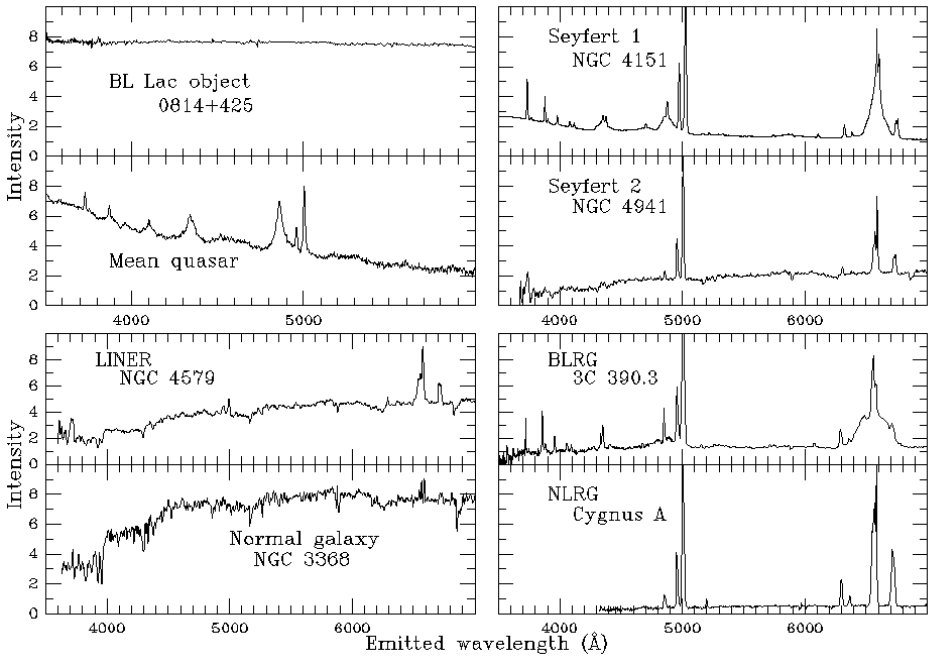

Figure 9: Optical spectra of various kinds of active galactic nuclei and a quiescent galaxy for comparison. Image credit to B. Keel.

As was already mentioned, our sample of AGN consists of > 600 objects, from

which we selected a sub-sample of ~300 AGNs to test the pipeline. Additionally, we had created a sample of stars, extracted from the HST tiles in the same way as the postage-stamp images of AGNs. We required this sample of stars to create empirical PSFs to be used for the 2-D modeling. Each postage-stamp image of both AGNs and stars was 128×128 pixels, which was sufficient to cover the source plus empty sky around it, required for the background measurements. Images were carefully cropped from the original ACS tiles, reduced with a pixel-scale finer than the original $0.03''$/pixel to provide a better sampling and resolution for the ACS PSF (Caldwell et al., 2008) (a detailed explanation for the HST data reduction and the choice of data-reduction parameters will be given in Chapter II). The diffraction limited optical system of HST, not restricted by the Earth's atmosphere, has on average 10 times higher angular resolution and 8 times finer pixel-scale than the WFI detector. Thus we removed from the sample all AGNs that have a close companion, that will be unresolved when observed with WFI. If the background source was located far enough from the AGN on the postage-stamp image, and did not affect the photometry of the targeted AGN host, it was masked out and ignored during the fitting process and background subtraction. In the case of stars, all the background sources, that were far enough from the star and did not affect its brightness profile, were masked out using patches extracted from the empty sky area. It is done in order to prevent empirical PSFs, averaged from these stars, from having uneven background patterns.

2.2 Background subtraction

After cleaning and masking out all the background sources in both samples, we had subtracted the local value of the background from all of the images. The method we used for background subtraction is very simple, yet very precise: background subtraction using curves of growth (CG). It allows the determination of the precise level of the sky background locally around each desired object, using aperture photometry. Basically, we performed least-square fitting of the distribution of measured flux values as a function of the aperture size to determine the local value of the sky[1]. In order to not overestimate the background, one has to make sure that each postage-stamp image has a sufficient area of empty sky, where the flux from the wings of the PSF is negligible. All of the apertures have to be centered precisely on the object. We used 2-D Gaussian fitting of the brightness profiles to determine the sub-pixel centers of galaxies. In the case of zero contribution from the background, for apertures of $r > R_{97}$, where R_{97} is the radius that encircles 97% of the flux in the PSF wings, the integrated aperture flux should remain constant with increasing aperture diameter. Thus, if the fit to the outer part $r > R_{97}$ of the CG has positive inclination, it means that the current value of the background is positive and should be subtracted from the image, and vice versa, if the CG is going down, the background is oversubtracted and should be added to this image.

[1] the global sky background value is normally subtracted by MultiDrizzle during the data reduction process

All the codes implementing our Pipeline and its procedures are written in Python, taking advantage of integrated MIDAS environment for some operations. The code iteratively performs least-square fitting of the CG outer part until the estimated background level becomes negligible or the maximum number of iterations is reached. Since 90% of our sample of AGNs are rather faint, with magnitudes of $m_{F606W} > 18$, the size of the postage-stamp images was selected in a way to have sufficient empty sky available for background determination. We manually checked the results of the background subtraction and corrected it if necessary. For example, we had to alter the limits for sky level estimation for the brightest and faintest objects in the sample, shrinking the ranges for measurements for the brightest objects due to the flux in the wings of the PSF, and allowing for larger ranges for the lowest signal-to-noise objects, where variations of the background become so large that they can spoil sky estimation. The iterative approach is required, because the estimated sky value makes a contribution to the pixel-count of the source as well. Therefore, at each iteration we subtracted 55% of the estimated value of sky, until the integrated input from the background becomes negligible or until the maximum number of iterations is reached. The fraction of 55% was determined empirically in order to minimize the number of iterations.

In the case of stars, we selected a sample of unsaturated moderate-luminosity stars, limited by $R < 24$. After we masked out the background objects, we subtracted the local sky background level as described above. We used these stars to create empirical PSFs for every AGN in our sample. For ACS, its PSF is variable both spatially and temporarily due to geometric distortion of the image and change of the focus distance of the secondary mirror, when telescope goes in and out of the Earth's shadow (see Sections 3.2 and 8 for details). To create PSFs which would reflect the spatial changes of ACS PSF and have high signal-to noise, we averaged 35 stars, applying a condition, that all of them have to be located not further than 40 pixels away from the position of investigated AGN independently on the tile as in Jahnke et al. (2004) and Sánchez et al. (2004). The final sample of stars suitable for averaging of PSFs consists of ~200 objects.

2.3 WFI low-resolution imaging data

All of the 5 epochs of raw WFI observations were obtained from the ESO data archive together with the corresponding calibration observations (bias, dark, standard, sky flat frames or dome flats, where flat field frames were not available). The data were reduced using a modified version of GaBoDS pipeline by Erben et al. (2005), adjusted by J. Kohnert.

We downloaded the archival CDFS data, observed in B, V and R bands. However, we analyzed only the R-band data, because among these filters, the central wavelength of the Rc162 filter (with the central wavelength of $\lambda_{ef} \sim 6517$ A) is the closest to F606W ACS filter ($\lambda_{ef} \sim 6060$ A), used for high-resolution observations. During the data reduction process, all the images, taken during all observing runs not more than 4 months apart, were combined together. The integrated ex-

posure time of the R-band image that we used to calibrate our pipeline is 9 000 sec, composed of 30 exposures of 300 seconds each. The data in the other bands is much shallower but can be used to extract color information.

2.4 Data reduction

Originally, the GaBoDS pipeline was developed to process data for the Garching Bonn Deep Survey, which was also carried out with the WFI at 2.2-meter telescope. This data reduction pipeline is almost entirely automated, and allows to process raw telescope telemetry into astrometrically and photometrically calibrated combined mosaic images. The modified version of the pipeline uses a Python wrapper, created by Jan Kohnert. The principal difference to the original GaBoDS pipeline is that it uses alternative software for combining images and for astrometric calibration and allows to reduce data obtained at various instruments. In this section is a detailed outline to the data reduction process employing the modified version of the GaBoDS pipeline, summarizing the GaBoDS Pipeline Handbook (Erben et al., 2005).

The signal of an object in a raw astronomical observation can be defined with the formula:

$$N(x,y) = (N_{obj}+Sky\times(1+Fringes))\times E_G\times E_A \times FF+Bias+Dark+Cosmics+CCD \quad (1)$$

To measure the pure flux of an object N_{obj}, one has to subtract all the additional sources of signal and noise in the image listed in this formula. These can be split into two classes: natural and instrumental. To natural components we primarily ascribe the brightness of the night sky (denoted *Sky* in the equation 1), modified by fringes. Fringes are an imprint of the interference pattern that occur in I-band observations (and longwards of I-band), thus it does not impact our R-band data. E_G and E_A are Galactic and atmospheric absorption coefficients, that determine how strongly the signal is weakened by intervening absorption along the line of sight by the Earth's atmosphere and by the gas and dust in the interstellar medium. The atmospheric absorption depends on the Zenith angle of observations: the smaller the ZA, the thinner the layer of atmosphere that light has to go through, and thus the weaker the absorption. The Galactic absorption depends on the direction of the line of sight. The distribution of gaseous clouds in our galaxy is well known and was published by Schlegel et al. (1998). A very important source of uncertainties in raw astronomical images are cosmic rays, which are high energy particles hitting the detector during observations.

The instrumental signal components in the formula (1) depend on particular qualities of a CCD. These components include flat field, bias and dark components. The flat field, denoted as *FF* in the formula (1), defines pixel to pixel sensitivity variation across the CCD. Bias level characterizes the basic level of signal in a detector and the dark level defines the thermal noise component. There is also an

additional term *CCD* in this formula that represents the defects of the detector, such as bad pixels and columns. After all of these extra signal components are properly removed the user can proceed with the photometric calibration.

All the steps which the Pipeline is performing to reduce raw telemetry data can be organized into the following scheme:

1. **Copying data and scripts to the specified destinations. Sorting data into a "data reduction tree (DRT)"** During this step, the pipeline copies and splits all the available data according to the types of observations and auxiliary files into corresponding directories.

2. **Creation of calibration images** During this step all the master images (bias, dark and flat) are created from the files located in the corresponding directories of the DRT. All the calibration files used during this step have to be visually checked prior to the pipeline run to make sure they are good quality: the pipeline *parameters file* (PF) allows to specify rejection limits for the range of pixel intensities for each type of calibration file, however a visual check is necessary to reject corrupt observations.

   There are three kinds of calibration images, which are required for the pre-reduction process. Bias images are zero time exposures of the CCD. They represent the minimum signal level that occurs in a detector during the CCD readout. As a rule, during every observing night 5 to 20 bias exposures are taken. There are two types of bias corrections that are applied to the data. The first correction takes into account immediate change of the bias level from one exposure to another. It is determined from the unexposed overscan region of every observed frame, and subtracted from all science and calibration images that are available, including bias calibration frames themselves. After that the overscan regions are trimmed, and all the bias frames are median-combined into a master bias frame, which is subtracted from the rest of the calibration and science frames.

   Similarly, a master dark image is combined from all of the suitable (bias-subtracted) dark frames. Together with bias calibration images, these observations should be done every night. Dark frames represent the noise that occurs in a semiconductor detector due to thermal electrons that are unavoidable at non-zero temperatures. Thus, the final number of counts depends on the exposure time and thermal conditions at which a CCD is kept. These calibration images are taken with the telescope shutter closed, with exposure times that are comparable to the exposure time of science observations. All suitable bias-corrected dark frames are averaged into a master dark frame, which is then subtracted from the flat field, science and standard observations.

   At the next step master flat field frame is created. It is used for correction of inhomogeneous surface brightness distribution in the images, which occurs due to variations of illumination and sensitivity across the CCD. Illumination

usually changes over large scales, while sensitivity may change from pixel to pixel and may also depend on the color of light which the CCD is exposed to. To account for these changes, two kinds of flat field calibration images can be taken: so-called sky flats, which are exposures of the sky during evening or morning twilight, or dome flats, which are exposures of a blank screen using white light lamps inside the telescope dome. In case of dome flats, it is difficult to take them for wide field cameras. In addition to that, the lamps, which are used to illuminate the screen inside the dome, have a maximum temperature of about 3000 K, which is insufficient for correcting science observations. Because of this, it is preferable to use sky flats instead. To get a good quality master flat field frame one has to take at least 5 exposures with different illumination ranges. It provides better sampling for the pixel-to-pixel intensity variation when a master flat field frame is combined. The master flat field was combined from at least 5 input flat field exposures, taken with certain exposure time steps to cover the broad illumination range of the CCD and normalized to unity.

In some cases, when the dithering pattern of science exposures has steps that are larger than the largest object in the field, these science frames can be combined into a so-called super flat field. This calibration frame is used to correct the immediate variations of sensitivity of the CCD at the given moment of observations. It is not necessary to perform super flat field correction, since the sensitivity of a CCD should not change significantly over short period of time. In the GaBoDS pipeline, super flat field frames are created by default, since they are used for illumination correction and global weight frames.

3. **Calibration of science frames** During this step all science and standard frames are bias-, dark- and flat field-corrected using the master frames created during previous steps of data reduction: from the overscan-corrected trimmed science and standard frames master bias and master dark images are subtracted. Then, the resulting frames are divided by the normalized master flat field frame. All of these steps of pre-reduction are performed in parallel, which allows to significantly speed-up the data reduction process.

4. **Manual masking** At this step all the science frames are visually inspected and masked in case they have defects, such as meteor trails, reflections of bright stars etc.

5. **Standard stars processing, and astrometric and photometric calibration of the science frames** Pre-reduced standard observations are used to create catalogs of standard stars required for photometric and astrometric calibration of science observations. When the stars, extracted from observed frames, are identified with a catalog of photometric standards, these identified pairs are used to compute high-order astrometric correction of the sky-pixel mapping of images. At this point, the pipeline produces plots that

help the user to check the quality of standard pair identification. Afterwards, matched standard pairs are used for absolute photometric calibration. Every pair of stars defines a system of equations, which is solved by the pipeline to derive the absolute photometric calibration parameters, such as photometric zero-point of observations (ZP), extinction coefficient (CE) and color term (CT):

$$m_i - m_{i,inst} = ZP + CE \times airmass_i + CT \times CI_i,$$

where i denotes each identified standard star, CI is the color index (e.g. (B-V)), m_i is the magnitude of each standard star in the catalog and $m_{i,inst}$ is the observed magnitude of the same star. The Pipeline solves this system of equations using a non-linear least-squares Levenberg-Marquardt algorithm with an iterative 3σ rejection. It allows to use rejected data-points later on if they match with a different solution. It is possible to interact with the pipeline to choose how many photometric coefficients should be fit. Finally, when a suitable solution is found, the image headers of the science frames are updated.

Astrometric calibration of the data observed with WFI is more complicated due to a large field of view (FoV). 8 CCDs which compose the WFI FoV are not perfectly aligned to each other which introduces a significant uncertainty in the astrometric solution (see Figure 10). Besides this, one has to take into account the spherical curvature of the sky projected on the detector. To solve these issues the modified GaBoDS pipeline uses SCAMP to create astrometric and photometric solutions (Bertin, 2010). SCAMP has a useful ability to deal with a large number of overlapping observations. Then, the astrometric solution is a result of a χ^2 minimization of differences of position between corresponding objects in overlapping fields.

For the photometric calibration it is important to take into account that individual observations could have been taken under different photometric conditions. That would lead to different photometric zero-points for each frame. Same as astrometric solution, SCAMP uses the χ^2 minimization of the magnitude difference of the matching pairs in the overlapping fields. For further details of the photometric and astrometric calibration, please refer to SCAMP 1.9 User's guide (Bertin, 2010).

6. **Final image co-addition** The final step of the data reduction pipeline combines all reduced images into a mosaic. This step is done with SWarp (Bertin, 2008). Basically, SWarp performs re-sampling of the images to combine them together, employing parameters specified in the updated headers of the images. The algorithm is similar to drizzling implemented in the MultiDrizzle data reduction pipeline, that is going to be discussed in detail in Section 7. More details on the usage of SWarp can be found in the SWarp User's guide by E. Bertin (Bertin, 2008) .

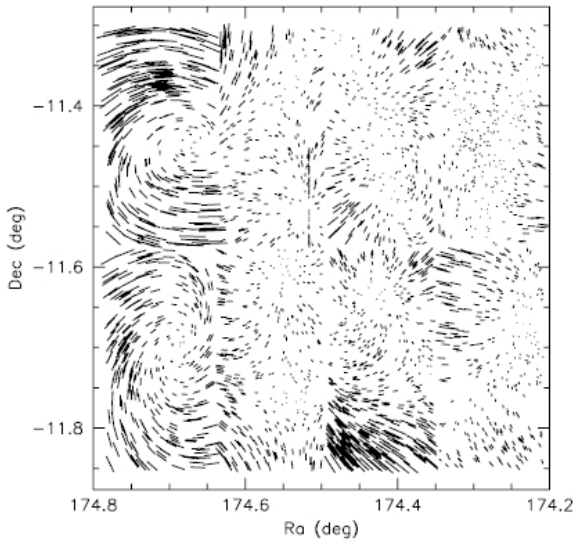

Figure 10: Difference between approaches for the astrometrical solution for a small field of view and for WFI. The maximum position difference in this image is about 6 pixels. This number is relatively small to compare with observations at other telescopes, but it proves the point, that a global distortion polynomial for all 8 chips of WFI camera does not work. Instead, a separate solution has to be created for each of the 8 chips individually.

The GaBoDS Pipeline is a convenient and versatile instruments pipeline. It allows easy interaction an correction of the data reduction parameters, giving an opportunity to have control over the reduction process with minimum effort. This pipeline utilizes parallel processing of the large amounts of data, employing the most efficient algorithms and ready-made software to produce best quality data product faster. However, every pipeline by its construction is a complex system, which requires a precise knowledge of the algorithms put in its basis and careful selection of the data reduction parameters. We used the GaBoDs pipeline to reduce and combine WFI data, as well as the first observed epoch of LBT data.

2.5 LBT observations

When we initiated the Deep variability survey we had one data set in hand, the multi-epoch ground-based WFI observations, with high-resolution HST imaging, X-ray follow up and photometric redshifts available. However, the principal part of the survey was to be carried out with the newly acquired LBT observations of the CDFN. This field, like CDFS, has HST imaging data available and it features the 2 Ms Chandra observations – one of the deepest X-ray observations conducted to date (Alexander et al., 2003). CDFN contains ≥ 500 X-ray sources over an area of 400 arcmin^2, which perfectly corresponds to the $23' \times 23'$ field of view of Large Binocular Camera of LBT. All of the X-ray detected point-sources in this catalog have spectroscopic follow-up available (Barger et al., 2005), however many of the X-ray emitting galaxies apparently do not show any signatures of AGN in their spectra.

We planned to pay 5 repeat visits to the field during 3 years, with $\sim$ 4000 seconds (+ 500 seconds for observational overheads) exposure per visit. The combined single epoch WFI observations have integrated exposure time of ~9000 sec. Even though the time of observations per epoch that we requested at LBT is less than half, the advantage in effective light-gathering power of LBT of ~15 times that of the 2.2-m telescope and lower sky surface brightness for observations in the B-band in a comparison to R-band, allow LBT observations a 2.5% better photometric accuracy at a $B = 24$ (AB) magnitude than that of WFI observations. 4000 seconds integrated exposure at LBT allows to obtain 10σ photometric measurement at $B \sim 26$ (AB), and reach ~1.5% photometric accuracy for an extended source of $B = 24$ (AB) magnitude. Moreover, the B-band corresponds to the rest-frame UV for objects at redshift $z \geq 0.2$: rest-frame UV emission of galaxies is less contaminated by stellar continuum, therefore it provides a better probe of the putative AGN emission. In such conditions, high quality photometric accuracy of the data would allow to detect variable nuclei at as few as 1% of Eddington accretion rate if it varies by a factor 2.

Due to various reasons (high competition, commissioning of new instruments, numerous technical problems at the telescope) we obtained only 2 (incomplete) epochs of observations spanning 2 years instead of the 4 epochs we applied for during 2008-2009. Accepted observing proposals were granted medium priority,

which in the absence of requirements of photometric conditions and excellent seeing, provided high probability that the data will be collected. Nevertheless, technical difficulties with the telescope operation, during which the red arm of the telescope was out of order, bad weather conditions and commissioning of a new instrument impeded acquisition of the complete data set we required to continue this project.

3 Analysis pipeline

3.1 Modeling of the high resolution data

We already gave an example how variability in galaxies can be extracted using aperture photometry in the studies of Trevese et al. (2008) and Sarajedini et al. (2011). By comparing the resulting catalogs of variable sources, drawn from overlapping fields, we concluded that aperture photometry with a fixed size aperture as a stand-alone analysis tool has two main disadvantages: i) when attempting to compute aperture flux one unavoidably gets AGN flux estimates contaminated by the host galaxy, which becomes crucial with decreasing AGN luminosity, and ii) this method does not take into account PSF instability, which exists for both space- and ground based observations, and particularly in the case of ACS observations can strongly impact the estimated significance of the variability detection. We suggest an alternative approach that minimizes the contribution from the host galaxy and also accounts for PSF variations. The method is based on 2-dimensional brightness profiles fitting of high-resolution data, that represents a reference frame for subsequent measurements. The best-fit parameters of a fitted galaxy are then degraded to match low-resolution ground-based data, and used to subtract the host galaxy model. In this way, one takes into account any changes of the PSF by convolving the created galaxy model with the time- and position-dependent PSF for each object in the sample, and can efficiently separate nuclear and host galaxy components. If the model is well constrained, measuring the residual flux of the host-subtracted galaxies will effectively reduce the threshold for variability detection.

To perform 2-D modeling of brightness profiles of galaxies in our sample, we used the GALFIT software, created by C. Peng (Peng et al., 2002, 2010). GALFIT is based on the algorithm that minimizes the χ^2 residual between an image of a galaxy and its model, which is parametrized with a set of user-defined variable parameters, that GALFIT simultaneously adjusts to find the best-fit solution. The alteration of parameters is controlled by a Levenberg-Marquardt minimization algorithm (damped least-square fitting algorithm), which is among the fastest algorithms to find the local minimum for a set of non-linear equations in a large parameter space. The definition of χ^2, normalized to the number of degrees of freedom, χ^2_ν, is:

$$\chi^2_\nu = \frac{1}{N_{dof}} \sum_{x=1}^{nx} \sum_{y=1}^{ny} \frac{(flux_{x,y} - model_{x,y})^2}{\sigma^2_{x,y}}, \tag{2}$$

$$model_{x,y} = \sum_{\nu=1}^{nf} f_{\nu,x,y}(\alpha_1 ... \alpha_n),$$

where N_{dof} is the number of degrees of freedom, *nx,ny* - are the dimensions of the image in pixels, $flux_{x,y}$ is the flux at *x,y* pixel of an image and $\sigma_{x,y}$ is the noise at a given *(x,y)* pixel. The model at each pixel is defined as a sum of the *nf* functions $f_{\nu,x,y}(\alpha_1 ... \alpha_n)$ contributing to the model with $(\alpha_1 ... \alpha_n)$ parameters. The number of

degrees of freedom is determined by the number of free parameters of each fitted component at every unmasked image pixel, therefore χ^2_ν will grow smaller for a larger number of components. However, additional components may not bear any physical meaning, especially for high-redshift galaxies, therefore to find the best quality fit we relied on the relative change of χ^2_ν for a fixed number of degrees of freedom depending on the input parameters of each model. We assumed a simple one-component Sérsic profile fit for all galaxies, or Sérsic+PSF model otherwise. Initial guess parameters for a fitted object were determined using SExtractor (Bertin & Arnouts, 1996). For a Sérsic profile component GALFIT requires several initial parameters, namely *(x,y)* the center coordinates, the effective radius, the minor-to-major axes ratio and the position angle, which are output directly from SExtractor. As an initial guess for the Sérsic index we used the $\frac{R_e}{R_{Kron}}$ ratio, with both radii estimates computed by SExtractor as well (Graham et al., 2005). If the visual check indicated the necessity, we additionally included a point-source component. Since the Levenberg-Marquardt algorithm finds only the local minimum of a set of equations it is important to use initial guess parameters as close to the real values as possible. Therefore we took special care in tuning the SExtractor output.

GALFIT requires 4 input images (image of a galaxy to fit, corresponding sigma-image, PSF image or array, and bad pixel mask) and a number of parameters of the input data, specific to a telescope to be specified. Using these parameters and initial guesses for each object, GALFIT creates a model, which is convolved with a PSF array to take into account atmospheric seeing during observations as well as instrumental features of a telescope. The noise map, utilized by GALFIT specifies the RMS at each pixel and can be provided by the user or, if not, is generated by GALFIT based on the *GAIN* and *READ-NOISE* values specified in the image header. In our case, noise maps for the observations were assembled from the weight images, produced during data reduction. However, the recipes for ACS and WFI data are slightly different, due to the complicated drizzling algorithms that are utilized for all HST observations, which alter the Poisson noise distribution of the non-drizzled images (for detailed explanation of MultiDrizzle operation see Chapter 7). For the ACS data we used for modeling galaxies in our sample, the noise maps are defined as follows:

$$\sigma_{x,y} = \sqrt{VAR_{x,y}} = \sqrt{\frac{\left(\frac{(SKY+75.)0.36^2 t}{WEIGHT_{x,y}} + FLUX_{x,y} t\right)}{t^2}}, \tag{3}$$

where $VAR_{x,y}$ is the variance at each pixel, *SKY* is the mean number of sky counts between 4 detector amplifiers in every tile, subtracted during the data reduction, t is the exposure time, $WEIGHT_{x,y}$ and $FLUX_{x,y}$ are the weight and the flux in cps/sec at (x, y) pixel, and 0.36 is a scaling factor due to re-sampling of data between 0.05″/pixel and 0.03″/pixel pixel-scales. For the bad-pixel mask we used masks which we prepared before for the background subtraction, where all the background sources that did not impact the photometry of the target galaxy where

masked out. The user can also use a so-called constraints-file, which is a simple text-file that specifies ranges in which free parameters are allowed to change during the GALFIT run. Another important GALFIT parameter is the size of the convolution box and the region of the image to fit. If the second parameter is self explaining, the first one determines the ranges of a frame that will be convolved with the PSF. The sizes of the convolution box and the PSF array to a great extent determine the speed of each iteration, as they are the most time-expensive steps GALFIT performs. Several other parameters GALFIT requires for operation are the PSF oversampling factor (1 for PSF averaged from the data itself), the pixel scale of the detector (not important for a Sérsic profile fits), and the photometric zero-point (ZP) of the observations, which for the HST is determined from the formula

$$ZP_{AB} = -2.5\log(PHOTFLAM) - 21.10 - 5\log(PHOTPLAM) + 18.6921,$$

3.2 PSF creation

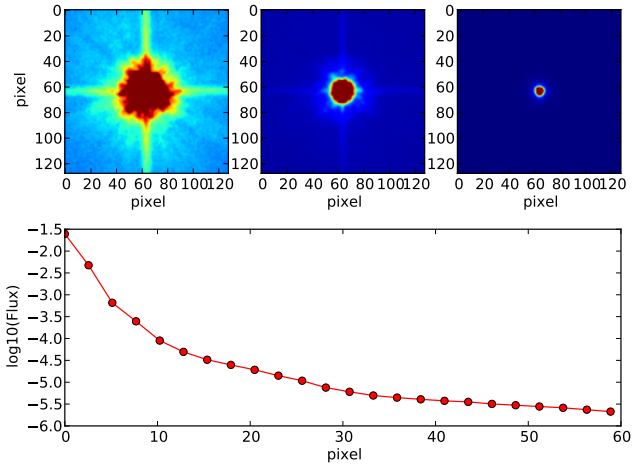

Figure 11: Plots in the top panel show the PSF image created from 266 stars in the field, applying different cuts. The leftmost image has the shallowest cuts and shows the extension of the PSF at large radii with 99.5% of the energy encircled, and its structure. The rightmost image shows the cuts through the core part of the PSF. In the bottom panel the brightness profile is shown. The x-axis is distance from the PSF center in pixels.

After a GALFIT run, the quality of the resulting best-fit model strongly depends on the PSF image, which determines the response of the imaging system of a telescope to a point source. For ground-based observations, the main factor which impacts the quality and stability of the PSF, besides the imaging system of the telescope, is atmospheric seeing, but for the HST observations, this factor is eliminated. However it is wrong to assume HST's PSF to be stable over a long

period of time across the entire field of view of the detector. Even though the full width at half maximum (FWHM) of ACS's PSF is much smaller than that of a ground based telescope, its surface brightness profile is not stable and varies both in time (due to so-called "telescope breathing", which is a change of the focus distance due to the temperature variations, when the telescope goes in an out of the Earth's shadow) and spatially (due to geometrical distortion by the optical system of the telescope). To account for these changes one either needs to create an empirical PSF model for a given position on a CCD at a certain time using stars extracted from corresponding observations, or to compute a parametrized model, which accounts for changes of the PSF profile for a user-defined position on the CCD and for a certain focus value of the telescope (for example using Tiny Tim software).

We used the PSF creation routine created by K. Jahnke (described in Jahnke et al., (2004)), which combines a certain number of nearby stars to get a high signal-to-noise empirical PSF model. This is a complex procedure written for MIDAS, utilizing fast "C" codes for image re-sampling written by L. Wisotzki. This code requires several input parameters, such as the position on a CCD where one wants to create a PSF and a list of stars available across the field with their *(x,y)*-pixel coordinates, to be used for PSF averaging. There are options to create a PSF from i) a certain number of nearest stars, or ii) using all stars within a certain radius. In the first option, all PSF models will have a certain signal-to-noise ratio, that depends on the number of stars used (the more stars the better S/N). In the second option, by selecting stars located very close to the desired position on the CCD, the user can take into account very small spatial variations of the PSF. These two options can be effectively combined, when the stars are distributed evenly and with a sufficient density across the field of view.

For each galaxy in our sample we used the option to combine the 35 nearest stars, making sure all were located less than 40 pixels away from the galaxy's central pixel position. Since the GEMS survey was observed during relatively short period of time, we neglected temporal variability caused by temperature variations, and assembled all stars from different tiles into the same catalog. This allowed us to reach sufficient density of stars in the field of view, to be able to take into account spatial variability of ACS PSF. Every star used for PSF construction was re-centered to a common center using a Hahn kernel for the pixel re-sampling, and after that radius-weighted-averaged into a final model of a PSF. Radius-weighted averaging requires a noise-suppression image

$$S \sim 6.283 s_0 \sigma^2, \tag{4}$$

where s_0 is a Gaussian function, centered on the star and with a width of

$$\sigma = (\frac{logI - logI_0}{2.35} + \sigma_0^{\frac{1}{4}})^4. \tag{5}$$

This relation is a fit to a PSF profile to translate a change in flux into a change in radius, where a certain signal-to-noise is reached (a factor of 2.35, $I_0 = 28.0$ and

$\sigma_0 = 32.0$ are optional empirical reference parameters, picked manually for GEMS data). All stars, after they are re-centered, are multiplied by their noise suppression images S, normalized to 1 and summed (the variance maps are treated correspondingly: every variance map is multiplied by the square of a noise-suppression image, normalized to the flux of a star, and summed). The PSF is then computed by dividing the sum of all stars over the sum of matching noise-suppression maps (correspondingly, the variance maps are created by dividing the sum of individual variance maps by the square of the sum of noise suppression maps). PSFs created in this way have an advantage of being local, which is important to account for spatial instability, and also have very high signal-to-noise ratio.

Figure 11 shows an example of a PSF, created as described above, without the restriction to locality, using all stars available across the field. In the top panel, the leftmost image is the one with the shallowest cuts, which encircles 99.5% of the intensity. High signal-to-noise shows the fine structure of the PSF and allows careful determination of the extension of the PSF-wings but does not trace variations due to imperfect geometrical distortion removal from the raw images. To show how significant the spatial variability of the ACS's PSF can be, we created 16 local PSF models, at a distance of 2000 pixels away from each other. Figure 12 shows the difference between the global PSF created from all stars in the field, and these local PSF models, each of which we averaged from 35 stars not further than 40 pixels away from the desired PSF location. It is clear that the profile of the PSF is significantly changing, and in order to get good fitting results and minimize the residuals due to PSF and data mismatch one needs to use local PSF models.

3.3 GALFIT run

We combined all the procedures which have to be executed to model a single galaxy into the same code. It was developed in Python programing language, using ESO MIDAS commands, which were enabled for the Python through the PyMidas interface. Images which will be fitted with GALFIT should be background subtracted as should be the images of stars used for PSF construction. The background subtraction is performed using the curves of growth method.

In order to show how the fit depends on the properties of a galaxy, we selected four galaxies with I-band magnitudes in the range $20 < V < 23$ AB and redshifts $0.5 < z < 0.9$. These galaxies belong to different morphological classes and have different brightness: a disk galaxy of $m_V = 22.3$, with a clearly resolved spiral structure; a bright elliptical galaxy of $m_V = 20.7$; a disk galaxy of $m_V = 22.6$ that has a significant inclination and no morphological features visible. Each of these galaxies was fitted twice –first with one component Sérsic profile and then with a combination of Sérsic+PSF.

For each galaxy our code performs the following steps:

1. First, the input parameter file for GALFIT is created. The first part of this file contains all the necessary information about the image, the instrument and the conditions of the observations.

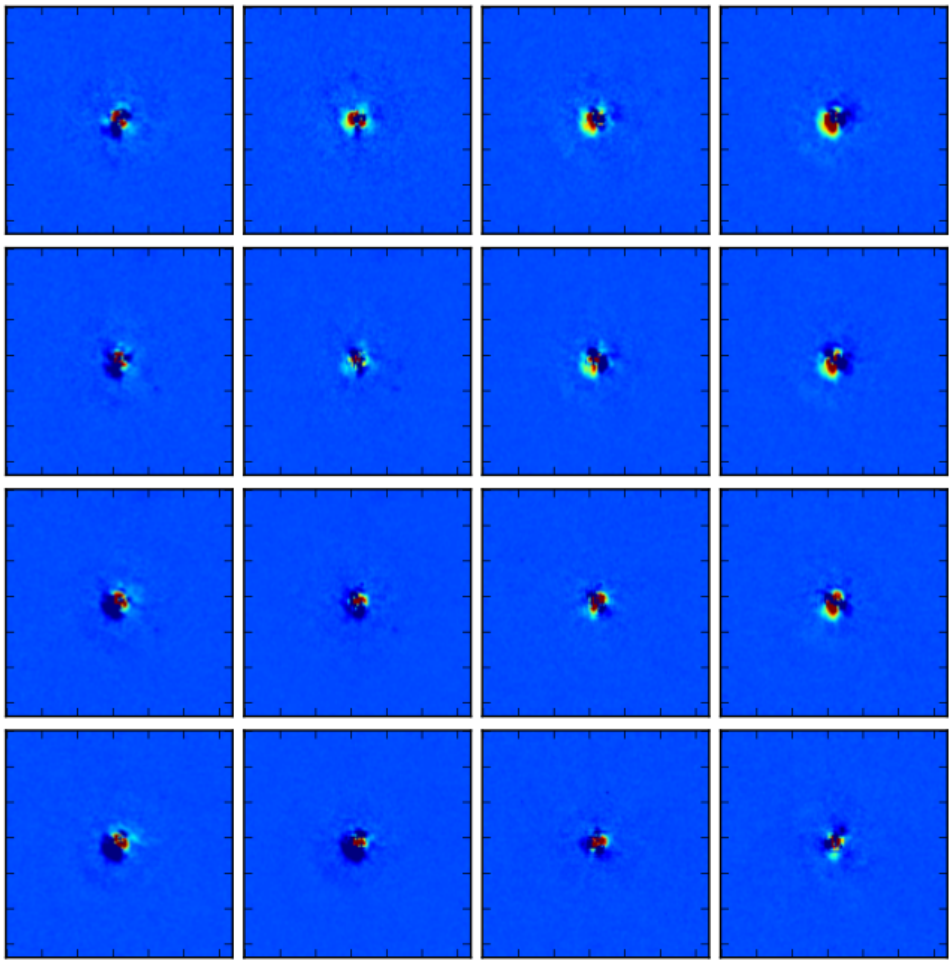

Figure 12: The difference between the global PSF created from 266 field stars ignoring the spatial variability of the ACS PSF due to imperfect geometric distortion removal, and the local PSF models, averaged from the 35 nearest $d < 40$ pixel stars. Each local PSF model is at least 2000 pixels away from the nearest neighbour.

2. An additional function creates a PSF model for the pixel-position of the galaxy employing the code used in Jahnke et al. (2004) described earlier.

3. In order to determine an initial guesses for the parameters of each galaxy, i.e. the effective radius, the Sérsic index, the minor-to-major axes ratio and the position angle, the image is ran through SExtractor.

4. After the parameter file is created, the code initiates the GALFIT run. If the fit was successful, GALFIT outputs a multi-extension FITS file, which contains the original image, the best-fit model and the residual image.

5. The code also produces a final catalog of best-fit parameters of each galaxy, including χ^2 values, which can be used as an additional criterion of the quality of the fit.

Besides visual inspection, one of the ways to judge the quality of the fit is the final value of the reduced χ^2_ν. If the sigma image follows the properties of noise and the estimated PSF reproduces the observations well, χ^2_ν should converge to unity, as can be seen from Equation (2): the difference between the galaxy and its model (convolved with the corresponding PSF) at each unmasked image pixel should be of the order of the variance at that pixel. We used a single component fit to reduce the number of degrees of freedom of each fit (or a Sérsic+PSF if a point-source was present). Given the small number of free parameters, a perfect convergence of a GALFIT run to a unity is highly unlikely, and the fit always requires visual inspection.

GALFIT performance is very sensitive to the input parameters. Even though these parameters are meant to be only a guess, they should approximate the actual values as good as possible. The GEMS Survey provides a catalog of morphological properties for many galaxies in this survey (Häussler et al., 2007). By calibrating parameters' guesses by SExtractor against this catalog, we could obtain initial guesses for objects that were not in it. All of the parameters required by GALFIT for a single Sérsic component (or a Sérsic+PSF) fit are therefore a direct output of SExtractor, except for the guess of the Sérsic index, for which the FLUX_RADIUS/KRON_RADIUS ratio serves as a good estimate (Graham et al., 2005), where FLUX_RADIUS is the SExtractor output for the effective radius (PHOT_FLUXFRAC parameter must be set to 0.5) and KRON_RADIUS is the SExtractor estimate of the radius that encircles 90% of galaxy's flux (Kron, 1980).

For every galaxy we performed two fitting procedures: a 1-component fit with a single Sérsic profile and a 2-component Sérsic+PSF fit. The need to include a PSF component arises when a galaxy hosts a prominent AGN which appears as a distinct bright central point source. For such galaxies a fit with a single component profile leads to a characteristic over-subtracted ring-like residual structure and a distinct residual point-source in the center, which are clear signatures that a second component should be used.

Figures 13 and 15 show examples of late type galaxies, at redshifts $z \sim 0.8$ and $z \sim 0.6$. A visual inspection of these galaxies enhanced by host galaxy subtraction

revealed clear signatures of spiral arms in the case of ID 39961 (Figure 13) and a dark dust absorption lane along the major axis in the case of ID 52472 (Figure 15). For both galaxies, the residual image of a single Sérsic profile fit clearly shows the presence of a faint point source. The change in the Sérsic index due to the inclusion of the PSF component into the fit indicates a change of the estimated morphological type of these galaxies from an early/spheroidal with Sérsic index $n \sim 3$ to a disk with exponential $n \sim 1$ profile and a faint central point-source component.

In case of the elliptical galaxy (ID 51199) shown in Figure 14, the difference between 1- and 2-components fits is very small. In both cases the galaxy does not change its morphological type and shows a good quality of residual already with a single-Sérsic fit.

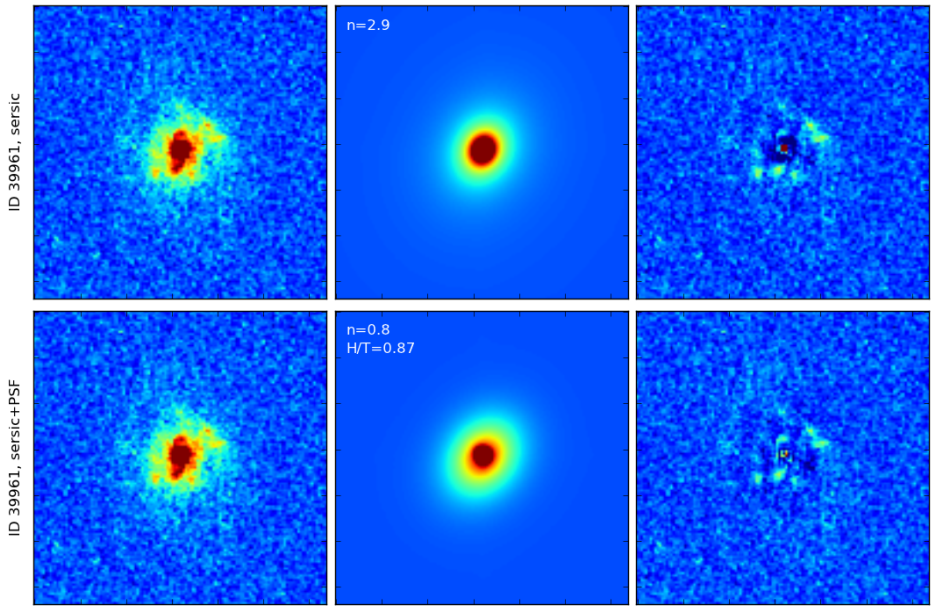

Figure 13: Spiral galaxy COMBO ID 39961, redshift $z = 0.812$. Top panel shows results of a single Sérsic component fit and the bottom panel is the Sérsic+PSF fit with a stated host-to-total ratio. In both panels from left to right: original ACS image cutout; GALFIT model; residual image.

3.4 Applying models to low resolution data

After high-resolution models were created, we downgraded the resulting parameters to match WFI data. WFI images of both galaxies and PSF stars were background subtracted during the sample selection, therefore all data was ready for the analysis. We performed the same PSF creation procedure that we performed for the ACS observations: we used the 35 nearest stars around the desired position to create high signal-to-noise PSF models. The number of stars available for this procedure is roughly the same as we had for the high-resolution data, however since

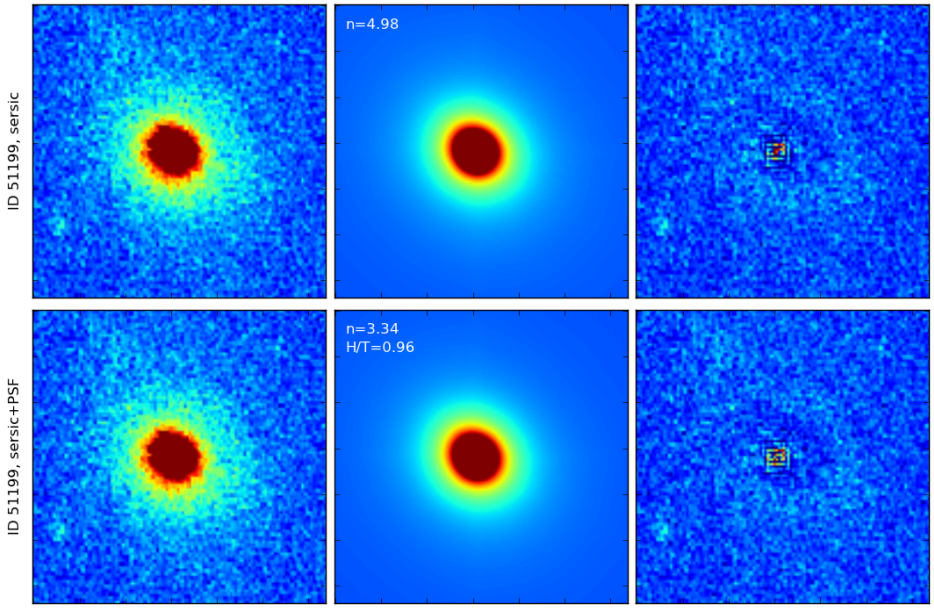

Figure 14: Early type galaxy COMBO ID 51199, redshift $z = 0.579$.

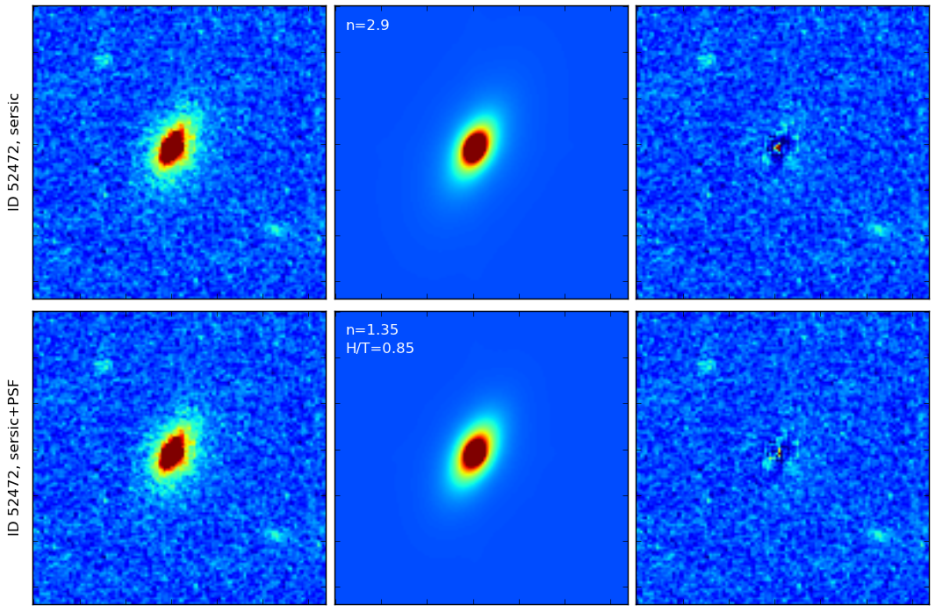

Figure 15: Late type galaxy COMBO ID 52472, redshift $z = 0.599$.

ID	e_1	e_2	d	ϕ
39961	0.016	-0.003	0.016	-5.3°
51199	0.013	0.012	0.018	20.8°
52472	0.001	-0.001	0.001	18.4°

Table 1: Measurements of the PSF anisotropies for WFI galaxies. The e_1 and e_2 (7 and 8) are ellipticities computed using the second-order moments returned by SExtractor, $d = \sqrt{e_1^2 + e_2^2}$ is the amplitude of the PSF anisotropy and $\phi = 0.5 arctan(\frac{e2}{e1})$ is its inclination, as in Figure 17

the FoV of WFI covers the entire survey area at once, the density of stars across the FoV is much lower. Nevertheless, the quality of the PSF in the co-added focal observations taken with WFI is very good. According to Erben et al. (2005) the largest PSF anisotropies of the WFI camera are of the order of 1% across the FoV with an average ellipticity of 0.003±0.008 throughout the field, as it is shown in Figure 17. These measurements were made in the R-band WFI data, co-added from 57 observations with ~500 sec exposure each. Indeed, if we compare the residual patterns and amplitudes of the deviations between the global PSF, constructed from > 200 field stars, and the local PSFs, created for the example galaxies (top left panel of Figure 16), they show very little difference, with amplitudes and directions of anisotropy following the map of Erben et al. (2005) in Figure 17 (red crosses mark galaxies' center pixel positions). Quantitative measurements of the amplitudes and the inclinations of PSF anisotropies for these galaxies are shown in Table 1.

In order to apply GALFIT models created for the ACS data to the WFI data, we introduced some changes to parameter files in order to account for the differences between the instruments and filters used for observations. The variance maps were derived from the weight maps in a similar way to how they were created for the ACS data in GEMS, without accounting for the change in the pixel scale (see Equation 3). In the case of photometric ZP, we used either the global ZP of the observations, computed by the GaBoDs pipeline or re-calibrated the data using standard stars (if the GaBoDs pipeline estimate had large variance). Then, if by keeping all the morphological parameters except the center of the galaxy fixed [1] for every galaxy in our sample, we can subtract its model at every epoch of observations. After that, aperture photometry measurements can be applied to every epoch of observations to detect any variable flux.

The very first problem we came across trying to realize this scenario is illustrated in Figures 18 and 19. Two spiral galaxies (ID 39961 on the left and ID 52472 on the right in Figure 18) appeared to be unresolved when observed with WFI, whereas the early type galaxy (ID 51199 in the center of Figure 18) has a very close companion, which was not visible in the ACS cutout. This galaxy is shown in Figure 19. Due to the absence of atmospheric effects and refined by drizzling pixel-scale, the FWHM of the ACS PSF is ~ 8.5 times narrower than that

[1] the value of the effective radius measured for ACS data has to be converted to the pixels-scale of WFI, namely $R_{e_{WFI}} \approx R_{e_{ACS}} \frac{0.03}{0.238}$

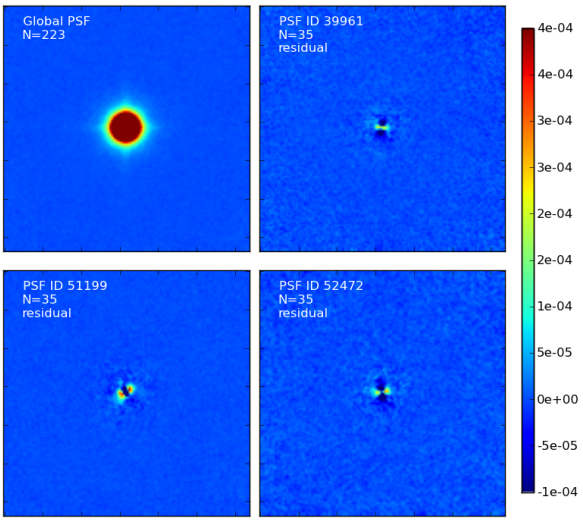

Figure 16: Top left panel shows the global WFI PSF created from 223 field stars. Other panels show residuals of subtraction of the local PSF models for above-mentioned galaxies from the global PSF, where each local PSF is averaged from 35 nearby stars. The deviation between these models is of the order of the RMS in the original 30×300 sec mosaic data.

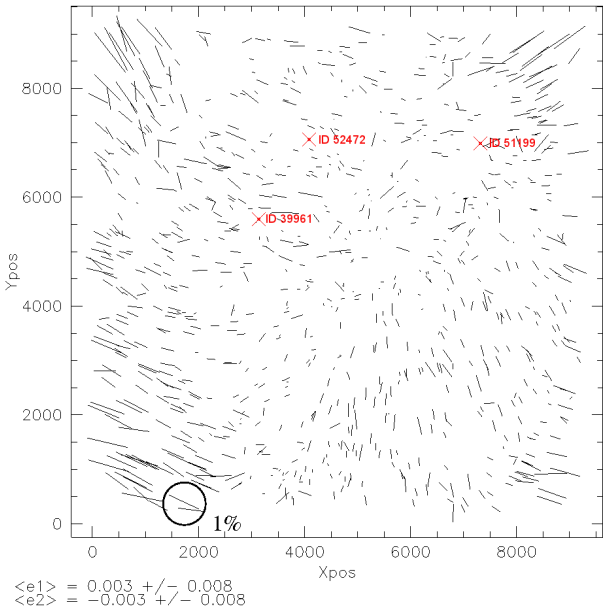

Figure 17: Anisotropies of the WFI PSF across the WFI field of view, measured in the WFI R-band mosaic, co-added from 57 images of 500 seconds each. The sizes of dashes represent the amplitude of the anisotropy, defined as $d = \sqrt{e_1^2 + e_2^2}$ and the inclinations are computed as $\phi = 0.5 arctan(\frac{e2}{e1})$. Image taken from Erben et al. (2005). Red crosses mark the location of example galaxies.

of the WFI PSF (0.12″ against 1.02″, see scales in Figure 20). This is why spiral galaxies, easily resolved with ACS appear to be unresolved with WFI.

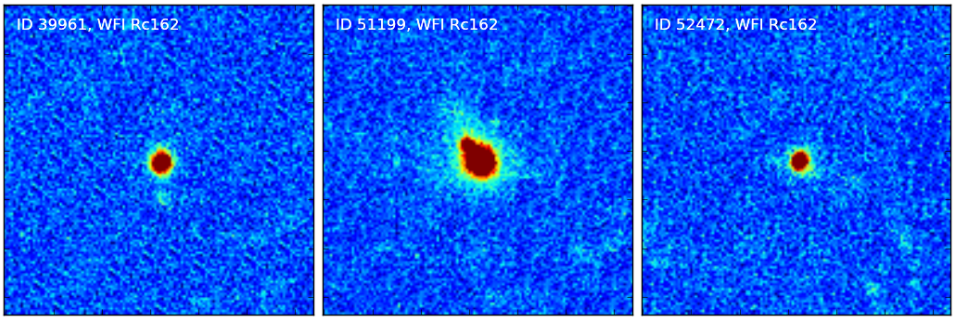

Figure 18: Example galaxies, from left to right: ID 39961, ID 51199, ID 52472.

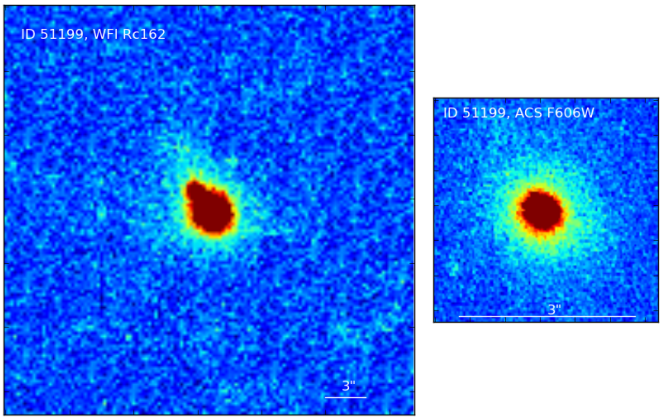

Figure 19: ID 51199, early type galaxy at redshift $z \sim 0.6$: on the left imaged by WFI and on the right by ACS. Scales are shown on the plot.

Nevertheless, we attempted to apply the high-resolution model of ID 51199 to the corresponding WFI observations. The result is shown in Figure 21. We applied a single Sérsic component fit, with all parameters except the center coordinates and effective radius fixed. We allowed the effective radius to vary, applying a 25% constraint, given that for the low signal-to-noise sources GALFIT may significantly underestimate the output values. The residual image shows a slightly over-subtracted residual in the central 2-pixel aperture, which may be due to variability.

We conclude that the given WFI data-set is not suitable for this kind of analysis. WFI is mounted on a 2-meter class telescope, which means that one requires long integration time to reach the depth required to study the variability of low-level AGNs. Additionally, one would have to select galaxies that have effective radii $r_e \gtrsim 0.6'' = 20$ pixel×0.03″/pixel in the ACS data set and $m_V < 21$ AB to

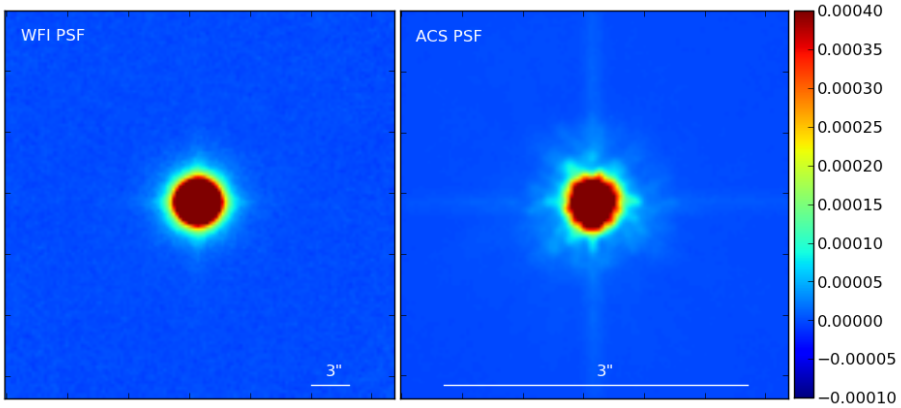

Figure 20: WFI and ACS global PSFs: both images are combined from the entire sample of stars ($N > 200$) available for this purpose. The pixel-scales of WFI and ACS are 0.238″/pixel and 0.03″/pixel correspondingly.

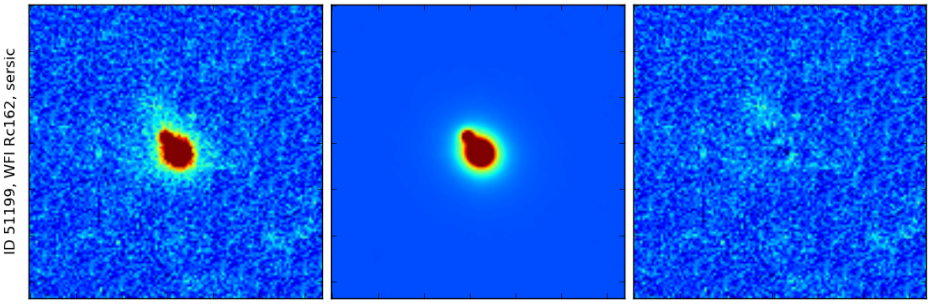

Figure 21: WFI observation of ID 51199, modeled with GALFIT. From left to right: original image, the model, created with parameters from high-resolution fit, and residual image.

make sure they will be resolved in the WFI data (the restriction on the effective radius is not only relevant to WFI but to LBC as well, since it has a pixel-scale of $0.226''$/pixel, comparable to WFI). According to the GEMS morphological catalog, CDFS has about 200 X-ray detected sources, whose angular sizes suggest that they can be resolved with ground-based telescopes.

How can we get better results? If we could obtain the LBT observations we applied for, we would have had a chance to get significant variability detections. If we compare the characteristics of the telescopes, the light gathering power of LBT is ∼15 times larger than that of WFI. Even though the combined exposure time of the WFI data is two times longer (at the longest epoch of observations) than what we proposed to obtain with LBT for a single epoch, the depth of LBT observations would be greatly higher. The exposure time calculator computes the signal-to-noise ratio for the LBC-Blue observation of an object of $B = 24.0$ AB with an integrated exposure time of 4000 seconds to be of the order of 60 (depending on the observing conditions and how we split the exposure), while for our WFI data for an $R = 24$ AB object signal-to-noise ratio is ~ 15. Thus, to obtain relevant results we would require LBT data.

4 Conclusions and perspectives

In order to create a survey of galaxies with low-level nuclear activity using variability to derive the accretion rate estimates, we created an X-ray pre-selected sample of galaxies in two fields in the Southern and Northern hemispheres: Chandra Deep Field North and South. Both fields have deep X-ray observations available, full coverage with HST, multi-band ground-based photometry, and spectroscopic or photometric redshifts available. By using X-ray pre-selected samples, we make sure that any variability if detected originates from an AGN, since independent detections of X-ray emission or continuum variability may be due to different reason. Since the penetrating power of X-ray emission is very strong even for column densities of $\times 10^{23}$ cm^{-2} (corresponding to a V-band absorption of $A_v \sim 150$ mag), it does not significantly affect the flux at $E \geq 5$ KeV band (Mushotzky, 2004). Therefore the probability to detect X-ray emission if it is present in a galaxy is very high.

For weak AGNs the X-ray emission is not significant and can be confused with X-ray emission originated by a starbrurst for example. In order to distinguish between AGNs and star-forming activity we intended to look for variability. In the presence of even weak X-ray emission, variability will be a distinctive property of the AGN activity, since integrated stellar populations do not vary in time. We planned to extract variability signatures of all X-ray emitting objects, and by measuring the amplitude of variability select those objects with accretion of the order of just a few percent of the Eddington's accretion rate. We employed a very sensitive method to extract variability, which would allow us to detect changes of flux even of a small amplitude. In order to find and measure variability we wanted to subtract host galaxy input from every object in our sample, applying models created using high-resolution HST/ACS data, to low-resolution repeated ground-based observations.

Our primary aim was to obtain repeated observations of Chandra Deep Field North, with its ~ 500 X-ray sources, using LBC. To establish the analysis pipeline before the observations would be acquired, we used archival multi-epoch photometric data from the WFI, which have a high-resolution HST followup available from the GEMS survey. While the work of establishing the image processing pipeline was conducted, both due to technical reasons and high competition for a share of observing time at LBT, we obtained only 2 incomplete epochs of data in one photometric band. Since after analyzing the the first epoch of the WFI data it became clear that its depth is not sufficient to obtain accurate results, this project was put on hold, until LBT observations are carried out.

Main accomplishments of this project:

1. We obtained archival raw ground-based WFI data and reduced it to use to set up the analysis code.

2. We developed the code for processing high-resolution ACS and low-resolution WFI/LBC data for our variability survey (the HST data were from the GEMS

Survey, the multi-epoch WFI data were downloaded from the public ESO archive, and the complete set of LBC data was never acquired). The code was developed in Python, and consisted of a set of functions to perform the full process of analysis of galaxies in the sample. 2-D surface brightness profiles modeling was carried out with GALFIT and the input guess parameters of galaxies were obtained by SExtractor. A separate function was executed to perform local background subtraction.

3. We aimed at subtracting the galaxy contribution from every object in the sample to discern even the smallest variation of flux if the central region of a galaxy. We created models of galaxies, using parameters obtained from high-resolution data modeling. We planned to measure the residuals of such models to determine the presence of variability, but the photometric depth of the WFI data was not sufficient to get reliable estimates of the variability amplitudes. We were forced to put this project on hold until we obtain the LBC observations.

Part II
Host galaxies of high-redshift X-ray selected AGNs in the COSMOS Survey

5 Introduction and motivation

In Part I of this thesis we explained pros and cons of several methods for selecting Active Galactic Nuclei. As we mentioned, the presence of X-ray emission is one of the most efficient ways to detect an AGN in both moderately obscured and unobscured AGNs at all redshifts. Indeed, detection of a Type 1 nuclei, which does not undergo significant absorption, is relatively straightforward within multicolor and spectroscopic (Coleman et al., 1980; Gavignaud et al., 2006; Richards et al., 2002; Maddox et al., 2012) surveys. For Type 2 AGNs identification becomes harder due to absorption, therefore they are most easily recognized by the presence of luminous X-ray emission ($L_{2-10\,\mathrm{keV}} > 10^{42}$ erg s^{-1}) and hard X-ray colours, since at this wavelength ranges and luminosities the contamination from star formation induced emission is negligible. X-ray selection remains very efficient up to line of sight column densities of neutral gas of $N \sim 10^{23}$ cm^{-2}, when AGN becomes Compton thick.

For the class of low-luminosity AGNs the launch of XMM and Chandra in 1999 enabled deep X-ray surveys, revealing a wealth of weakly accreting black holes in galaxies which otherwise show no presence of activity. X-ray bright optically normal galaxies, extremely red galaxies, sub-millimeter galaxies, star forming galaxies at high redshifts and Lyman break galaxies do not reveal significant emission lines due to obscuration and/or because the nuclear emission is overpowered by the host galaxy. Thus, combining multiwavelength optical, infrared, radio and X-ray data is crucial in order to understand the nature and evolution of AGN of all classes of luminosity at different stages of their evolution.

There are two most favoured scenarios that attempt to explain the onset of AGN activity in a galaxy and its evolution. The first scenario favours hierarchical structure formation in the early universe, where the key role in building up growth of present day galaxies belongs to major mergers (Toomre & Toomre, 1972; Kauffmann et al., 1993; Cole et al., 2000; Kauffmann & Haehnelt, 2000; Di Matteo et al., 2005). After two or more galaxies of similar mass merge, their galaxy product has a large reservoir of gas available to fuel the Super Massive Black Hole (SMBH) in the galactic center and also to drive a nuclear star burst (Hernquist, 1989; Barnes & Hernquist, 1991, 1996; Cox et al., 2006, 2008). At this primary stage of QSO evolution right after a coalescence, the nucleus is surrounded by optically thick gas and dust that were driven towards the center by the merger. Soon after a quasar phase

starts, its radiative emission would expel (or ionize) gas and dust around its nuclear regions exhausting the accretion and quenching the nuclear star formation (Silk & Rees, 1998). In this scenario the obscuration of an AGN can be explained by early evolutionary phase at which we observe the QSO, rather than from the point of view of AGN unification scheme. This is a simplified scenario, and the evolution of a given AGN significantly depends on numerous parameters, like BH mass and Eddington ratio, star formation in the host galaxy, AGN duty cycle length, or the environment in which AGN resides.

As the universe expands, mergers become less common (Toomre, 1977), and in order to explain the ubiquity of nuclear activity in the local galaxies one requires mechanisms other than major mergers. An alternative to the major mergers scenario, promotes secular mechanisms like minor mergers, galactic and nuclear bars, supernova explosions, etc. to regulate the transport of gas within a galaxy. This gas can feed the SMBH (Martini, 2004; Jogee, 2006; Hopkins et al., 2009) as well as form stars and build up a stellar pseudo-bulge component (Kormendy & Kennicutt, 2004; Jogee, 2006). The time-scales on which the secular evolution operates are much longer than the dynamical friction time-scales of major mergers. Secular processes are also less violent and are considered a main driver of nuclear activity in the classes of low- and medium-luminosity AGNs at low redshift. Indeed, Ueda et al. (2003) have shown the AGN luminosity function to be redshift-dependent: the luminosity function of AGNs of $\log_{10}L_x = 41.5 - 43$ erg s^{-1} peaks around redshift $z = 0.3$, that of medium-luminosity AGNs of $\log_{10}L_x = 43 - 44.5$ erg s^{-1} around redshift $z = 1$, and that high-luminosity AGNs of $\log_{10}L_x = 44.5 - 48$ erg s^{-1} around redshift $z = 2$. In this case it seems natural to assume that low-luminosity AGNs are triggered by secular evolution and high-luminosity AGNs by major mergers.

However, according to recent developments, the role of secular evolution-driven activity with growing AGN luminosity may be significantly underestimated. For instance, at redshift $z < 1$ a very low ($< 15\%$) fraction of disturbed AGN host galaxies and a 45% fraction of early-type morphologies suggests that these galaxies have not gone through a recent major merger phase (Cisternas et al., 2011; Gabor et al., 2009). This trend persists at higher redshifts. Schawinski et al. (2011, 2012) and Kocevski et al. (2012), for example, analyzed the morphology and disturbance rates among the AGN host galaxies with $\log_{10}L_x = 42 - 44.5$ erg s^{-1} at redshift $z \sim 2$ and found that about 50 to 80% of galaxies in their samples have disc-dominated morphologies. Also, about 50% of the galaxies in Kocevski et al. (2012) sample show no signatures of a recent interactions.

Both of these SMBH fueling mechanisms clearly play an important role in AGN evolution. However, it is not clear to what extent each process dominates at each class of AGN luminosity. Both mechanisms are efficient at transporting gas to the center of the galaxy, where it is then consumed by accretion and star-formation (Hopkins & Hernquist, 2009). The growth of the two components –SMBH and bulge –should occur at the same stage of galaxy evolution. This is suggested by the existence of numerous scaling relationships between the SMBH mass or AGN

luminosity and the properties of stellar bulge, like mass, luminosity or velocity dispersion (Kormendy & Richstone, 1995; Magorrian et al., 1998; Ferrarese & Merritt, 2000; Gebhardt et al., 2000) etc. To get a better understanding of the evolution processes as well as to study AGN triggering mechanisms one would need to create larger samples of objects especially at redshifts $z > 1.5$, around the peak of AGN accretion where most of the black hole growth occurs.

However, there is a major complication when attempting to study host galaxies of luminous QSOs at high redshift, namely the QSO may be so bright that it would overshine the host. To improve the contrast between the central point source and its host galaxy, the observations should be carried out at the maximum of the rest-frame emission of stellar component of a galaxy, which is rest-frame B to R bands, which corresponds to J to K band emission at redshift $z > 1.5$ (Shields et al., 2006; Alonso-Herrero et al., 2008; Schramm et al., 2008; Kotilainen et al., 2009; Schawinski et al., 2011, 2012).

Before 2007, when the Cosmic Evolution Survey (COSMOS) (Scoville et al., 2007) was carried out, high-resolution imaging surveys were performed on in small areas of sky, which could provide only small samples of high-redshift QSOs: the GEMS survey has a coverage of $\sim$ 0.22 deg^2 with $\sim$ 140 suitable high-redshift AGN candidates, a subsample of which was studied by Jahnke et al. (2004), and GOODS is $\sim$ 0.08 deg^2. The COSMOS Survey has the advantage of covering 1.64 deg^2, employing high-resolution imaging with ACS/HST in the broad-band F814W filter, which at the redshifts $z > 1.5$ corresponds to rest-frame UV emission, probing star-forming regions of AGN hosts. In addition, COSMOS has ground-based multiwavelength optical coverage of its entire area, employing large and medium-sized telescopes, such as Subaru, Canada-France-Hawaii Telescope (CFHT), United Kingdom Infrared Telescope (UKIRT), the Sloan Digital Sky Survey (SDSS) observations and more. Deep X-ray observations performed with XMM-Newton and Chandra in the central region of the field provide a catalog of X-ray detected sources, which together with infrared observations by the Spitzer Space Telescope (IRAC and MIPS instruments) and radio follow-up with the Very Large Array telescope provide an excellent base to study AGNs. Figure 22 shows the photometric depth of various observations carried out for the COSMOS Survey. These photometric data have been employed to estimate high-quality photometric redshifts through SED fitting. The COSMOS field has been partially observed spectroscopically with (i) the Very Varge Telescope (VLT) at ESO using the VIMOS spectrograph (540 hours) producing a catalog of 25 000 galaxies at redshifts $0.3 < z < 1.0$ with $I_{AB} < 22.5$ and 12 500 star-forming color selected galaxies at redshifts $1.4 < z < 2.5$ (zCOSMOS program); (ii) IMACS spectrograph at the Magellan Observatory (5 nights) sampling > 3000 high-redshift galaxies and AGNs; (iii) short programs at Keck, Galileo etc.

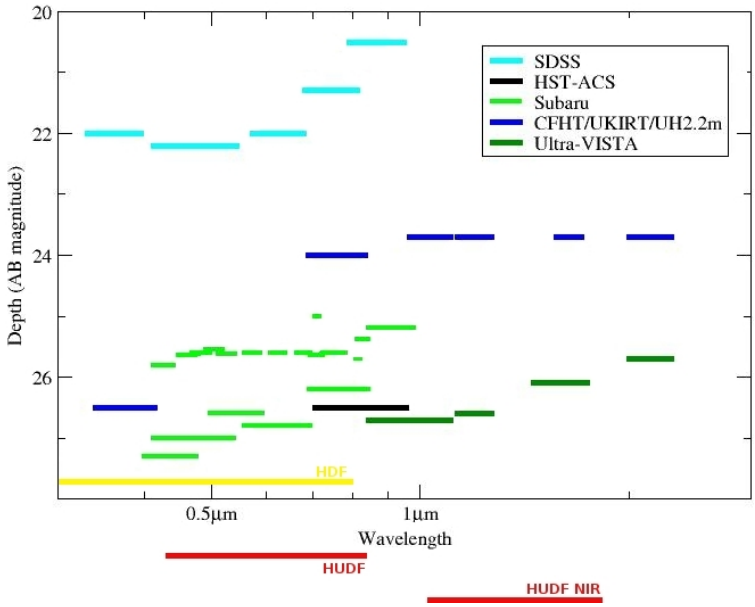

Figure 22: 5σ depth for all existing COSMOS observations. The depth for each campaign is measured as the flux in a 3“ aperture, except for ACS observations, where a 0.15“ aperture is used. Figure taken from the COSMOS web page.

6 Sample

The COSMOS Survey has a complete coverage of its area with the XMM-Newton X-ray telescope (XMM-COSMOS, Hasinger et al. (2007); Cappelluti et al. (2009)) through 55 pointings with a total exposure time of $\sim$ 1.5 Ms. The catalog of Brusa et al. (2010) comprises multiwavelength data for 1848 point sources, which include three XMM X-ray bands: soft (0.5-2 keV), hard (2-10 keV) and ultra-hard (5-10 keV) with sensitivity down to flux limits of $\sim 5 \times 10^{-16}, \sim 3 \times 10^{-15}, \sim 7 \times 10^{-15}$ erg cm^{-2} s^{-1}, correspondingly in each band; *r*- and *I*-band AB magnitudes from the Optical and NIR catalog (Capak et al., 2007); *K*-band magnitudes from the COSMOS-WIRCam Near-Infrared Imaging Survey (McCracken et al., 2010); 3.6 μm, 4.5 μm, 5.8 μm and 8 μm fluxes from Spitzer IRAC observations for the Photometric redshifts catalog of Ilbert et al. (2009), and 24 μm magnitude from the Deep Spitzer Imaging catalog with MIPS camera by Le Floc'h et al. (2009). The Brusa et al. (2010) catalog also contains additional information besides photometry, such as X-ray hardness ratios of every source detected in the hard- and soft-X-ray bands, photometric and spectroscopic redshifts (when available), together with a flag indicating which instrument the spectra was taken with, the spectroscopic class of the object (broad line AGN (BL AGN), narrow line AGN (NL AGN), normal/star-forming galaxy), and the SED template used for photometric redshift determination.

To identify all the XMM detections, they were matched with other catalogs

in this field. 26 sources, out of the total number of XMM detections above the likelihood threshold adopted for these XMM observations, match with the diffuse X-ray sources in the catalog of Finoguenov et al. (2007) and thus they were excluded from the XMM point-source catalog. The $\sim$ 0.92 deg^2 area in the central part of the XMM field was observed by Chandra with a total exposure of $\sim$ 1.8 Ms, which results in these observations being 3-4 times deeper than those of the XMM-COSMOS survey. In the overlapping area of XMM and Chandra observations there are 945 sources (which constitutes $\sim$52% of the 1822 point-sources in the entire field), 875 of which were detected by Chandra at exposures larger than 30 ks. These X-ray detections comprise the C-COSMOS Point-Source catalog by Elvis et al. (2009). The remaining 70 XMM sources, that were not detected by Chandra, are very faint hard-band detections which belong to regions that have only 30 to 50 ks exposures by Chandra. 25 sources detected by both XMM and Chandra were resolved into two distinct objects by Chandra due to the larger FWHM of the XMM PSF, thus they were excluded from the final catalog. The remaining 1797 X-ray point-sources were then subjected to cross-identification in optical bands.

In order to identify X-ray sources with the optical I-band counterparts, the statistical method of likelihood ratios was used. This method is described in details in Brusa et al. (2007). It computes the ratio of probabilities that a match is correct by weighting the distance between X-ray and optical detections, the surface density of false background objects and the brightness ratio of the matched pair. The resulting identification are divided into 3 classes: "reliable IDs", "ambiguous IDs" and "not identified" with 80%, 10%, 10% of sources in each class. After that, the positions of the I-band detections were matched to the positions in the K-band and IRAC catalogs. Most of the sources which were not identified in the I-band turned out to have bright K-band and/or IRAC counterparts, and were transfered to the class of "reliable IDs". On the contrary, some objects were changed from being reliably identified to ambiguous matches due to the uncertainty of their NIR and IR matched positions. The final numbers of sources in each class are: 1458 reliably identified sources (81.3%), 319 questionable matches (17.7%) and only 20 non-identified objects that constitute 0.9% of the total number of X-ray detections. To confirm positions of the XMM-detected sources, the identification has been also done using highly accurate Chandra X-ray positions in the overlapping area of two Surveys. We summarize all the results of cross-identification in the Table 2 taken from Brusa et al. (2010).

The resulting 1797 sources constitute the X-ray point-source catalog, which we used to select our AGN sample. To study host galaxies of high-redshift AGNs we selected all X-ray-detected point-sources with redshifts $z \geq 1.5$. At this redshift the HST F814W filter corresponds to the rest-frame UV, which is schematically shown on the Figure 23. Assuming that we are capable to properly remove AGN continuum of each object in our sample, we can therefore probe the emission of star-forming regions of the AGN host galaxies. Recent studies suggest that host galaxies of luminous AGNs have enhanced rates of star-formation in comparison to non-AGN galaxies (Silverman et al., 2009; Santini et al., 2012), and we were

Sample	Total Sources	Reliable (%)	Ambiguous (%)	Undefined (%)
Chandra area (before Chandra check)	850	712 (83.8%)	135 (15.9%)	3 (0.3%)
Chandra area (after Chandra check)	850	829 (97.5%)	21 (2.5%)	0 (0.%)
Chandra area (before Chandra check)	1797	1458 (81.1%)	319 (17.8%)	20 (1.1%)
Chandra area (after Chandra check)	1797	1577 (87.7%)	203 (11.3%)	17 (1.%)

Table 2: Optical identification of the XMM point-sources. Table is taken from Brusa et al. (2010).

expecting a significant rate of successful detections given that we will be able to separate the AGN rest-frame UV emission from the host galaxy emission. We also created a sample of quiescent non-AGN galaxies, matched by their absolute magnitude and redshifts to the AGN hosts (after the hosts' properties were extracted) to compare their morphological properties and environments.

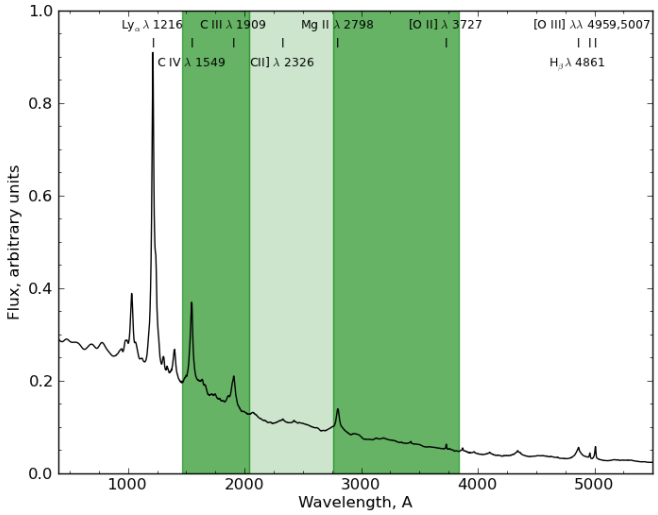

Figure 23: QSO composite spectra with spectral coverage by F814W HST filter, de-redshifted for the range $1.5 < z < 3.7$. Thick green filling marks the range of the de-redshifted filter for redshifts $z = 1.5$ and $z = 3.7$. The main QSO emission lines are marked.

6.1 QSO sample

We aimed at studying the host galaxies and the environments of the medium- to high- luminosity AGNs around the peak of the AGN luminosity function at redshift $z \sim 2$. At this epoch the most active SMBH growth takes place, providing perfect conditions to study the growth of the stellar component in galaxies. We also wanted

to compare morphological properties and environments of the AGN hosts to non-AGN galaxies to see if there are any prerequisites for the activity in these galaxies.

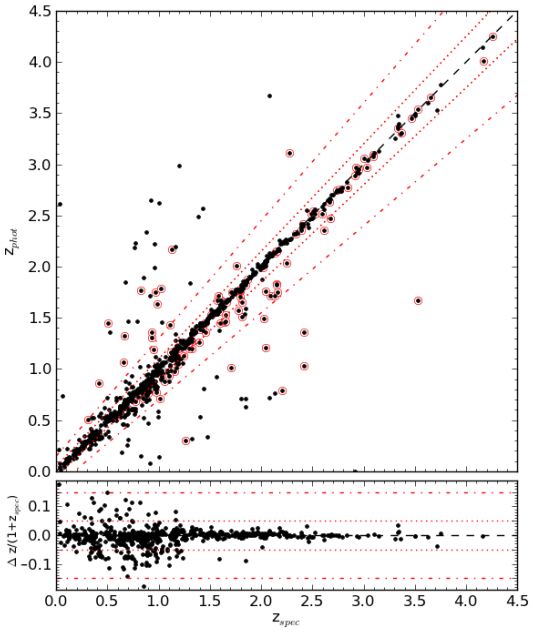

Figure 24: Comparison between spectroscopic and photometric redshifts for a sample of 844 objects from Brusa et al. (2010). The black dots encircled with red are objects with $I_{AB} > 22.5$. The black dashed line shows the $z_{\rm spec} = z_{\rm phot}$ relation, the red dotted line shows $z_{\rm spec} = z_{\rm phot} \pm 0.05(1+z_{\rm spec})$ and the red dash-dot line is $z_{\rm spec} = z_{\rm phot} \pm 0.15(1 + z_{\rm spec})$.

We applied a redshift cut of $z \geq 1.5$ to the X-ray point-source catalog of Brusa et al. (2010) to select a sample of high-redshift AGNs of both Type I and II. The division into classes is based on the level of X-ray absorption, that can be approximated by the X-ray hardness ratio of AGNs, defined as $HR = \frac{H-S}{H+S}$, where H is the hard X-ray (2-8 keV) band photon count and S is the soft X-ray (0.5-2 keV) band photon count (Wang et al., 2004): Type I AGNs have $HR < -0.2$, which corresponds to Hydrogen column densities of $N_H \sim 10^{23}$ at $z > 1.5$, and Type II have $HR > -0.2$. We performed the selection using spectroscopic redshifts when available, and photometric redshifts otherwise. After applying the redshift cut we obtained a sample of 627 X-ray point-sources, of which 249 (39.7%) have spectroscopic redshifts available. For the rest of objects we adopted photometric redshifts from SED fitting by Salvato et al. (2009). To determine photometric redshift for each object the authors performed SED fitting, with the SED constructed from up to 30 photometric data-points. They used a library of 30 SED templates, suitable for identifying stars, galaxies, star-forming galaxies and various types of AGNs. The solutions for the SED type, photometric redshift and intrinsic reddening were computed with the LePhare code (see Arnouts & Ilbert (2011) for the latest version of

the code), that uses χ^2 minimization to derive the best-fit parameters. The accuracy of the derived photometric redshift measurements with this code is $\frac{\sigma_{\Delta z}}{(1+z_{spec})} \leq 0.019$ for $I_{AB} < 24.5$ for both Type I and Type II AGNs up to the redshift $z \leq 4.5$ with only $\leq 6.3\%$ outliers. Figure 24 shows the relation between the spectroscopic and photometric redshifts for the 844 X-ray detected galaxies from the Brusa et al. (2010) catalog, which have both redshift measurements available.

Figure 25 shows the spatial distribution of objects across the 1.64 deg^2 COSMOS field as it was observed by the ACS/HST. ACS tiles are drawn with grey squares. We show the spatial distribution of high-redshift QSOs with $z > 1.5$ in the COSMOS field with the following the color-coding: blue stars show AGNs with resolved hosts we will report in this study (N=249), red filled circles are AGNs for which we could not resolve host galaxies (N=97), and empty red circles denote AGNs which were excluded from the analysis. The RA and dec coordinates of each of the 627 sources in the XMM point-source catalog by Brusa et al. (2010) are the coordinates of the optical/IR detections in the Capak et al. (2007) photometric catalog, matched to XMM and Chandra detections. We calculated pixel positions of these 627 AGNs in each tile using standard World Coordinate System (WCS) conversion and cut out a 108×108 pixel^2 postage-stamp image for each object, centered on the AGN. 81 AGNs turned out to be located outside of the ACS mosaic coverage and another 62 objects were excluded due to an initial magnitude cut of $I = 25.5$ mag, leaving 484 AGNs in the sample. After that, we visually checked all images to clean the sample from image artifacts, such as residual cosmics and bad pixels close to the borders of tiles. We also excluded those AGNs which had bright/saturated nearby stars (in total ~80 AGNs). During visual check we additionally excluded 66 AGNs which appeared too faint for the surface-brightness fitting, setting the final magnitude cut to $I \lesssim 25.0$ mag AB.

After the visual check was done, the final sample totalled 346 AGNs, with $I \lesssim 25$ mag AB. Several AGNs in the sample were observed in two adjacent tiles, so we used pairs of images to compare the consistency of fits for a given AGN. For each AGN cutout, which contained background stars or galaxies we manually created a mask to be used for background subtraction with the curves of growth method and for the surface-brightness profiles fitting.

All the AGNs in our sample are bright in X-ray as shown in Figure26, with luminosities $L_X > 10^{43}$ erg s^{-1} in the soft (0.5-2) keV band and $L_X > 10^{43.7}$ erg s^{-1} in the hard (2-10) keV band. Our final sample is dominated by unabsorbed Type I AGNs ($N_I = 191$) and Type II account for only 44 out of 235 objects that have detections in both X-ray bands. Among the 346 objects in our sample 211 were targeted spectroscopically, and 13 of those appeared to be narrow line (NL) AGNs.

6.2 Quiescent galaxies sample

In order to compare properties of AGN hosts to non-AGN galaxies, we created a sample of quiescent galaxies matched in absolute magnitude-redshift space to our AGN host sample. We computed the absolute magnitudes of the AGN hosts after

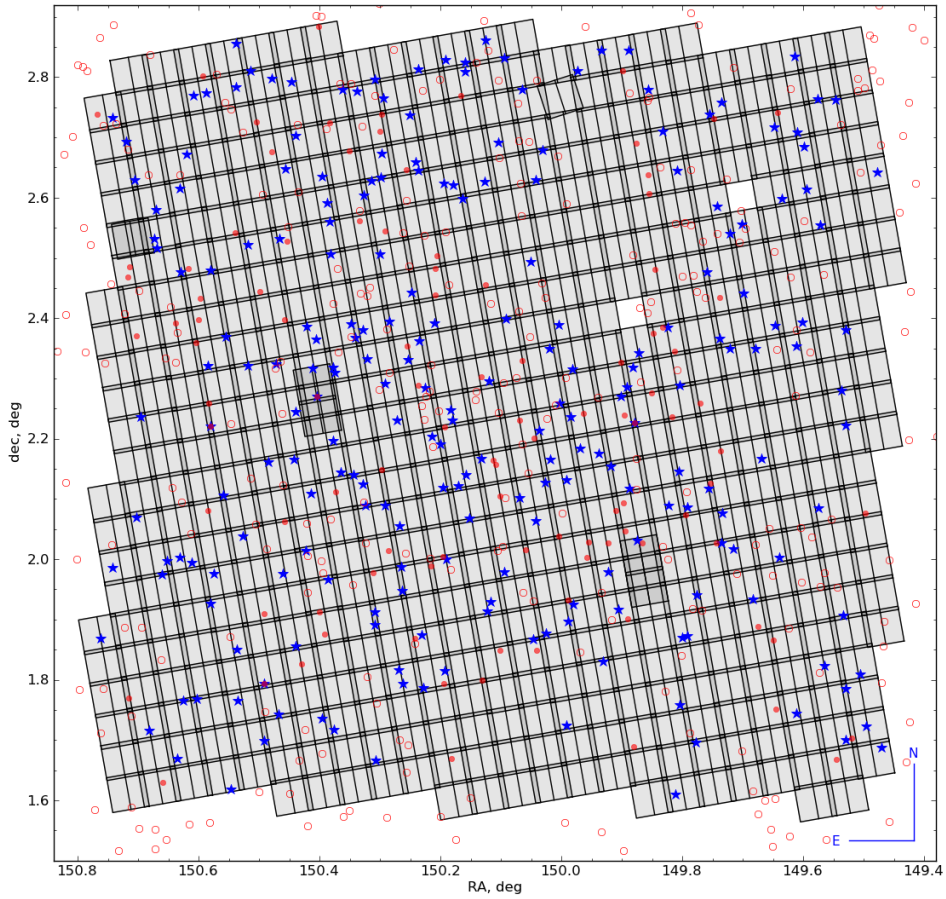

Figure 25: Map of the COSMOS field with its coverage by the ACS/HST mosaic. Blue stars mark positions of $z > 1.5$ AGN with resolved host galaxies (249 objects), red filled circles are $z > 1.5$ AGNs with unresolved hosts (97), and red empty circles mark positions of $z > 1.5$ AGNs which were excluded from the analysis (total number of 295: 81 are outside of the mosaic coverage; ~120 AGNs were excluded due to a magnitude cut of $I \lesssim 25.0$ mag AB; the rest were excluded due to image artifacts and proximity to the borders of the tiles).

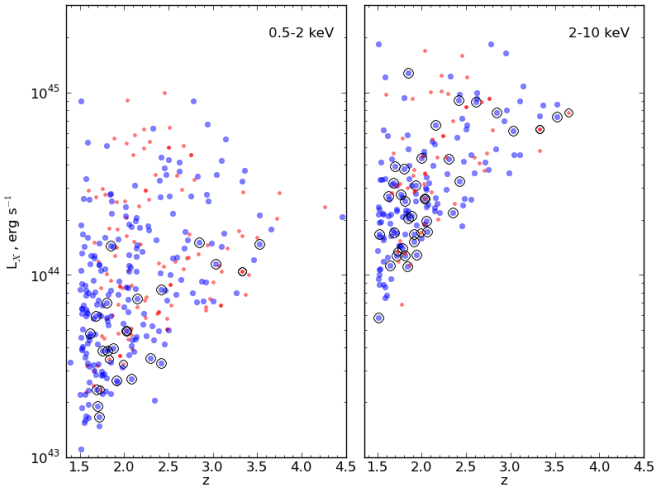

Figure 26: X-ray luminosity-redshift planes for the complete sample of QSOs: the left panel shows the soft X-ray band (0.5-2 keV), the right panel the hard X-ray band (2-10 keV). Blue circles are AGNs with resolved host galaxies, and red circles are AGNs which do not have a reliable detection of a host. Black circles denote Type II AGNs in both subsamples.

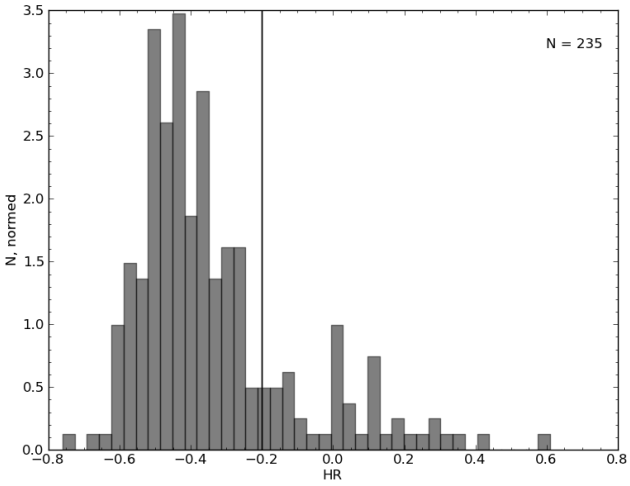

Figure 27: The normalized distribution of X-ray hardness ratios for our sample of QSOs (235 objects which have detections in both hard and soft X-ray bands). The vertical line at HR=-0.2 represents the commonly adopted threshold for Type I/Type II division. Sources detected only in the soft (0.5-2 keV) band or hard (2-10 keV) band are not shown.

completing the brightness profiles fitting, which will be described later in Sections 9 and 11. To select all the quiescent galaxies in the field we matched the photometric redshifts catalog by Ilbert et al. (2009) and the ACS I-band photometry catalog. This provided us with a sample of $I < 25.0$ objects, classified from the SED fitting as galaxies, with photometric redshifts estimates and apparent I-band magnitudes available for all objects in this parent sample. From this sample we selected all galaxies with redshift $z > 1.5$. To avoid AGN contamination we matched this catalog to X-ray point-sources catalog of Brusa et al. (2010), which we used to select AGNs. The resulting parent sample consists of about $\sim$ 22000 objects.

We created a sample of quiescent non-AGN galaxies, selecting the four nearest matches for each AGN host in the quiescent galaxies parent sample in the rest-frame absolute magnitude-redshift space. To do this we converted all the apparent $F814W$ magnitudes of quiescent galaxies into absolute rest frame magnitudes using the conversion formula $M = m_{rf} - 5((log_{10}D_L) - 1) - K$, where D_L is the luminosity-distance in the units of parsecs, m_rf is the apparent rest-frame AB magnitude (for the details see Section 11.1), and K is the K correction. We assumed $K = 0$ for the star-forming galaxies at redshift $z > 1.5$. This will be explained later in Section 11. The luminosity-distance of every galaxy in the galaxies parent sample was computed using Ned Wrights cosmology calculator (Wright, 2006), using standard cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$.

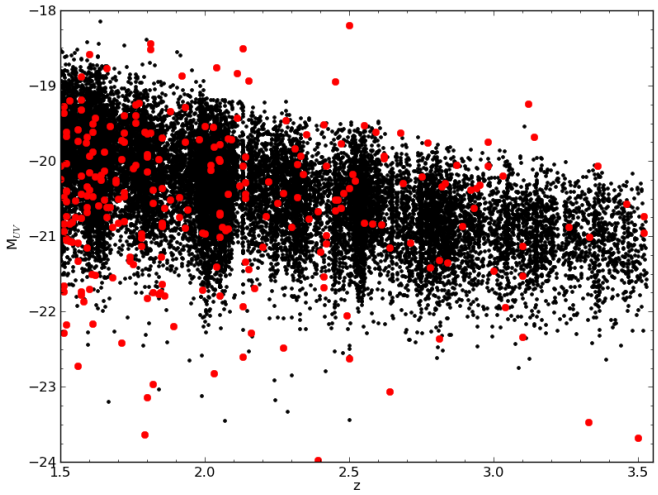

Figure 28: The distribution of the quiescent galaxies in the parent sample and estimated properties of AGN hosts in the rest-frame UV absolute magnitude-redshift space. Black dots are quiescent galaxies, red circles are recovered AGN hosts. For galaxies photometric redshifts are from the SED fitting by Ilbert et al. (2009).

Figure 28 shows the parent sample of galaxies and recovered hosts in the absolute magnitude-redshift space. For each AGN host we selected the four nearest galaxies, that were not more than 0.05 units away from the AGN host in this magnitude-redshift space, and identified their pixel-positions in the ACS tiles the

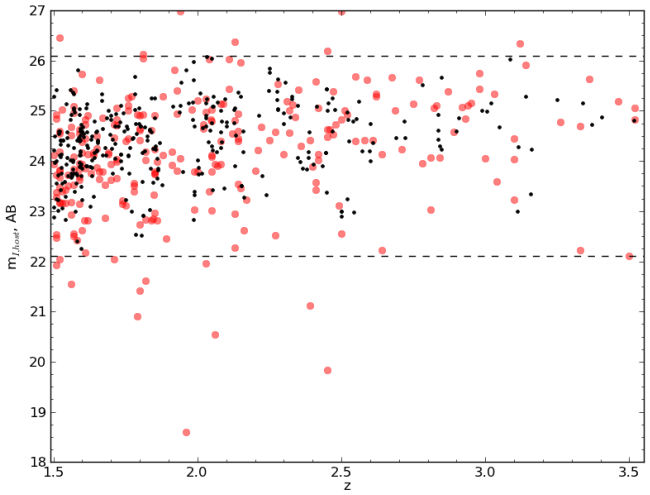

Figure 29: The distribution of the AGN hosts and matched quiescent galaxies in the F814W observed magnitude-redshift space. The horizontal dashed magnitude limits mark the adopted cuts, imposed by the lack of matching objects outside this range.

same way we did it before for AGNs. After that we performed visual check to remove possible false identifications, data-reduction artifacts, and very low surface brightness objects. We made sure that in our sample of AGN hosts, every entry has at least one matching quiescent galaxy, unless there were no close matches. One can see that the AGN hosts sample has some objects with very faint recovered absolute magnitudes which did not have matches in the quiescent galaxies due to the magnitude cut off at $I < 25.0$. Hence, we got a final sample of $\sim$ 300 quiescent galaxies, some of which were imaged in several overlapping tiles. Figure 29 shows the final sample of quiescent galaxies (plotted in black) and the sample of AGN hosts in the apparent magnitude-redshift space. The horizontal dashed lines mark the magnitude range where we could find matching pairs for AGN hosts. A 2D Smirnov-Kholmogorov test performed on the final sample of quiescent galaxies and (reduced) sample of AGN hosts, excluding hosts which did not receive a matching galaxy, showed that they are consistent with being drawn from the same parent distribution. At the low-redshift end of this distribution, however, there is an excess in the number of matches for each AGN host, which does not affect the statistics in our results.

Finally, for all the galaxies we selected for comparison, we performed the background subtraction and masked out the background objects using the same procedure that we employed for the AGN sample.

7 HST Data reduction

The HST provides a unique opportunity to study distant galaxies with exceptional resolution and sensitivity, unrestricted by the Earth's atmosphere. Extragalactic surveys like the Hubble Deep Fields (North and South) (Williams et al., 1996, 2000), GOODS (Giavalisco et al., 2004), UDF (Beckwith et al., 2006), GEMS (Rix et al., 2004), COSMOS (Scoville et al., 2007) made it possible to study galaxy evolution out to redshift $z = 6$. In particular, the COSMOS Treasury Survey covers the largest field ever imaged with HST, which has created large statistical samples of galaxies down to $m_{I,AB} = 28$ for point-source depth and $m_{I,AB} = 26$ for galaxies at a 5σ significance.

COSMOS was observed using the Wide field Chanel (WFC) of the Advanced Camera for Surveys (ACS). The WFC detector employs two 16-Megapixel CCDs, with a field of view of 202×202 arcsec^2 and native pixel-scale of $0.05''$/pixel, tilted by 22°with the respect to the focal plane of the telescope. The large field of view and off-axis location cause large geometric distortions in the images taken with this instrument and special care must be taken to correct them. In this Section the adopted strategy for reducing ACS observations is discussed. We will also discuss the effect of Charge Transfer Efficiency on ACS observations.

The main goal of this section is to review the main basics of MultiDrizzle algorithms to familiarize the reader with the technical background of reducing raw ACS data obtained with the diffraction-limited system of the HST[1][2].

7.1 Pros and cons of Dithering for the HST observations

The COSMOS HST Treasury Survey (see Scoville et al. (2007) for description) was granted 640 orbits during HST Cycles 12 and 13, 575 of which were dedicated to the COSMOS Evolution Survey (and the rest for a Supernovae search). A sky area of $\sim$ 1.64 deg^2 was observed with F814W (I-band) filter of WFC of ACS. Each pointing employed a 4-point dithering pattern, using 507 seconds exposures per dithering step. Such a strategy for observations is particularly useful for combining data in a way to improve sampling of HST's point spread function (PSF). A PSF is properly sampled when it satisfies the Nyquist criterion: the full width at half maximum (FWHM) of the PSF should be at least two resolution elements (pixels). In the case of ACS, the native pixel-scale of its detector is $\sim 0.05''$/pixel and the diffraction limit is $\sim 0.1''$. Thus, such marginal sampling does not provide sufficient quality of data, suitable for high-precision photometric measurements. Therefore, ACS PSF sampling requires improvement. It is possible to restore losses due to low-sampling and increase the spatial resolution of the data using the combination of shift-and-add techniques and interlacing, which are enabled in the drizzling algorithms for combining dithered data (Fruchter & Hook, 1997).

[1] http://www.stsci.edu/hst/HST_overview/documents/multidrizzle
[2] http://www.stsci.edu/hst/HST_overview/documents/datahandbook

However, although dithering and drizzling are very useful tools to perform linear reconstruction of pixels, applying dithering to data has also some drawbacks, which observer should keep in mind when choosing the observing strategy. One of those it is the length of the available exposure time: if one aims at certain depth of photometric data with limited time available, splitting exposure into parts will result in increasing the read-out time and re-pointing the telescope, which will shorten the effective science exposure time due to observational overheads. The choice of observing strategy also depends on the importance of spatial resolution over photometric accuracy or vice versa: due to intra-pixel sensitivity deviations in the CCD the time-dependent monitoring of the photometric time-series in the presence of dithering can become more complicated. Besides, since ACS is situated off the primary optical axis of HST, ACS images have very large geometric distortions of up to 8% across the WFC detector area. The projections of pixels on the sky change shapes from squares to rhombuses, with interior angles changing by up to 5° from the right angle depending on the position on the detector. Therefore, to combine dithered images with arbitrary offsets one needs to use an algorithm which is able to deal with data subjected to large geometric distortion.

There are various patterns in which one can employ dithering, depending on the aim of a project. For the COSMOS survey a "Full" dithering scheme was used. It is a 4-point dithering scheme, providing: (1) $\frac{1}{2}$-pixel sub-sampling along both WFC ACS detector axes, which is sufficient to restore the image resolution and all the losses due to PSF undersampling and (2) integer pixel shifts, large enough to cover the gap between the detectors and remove hot pixels and columns. This scheme is shown on Figure 30. Dividing the total exposure time into four parts (of 507 sec per dithering step), this dithering scheme also provides an efficient base for removing cosmic rays. Cosmic rays impact on average $\sim$ 4% of the detector pixels during a $\sim$ 2000 seconds exposure, therefore splitting a long 2028 seconds exposure into 4 parts of 507 seconds each will ensure less than $<$ 1 pixel impact by cosmic rays occurs during all 4 exposures according to a statistical binomial probability.

As for the photometric accuracy, it was mentioned above that HST detectors have sensitivity variations across individual pixels, which are hard to take into account. Even though this deviations are not very significant, they may cause difficulties for studies which require precise knowledge of brightness distribution, such as in weak gravitational lensing studies. This restriction also demands observers to be cautious when choosing dithering strategy.

7.2 Drizzling as a tool to combine dithered data

Drizzle was created in 1997 (Fruchter & Hook, 1997) and was first applied for reducing Hubble Deep Field data. It is designed to combine advantages of interlacing and shift-and-add algorithms in a presence of both $\frac{1}{2}$-pixel and large integer-pixel dithering shifts in case of undersampled distorted data. The engagement of these two algorithms is determined by how well a PSF can be sampled with the shifts

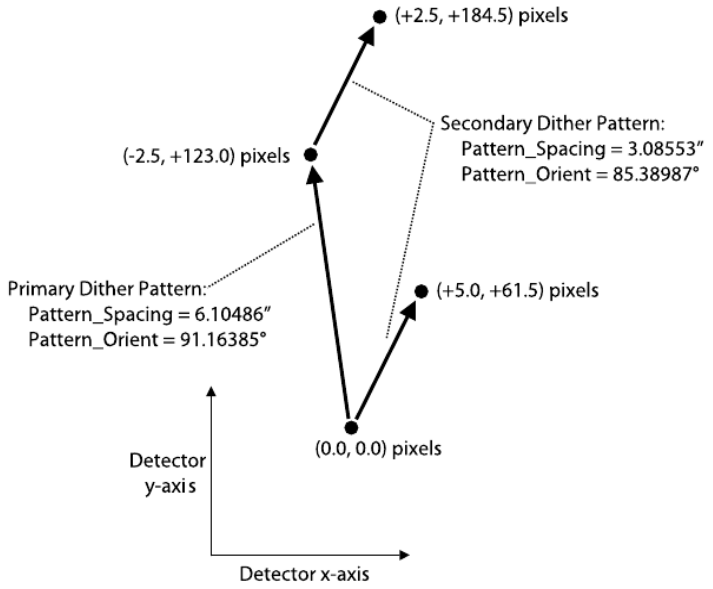

Figure 30: Scheme of the "Full" 4-point dithering scheme used for COSMOS observations. Image is taken from Koekemoer et al. (2007).

of employed dithering pattern. The global drizzling algorithm is basically controlled by the *pixfrac* parameter, which in the two extreme cases of *pixfrac=0* and *pixfrac=1* corresponds to pure interlacing or pure shift-and-add image transformation techniques, respectively.

A good way to visualize the case of *pixfrac=0* is to imagine a CCD image as a footprint of the sky, convolved first with optics of the telescope and then with a pixel response function (which ideally is a square of the size of a pixel, but in reality is distorted to the shape of a rhombus), and then sampled with a delta function at the center of each pixel. This case is basically pure interlacing, when output sub-sampled pixels are combined from the corresponding input pixels of dithered images. Even though interlacing gives an advantage of increased resolution, it is not feasible to use it as a stand-alone recipe. This is mainly, because it requires very precise dithering offsets, which even if the pointing accuracy of the telescope was ideal, would not be feasible due to geometric distortion of the data.

The other extreme case of *pixfrac=1* corresponds to pure shift-and-add algorithm. Each input pixel of the dithered exposure is transformed into a block on a finer sampled grid and shifted to its place according to the dithering offset on the output image. This algorithm has the advantage of being able to deal with arbitrary shifts, but it introduces additional blurring in the output image and correlation in noise due to convolution with the original pixels.

Drizzling allows the user to combine both of these techniques. Its algorithm is rather straightforward. Each pixel of every dithered exposure should be regarded

as an element of a combined higher-resolution image, convolved with a PSF and a pixel-response function of the CCD. “Old” pixels from the dithered exposures are shrunk by a factor in the range 0 to 1 given by the value of the *pixfrac* parameter. These shrunk pixels are called “drops” and are mapped onto the new sub-sampled output pixel grid, with the pixel positions determined by the shifts and rotations in the dithering pattern, taking into account optical distortion of the camera. This is schematically shown in Figure 31.

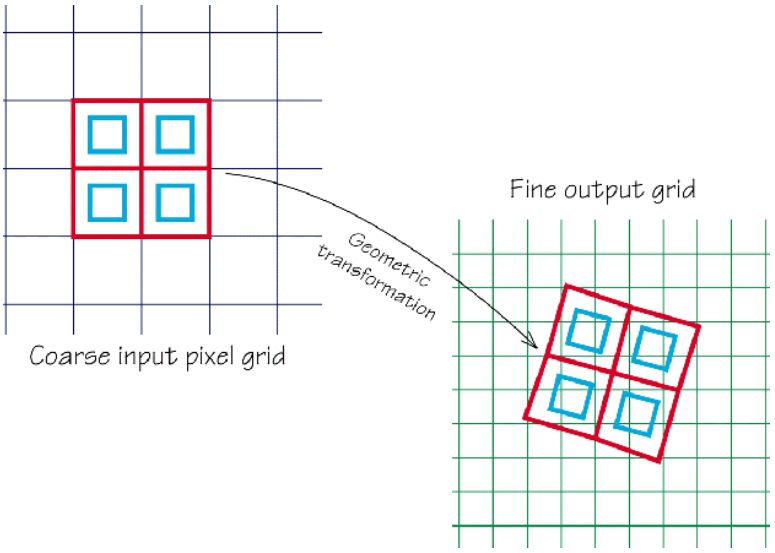

Figure 31: Scheme of the drizzling algorithm. Image is taken from MultiDrizzle Handbook (Fruchter & Hook, 2002).

The value of the output pixel in the final image is assessed from the values of the input pixels, weighted proportionally to the overlap area between these drops and output pixels. If the size of a drop is small enough, then not all output pixels will have contribution from every input image, thus avoiding degradation of the image sampling, unlike in the case of a pure shift-and-add algorithm. Therefore, one must ensure that the size of the drop is not too large, but, on the other hand, large enough so that the coverage on the output image is reasonably uniform. The degree of sub-sampling is controlled by the pixel *scale* parameter, which is the linear size of the output pixel in the units of “arcsec/pix”. Mathematically, drizzling can be described as follows. When the pixel (x_i, y_i) from the input image i (with a value d, and a weight w), is drizzled onto the pixel (x_o, y_o) of the output image o (with the value I and the weight W), and the fractional pixel overlap area between them is $0 < a_{x_i,y_i,x_o,y_o} w < 1$, the resulting values of the output pixel and its weight I' and W' are calculated as:

$$I' = \frac{d a_{x_i,y_i,x_o,y_o} w s^2 + IW}{W'}$$

$$W' = a_{x_i,y_i,x_o,y_o} w + W,$$

where s^2 is a factor introduced to conserve the surface intensity. During drizzling, these formulas are applied to every pixel of every dithered image.

When drizzling divides the power of the input pixel into several output pixels, the noise in this adjacent pixels will be correlated. It is important to have numerical estimates of this correlated noise and create proper variance maps to be used by GALFIT. Due to such correlations, measurements of noise in the output image are underestimated to compare with original pixels. This is illustrated in Figure 32.

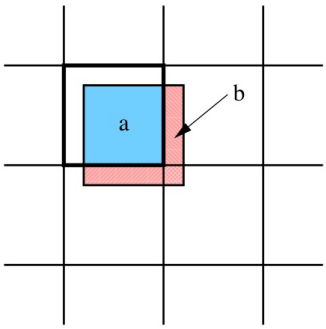

Figure 32: Scheme of the distribution of noise between adjacent pixels in the output drizzled image. Image is taken from MultiDrizzle Handbook (Fruchter & Hook, 2002).

The original pixel (shown in color) is drizzled onto the output pixel grid overlapping with 4 output pixels. If we designate the overlapping area between the original and primary output pixels as a (shown in blue) and those with other three pixels as b_1, b_2, b_3, then $b = b_1 + b_2 + b_3$ and $a + b = 1$. Then, if the noise of the original pixel is ε, the total variance added to the output image is $(a + b)^2\varepsilon^2$. However, if we sum up the input of the variances added pixel-by-pixel, it will be $(a^2 + b_1^2 + b_2^2 + b_3^2)\varepsilon^2$, which is less than $(a+b)^2\varepsilon^2$. Therefore the total noise power in the output drizzled image is underestimated by an amount of:

$$\begin{aligned}(a+b)^2 - (a^2 + b_1^2 + b_2^2 + b_3^2) = \quad & (a + (b_1 + b_2 + b_3))^2 - (a^2 + b_1^2 + b_2^2 + b_3^2) \\ = \quad & 2\,(ab1 + ab2 + ab3 + b_1b_2 + b_1b_3 + b_2b_3).\end{aligned}$$

These terms represent the correlated noise in the image and can sum up to a significant value. For an image of $N \times N$ pixels the noise correlation ratio is defined as

$$\lim_{N\to\infty} \frac{a^2 + b_1^2 + b_2^2 + b_3^2}{(a + b_1 + b_2 + b_3)^2} = R$$

and it indirectly represents the quality of the performed drizzling. This value strongly depends on the data reduction parameters *pixfrac* and *scale*, that were used to drizzle-combine the data, as well as on the dithering pattern used for observations. Also, given that the distortion of the ACS WFC detector changes across the field of view, the noise correlation ratio can vary across the image.

7.3 The geometric distortion and its consequences for the drizzling

The large geometric distortion of images taken with ACS makes a significant impact on all aspects of data reduction, starting from selection of dithering strategy, to photometric and astrometric accuracy of the calibrated images. The WFC detector is mounted $\sim 6'$ off the primary optical axis of the telescope and is tilted by 22°with the respect to its focal plane. This results in 8% elongation along the detector diagonal, which corresponds to the largest distortion displacement of ~ 82 pixels≈ 4 arcsec. Figure 33 shows a plot taken from the MultiDrizzle Handbook depicting a non-linear component of the ACS distortion modeled with a quadratic fit to data taken with the F475W filter. The arrows indicate the direction of the distortion and their amplitudes are larger by a factor of 5 with respect to the axes scale.

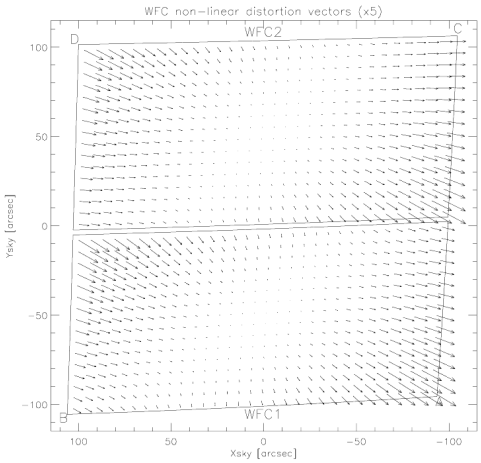

Figure 33: ACS WFC distortion pattern over the entire FOV of the WFC camera. The direction of the arrows marks the direction of elongation and the amplitude of arrows is 5 times overstated. Image is taken from MultiDrizzle Handbook (Fruchter & Hook, 2002).

Such a large distortion causes a change in the degree of sub-sampling across the image. Even a relatively small 5-pixel dithering shift at the center of the field of view transforms into additional 0.4 pixel shift at the edges of the detector. In order to produce more uniform sub-sampling several offsets have to be done, as for example the 4-point dithering scheme of the COSMOS Survey.

Geometrical distortion of the WFC detectors can be modeled with a polynomial

of the order "m" relative to some reference pixel (x_0, y_0):

$$X = \sum_{i=0}^{m} \sum_{j=0}^{i} a_{ij}(x - x_0)^j (y - y_0)^{i-j} =$$

$$= a_{00} + a_{10}y + a_{11}x + a_{20}y^2 + a_{21}xy + a_{22}x^2 + a_{30}y^3 + a_{31}xy^2 + a_{32}x^2y + a_{33}x^3 + \ldots$$

$$Y = \sum_{i=0}^{m} \sum_{j=0}^{i} b_{ij}(x - x_0)^j (y - y_0)^{i-j} =$$

$$= b_{00} + b_{10}y + b_{11}x + b_{20}y^2 + b_{21}xy + b_{22}x^2 + b_{30}y^3 + b_{31}xy^2 + b_{32}x^2y + b_{33}x^3 + \ldots$$

where (X, Y) and (x, y) are pixel coordinates in the undistorted and distorted frames respectively. The linear component of the distortion can be reduced to

$$X = a_{11}x$$

$$Y = b_{10}y + b_{11}x.$$

To better understand the origins of such reduction, the projection of the detector on the undistorted coordinate system (X, Y) are shown in Figure 34. The a_{00} and b_{00} components vanish, because both distorted and undistorted coordinate systems have the same origin. The a_{10} component is also zero, because the undistorted Y-axis is directed along the distorted y-axis of the detector.

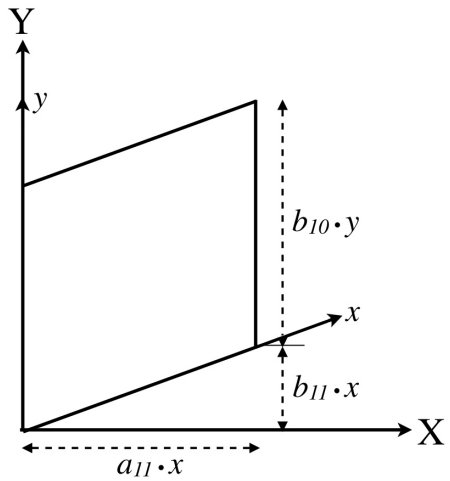

Figure 34: Projection of the *x,y* distorted coordinates onto "real" *X,Y* coordinate system in terms of linear components of the distortion polynomial. Image is taken from MultiDrizzle Handbook (Fruchter & Hook, 2002).

The distortion coefficients for higher order components of every instrument detector on-board HST are stored in a special reference file, called "The Instrument Distortion Correction Table", IDCTAB, which MultiDrizzle (and Tweakshifts) software refers to during a run to correct for geometric distortion of an

image. Moreover, ACS distortion is not stable over time due to changes in the telescope's orientation. The residual distortions have a linear time-dependent component, which can be account as:

$$\alpha = 0.095 + \frac{0.090(obs.date - 2004.5)}{2.5}$$

$$\beta = -0.029 + \frac{0.030(obs.date - 2004.5)}{2.5},$$

where *obs.date* refers to the observations date in units of decimal years. These coefficients are written to the fits header keywords *TDDALPHA* and *TDDBETA* of the calibrated file, which are used by MultiDrizzle to correct the WCS of the final co-added image.

The distortion model of ACS can be modeled to a rather good precision. The 4-th order polynomial has been well calibrated by the STScI instrument team and recorded to the *IDCTAB* to be subsequently used by MultiDrizzle for the geometric distortion correction. For convenience, the distortion correction coefficients are also included to the image headers using the Simple Imaging Polynomial (SIP) convention, which allows various visualization packages to display distortion-corrected positions in the frame. The usage of the SIP coefficients is indicated in the *CTYPE1* and *CTYPE2* keywords, which record the coordinate type for the x- and y- axes of the image.

After the astrometric registration and correction, each pixel coordinate is computed using a CD matrix, recorded in the image header. This matrix includes linear terms of the distortion model, as well as rotation, scale and skew:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} CD1\_1 & CD1\_2 \\ CD2\_1 & CD2\_2 \end{pmatrix} \begin{pmatrix} u + f(u,v) \\ v + g(u,v) \end{pmatrix},$$

where *u,v* are the original coordinates in the image, and *f(u,v)* and *g(u,v)* are nonlinear terms of distortion:

$$f(u,v) = \sum_{p,q} A\_p\_q u^p v^q, \; p + q \leq A\_ORDER$$

$$g(u,v) = \sum_{p,q} B\_p\_q u^p v^q, \; p + q \leq B\_ORDER.$$

A_p_q and *B_p_q* are header keywords for the polynomial terms $u^p v^q$, and *A_ORDER* and *B_ORDER* are keywords stating the order of the polynomial used for the correction.

7.4 Astrometric image registration

Another crucial question of dithering is the ability of the telescope to provide sufficient pointing accuracy for the repeatable observations, which determines the precision of the image alignment. Aligning dithered exposures within a single tile is

important for efficient cosmic ray rejection, as well as sub-sampling procedures, which were discussed above. Since each tile was observed during single orbit and the dithering offsets where smaller than $\sim 9'' = 180$ pixels, the telescope was retaining the same guiding stars during the entire orbit. HST usually requires two guiding stars to minimize tracking inaccuracies, so if both were acquired at the start of the orbit, they provided positional accuracy between dithering steps better than $\sim 2-3$ mas. Observers also have to make corrections in order to take into account the misalignment introduced by inaccuracy in the positions of the guiding stars: before Cycle 14, the HST Guide Stars Catalog did not take into account proper motions of guide stars, which could result in $\sim 1'' - 2''$ uncertainty of absolute astrometric accuracy. In order to correct these uncertainties we employed the Tweakshifts procedure. Tweakshifts is a Python code, wrapping IRAF-based tasks, which computes shifts between the WCS of two or more co-added images. It automates the procedure of producing a file, that contains computed shifts between coordinate systems of the combined images, that can be supplied to MultiDrizzle during data reduction process to ensure correct astrometric registration. There are two ways in which Tweakshifts can operate: through cross-correlation of images (suitable for observations which have no or too few point sources) or through matching images to a reference catalog, the option which we used for computing shifts for the COSMOS data. The latter is done through identifying point sources in the dithered exposures with a catalog (via daofind or SExtractor), and computing shifts from only certainly matched sets of objects, additionally applying distortion corrections to their estimated positions.

The WCS for each FITS image is stored in its header using the standard keywords: *CRVAL1* and *CRVAL2* for RA and dec coordinates of the reference pixel, *CRPIX1* and *CRPIX2* for the x,y pixel positions of the reference pixel. The plate scale information and rotation are stored in the CD matrix, which is computed as:

$$\begin{pmatrix} CD1\_1 & CD1\_2 \\ CD2\_1 & CD2\_2 \end{pmatrix} = scale \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} A10 & A11 \\ B10 & B11 \end{pmatrix},$$

where *scale* is the final pixel-scale of the undistorted image, θ is the value of the *ORIENTAT* header keyword, and the terms *A10, A11, B10, B10* are the linear terms of the distortion model from *IDCTAB* file.

7.5 Data reduction. Step 1: Initial calibration

The Multimission Archive at the Space Telescope Science Institute (STScI) (MAST) provides a possibility to download COSMOS HST data, which can be either raw, pre-reduced or already fully combined with MultiDrizzle. Since the fully processed data-products available in the archive have been reduced using the default deta reduction parameters (default pixel-scale of $0.05''/pix$ and *pixfrac=1*), which does not provide sufficient PSF sampling and stability, we downloaded the calibrated data in order to combine it with MultiDrizzle. These calibrated data is bias and dark subtracted, gain-corrected, flat-fielded and has all the bad pixels identified.

By querying data from the MAST, the user receives the *_flt_* files, which are calibrated and flat-fielded single dithering step exposures, that have not been corrected for distortions or cleaned from cosmic rays. Each *_flt_* file contains 7 extensions: the global header, two science extensions for both CCDs, two error arrays, with the RMS noise computed as

$$\sigma_{CCD} = \sqrt{\frac{(DN-bias)}{gain} + (\frac{readnoise}{gain})^2}, \quad (6)$$

and data quality arrays for each CCD, which indicate the locations of the bad pixels. The user also obtains the supplementary files, which are required by MultiDrizzle, and the association *_asn_* file, with a list of images associated with the same tile and the type of these observations (i.e. dithered data or CR-SPLIT observations). Calibration tables, used by *calacs* are delivered to match the requested data-set to be used by MultiDrizzle. Detailed information about calibration procedures and intermediate data types can be found in the ACS Data Handbook [1].

7.6 Charge Transfer Efficiency

There is one additional issue, which needs to be considered concerning the accuracy of photometric measurements on ACS data –the so-called Charge Transfer Efficiency (CTE) (or Inefficiency). This effect is a result of degradation of the ACS detectors due to charged particles. Every time a charged particle hits the CCD it creates a charge trap in the semiconductor, which can capture photo-electrons and release them only after some characteristic time. At the end of each exposure during the readout, electrons are being transferred, row by row, to the nearest amplifier (ACS has 4 amplifiers, located at the top corners of the top chip, and bottom corners of the bottom chip). When some of these electrons are trapped on their way to an amplifier, this results in their virtual "migration" to a pixel displaced from the real location by the distance determined by the time delay. Thus, CTE forms a trail behind every source in the image. The strength of this effect depends on the number of traps and the distance between the pixel and its nearest amplifier, as well as on the temperature of the CCD, responsible for the characteristic time after which trapped photo-electrons are released. This effect obviously gets stronger over time, since the number-density of traps grows with the time telescope operates. The strongest discrepancies arise in cases of faint sources and low sky background values, since in that case the source counts and the CTE counts are comparable and have a rather high contrast with the sky.

In general, it is hard to account for the above effect, because the amount of flux drawn away from the source depends non-linearly on the flux, size and shape of the source. But if one knows the trap density over the detector and the characteristic time by which these traps can retain electrons (which depends on the temperature of the detector), it is possible to model the CTE effect using a so-called forward readout process. To check how strong the effect of CTE for COSMOS data is,

[1] `http://www.stsci.edu/hst/acs/`

we applied the model for CTE correction (Massey et al., 2010) to a set of raw non-calibrated images associated with observations of a single tile, adopting the electrons trap densities from Massey et al. (2010). The IDL code is available for public download at the Richard Massey's web site[1]

The output of the IDL routine is a CTE-corrected non-calibrated image, which has to be run through *calacs* for calibration, and after that through the standard data-reduction MultiDrizzle pipeline. Figure 35 shows the residual charge trails, computed by subtracting the CTE-corrected drizzled data from the one without the correction. To estimate the impact of CTE on photometry, we ran both CTE-corrected and non-corrected images through the SExtractor to create catalogs of detected objects, and their aperture magnitudes (we used 8-pixel apertures). Figure 36 shows the difference between the photometry of matched sources in these two images, showing that as expected, accounting the CTE should result in brighter objects. However, this plot shows a significant number of sources of the faintest magnitudes become fainter. Visual check of these positive residuals showed that all of them are faint galaxies. Given that we expect the GALFIT uncertainties for the magnitude measurements to be larger than the uncertainties due to CTE for the faintest objects in this tile (0.1 mag), we believe it is not needed to perform a very time-expensive CTE correction for the entire field.

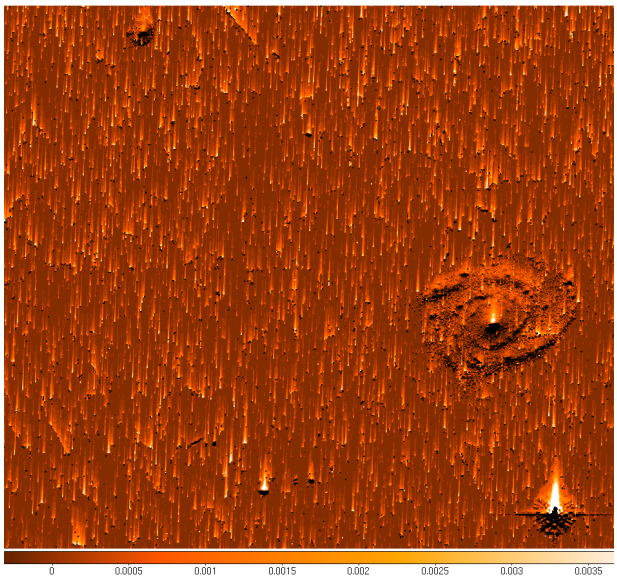

Figure 35: A cutout from the difference image between originally reduced data without CTE taken into account and data, to which Massey et al. (2010) CTE correction model was applied.

[1] http://www.astro.caltech.edu/ rjm/acs/

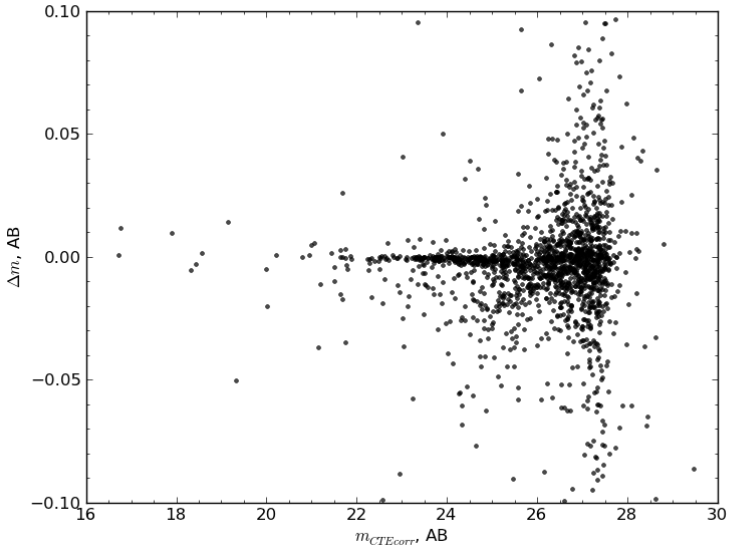

Figure 36: The difference in the photometric measurements of the matched sources in the CTE-corrected and original data.

7.7 Data reduction. Step 2: MultiDrizzle

MultiDrizzle is a fully automated pipeline needed to produce geometric distortion corrected, sky subtracted and clean from cosmic rays drizzled images, which may optionally be combined into a mosaic. To run MultiDrizzle, one requires 3 types of files to operate: (1) an association file, describing the observing set, which the user wishes to process, (2) fully calibrated $\_flt\_$ images and (3) two reference files, *IDCTAB* for geometric distortions correction, and *MDRIZTAB*, which contains the initial MultiDrizzle parameters.

MultiDrizzle performs a set of operations in order to acquire the final image: initialization, bad-pixel mask creation, sky subtraction, drizzling of separate frames, median image creation, blot image creation, cosmic-ray mask creation and then it combines the final image. Each procedure of the pipeline can be turned on and off, but the tasks are not independent. Once a step is completed, a sequencing step can be run over and over using different data reduction parameters. Therefore, MultiDrizzle requires temporary files, produced during early steps of the data reduction process, to be stored after the completion of each reduction step in order to allow the user to restart separate tasks. Each step has a variety of parameters that can be altered according to the requirements. In the following we are going to describe each step of the reduction sequence and some of its most important parameters, which we modified to improve the quality of the output data [1].

[1]Parameters which are not specified in the command line of MultiDrizzle, are either assigned their default value or simply skipped if they are not crucial for the MultiDrizzle. The complete list

Initialization At this step MultiDrizzle identifies the files it is provided as the input. The list of the *_flt_* images to be combined should be specified under the **Input** parameter. These images should be calibrated and belong to the same tile. The **updatewcs** switch should be set to 'Yes' in order to correct the WCS of the image for time-dependent geometric distortion, and populate the image header with the SIP keywords. To align dithered images, they should be matched with a reference catalog using *Tweakshifts*. It produces the **shiftfile**, which should be supplied for the MultiDrizzle. Figure 37 shows an example of a science extension of a fully calibrated ACS exposure from the COSMOS survey prior to the data reduction.

Static Mask Creation During the following step the pipeline creates a bad pixel mask, in which every pixel with a value, which differing from the mode by more than **static_sig***RMS, is flagged. Such bad pixels may originate due to the over-subtraction of bad pixels in the dark image, or too high sky levels estimated during pre-calibration. Thus, the **static_sig** keyword controls the quality of the rejection. If this value is set too high the mask may contain real objects. The final bad-pixel mask is combined from these flagged pixels and the data quality (DQ) mask produced during the calibration.

Sky subtraction During this step MultiDrizzle estimates the Sky background level in all images and subtracts it. To determine the sky background level, the pipeline performs iterative σ-clipping for each amplifier. At the first iteration it considers the full range of pixel intensities for a given amplifier readout region to determine the standard deviation σ of this distribution, and then it rejects all the values which differ from the mean of this distribution by the value specified in the **skylsigma** and the **skyusigma** keywords. After the **skyclip** iterations, the median, mean or mode (specified by the **skystat** keyword) of the pixel distribution is adopted as the sky value for corresponding amplifier.The global sky level over the image is then settled to the lowest sky level among four amplifiers.

Drizzling of Separate Frames At this step separate images get drizzled onto separate undistorted grid, which, if combined, would correspond to the final image product. This step is used to refine registration of the input images. Using the table with the distortion coefficients (IDCTAB) and the static mask, produced at the previous step, the input distorted images are mapped on the refined output pixel grid, applying distortion coefficients to derive the pixel positions in the undistorted grid. At this step the user has to specify the new drizzling parameters to improve the PSF sampling, namely the **driz_sep_pixfrac** the **driz_sep_scale**. The first parameter controls the linear size of the output pixel and the second parameter determines its absolute size. Another important parameter, which determines the kernel to be used

of parameters can be found in the MultiDrizzle Handbook

by MultiDrizzle for the pixels re-sampling is **driz_sep_kernel**. In Section 8 we are going to explain how the optimal data reduction parameters were determined.

Median Image creation Median image is created as an estimate for the final image. By comparing the corresponding pixels of the median image and separately drizzled frames, the pipeline improves the bad pixel registration and rejects cosmic rays. This frame is median-combined of aligned drizzled images, produced at the previous step of the data reduction.

Blot Image creation Once median image is created, MultiDrizzle transformed it back to original distorted coordinates. This distorted median image is called blot image and it is used for the direct comparison with original undistorted exposures to identify cosmic rays (CR).

Cosmic Rays Rejection For each of the four 507-seconds exposures $\sim 80000 - 250000$ pixels out of 4096^2 array were affected by CR. The 4-point dithering provides very efficient CR identification, with $\ll 1$ pixel affected simultaneously at all 4 exposures. To identify and reject cosmic rays in every image, the pipeline compares the blot image, created at previous reduction step, to original raw exposures. Rejection process is controlled by the $\sigma - clipping$, asserting the *RMS* of each pixel from its count rate, background sky counts and the CCD readout noise. To avoid confusion by bright stars during clipping, rejection is softened using a derivative image constructed from the median image in a way, that every image pixel has a value of the maximum gradient between the adjacent pixels. Thus, this derivative image is only significant for the pixels, which have large gradients with the neighboring pixels. A pixel is flagged as the cosmic ray if the difference between the input pixel and corresponding pixel of the blot image exceeds the sum of the σ-criterion, specified by **driz_cr_snr** keyword, and this gradient map, scaled by a factor specified in the **driz_cr_scale** keyword. As an output of this procedure the MultiDrizzle produces a bad pixel mask, which is used for the final image combining.

Final Image Combining At the final step of the data reduction, the input images are drizzled onto the output pixel grid, applying cosmic ray and bad-pixels masks. This process is controlled by a similar range of parameters as during drizzling of the separate images: by **final_pixfrac**, **final_scale** and **final_kernel** parameters, which we also set to *0.8*, *0.03 arcsec/pix* and *gaussian* correspondingly as in previous section. The final combined image contains 3 extensions, with the science, the weight and the context images, which maps exposures that contributed to the output image. Weight maps contain all of the noise components, including the estimated sky background level, the readout noise, the dark current and also the static bad-pixel masks created by MultiDrizzle.

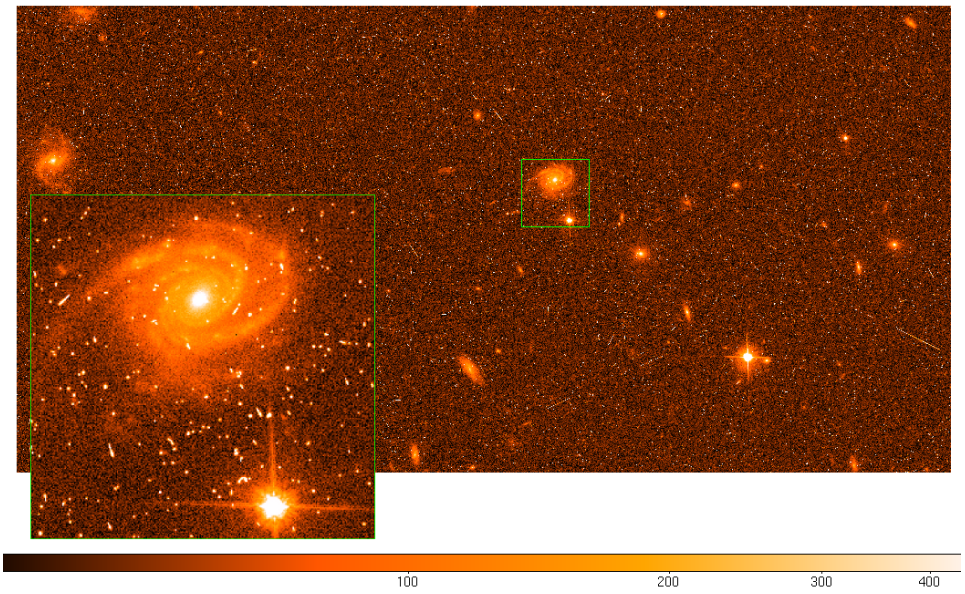

Figure 37: A single extension of an input ACS exposure from the COSMOS Survey, distorted and with the presence of numerous cosmic rays and bad pixels.

File Suffix	Description
_single_sci_	The individually drizzled, corrected for distortions single exposure
_single_wht_	The corresponding weight image of an individually drizzled single exposure
_single_mask*_	The bad-pixel mask before the CR-rejection is run; for each chip
_final_mask*_	Final bad-pixel mask, which includes CR; for each chip
_sci*_blt_	Blotted image; for each chip
med	Median image, median combined drizled images
_coeffs*.dat	Text file, extracted from the IDCTAB to correct for geometric distortions

Table 3: A list of intermediate data-products output by MultiDrizzle. "*" stands for the single chip number.

After all of the data-reduction steps were executed, using the alternative data-reduction parameters, we obtained a final image for each of the 575 tiles observed for the COSMOS Survey. These images present an improvement over the originals in terms of PSF sampling and image resolution, stability of the stochastic alignment of objects' ellipticities. The difference in the image quality is appreciable even to the naked eye, as can be seen in Figure 38

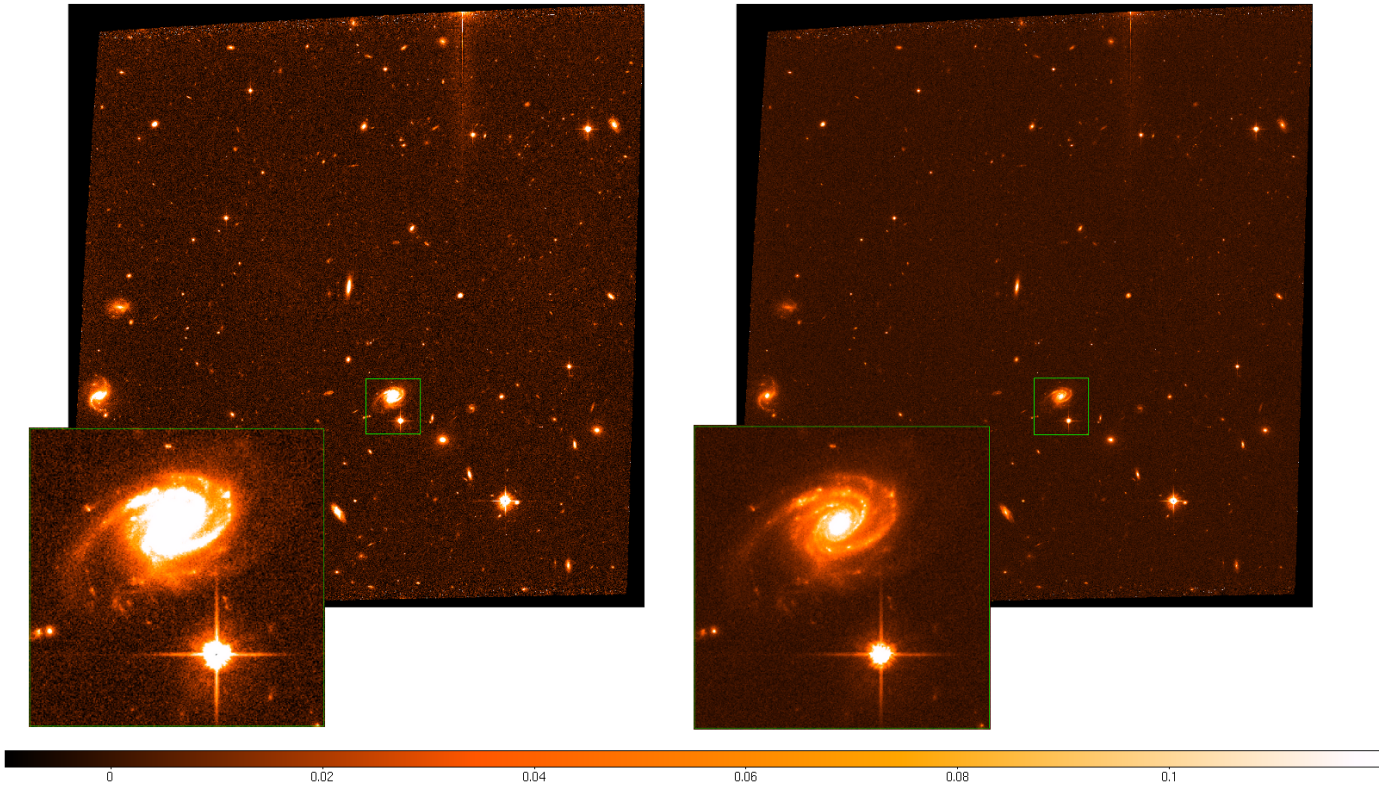

Figure 38: On the left: Original MultiDrizzle data reduction, with *pixfrac=1.0* and *scale=0.05 arcsec/pix* and *square kernel* as available in the HST Archive. On the right: Our data reduction with modified parameters: *pixfrac=0.8* and *scale=0.03 arcsec/pix* and *gaussian kernel* in order to improve PSF-sampling, noise, etc.

8 Point Spread Function for the ACS data

When the ACS was enabled for observations in 2002, it provided new opportunities for observational studies, which required high-resolution imaging and large fields of view. Even though the HST mirror is not big in comparison with the largest ground-based telescopes, the absence of the atmosphere brings it way ahead of the competition due to the better quality of the observations. The PSF for any diffraction-limited imaging camera of HST in any observational band is much narrower than that of a matching band of any ground-based telescope without AO-assistance. The width of the PSF is determined by the optical system of the telescope and the Pixel Response Function (PRF) of the detector; depending on the filter, it can reach as low as *0.05*, which is equal to the linear size of a pixel of ACS. But even though the PSF of ACS is so narrow and is free of the atmosphere impact, it has two variable components, which we briefly mentioned before in Section 3.2. On the one hand, the PSF changes over time due to variation of the distance between primary and secondary mirrors of the telescope, and on the other hand the PSF varies spatially due to the change in the degree of geometric distortion across the detector. We have already described the properties of the geometric distortion of the ACS in the previous section, which can be easily modeled and corrected for using observations of stellar fields (Meurer et al., 2003; Anderson et al., 2006). The time variability, caused by changes of the focus distance of the telescope (or so-called " breathing"), occurs due to temperature variations, when the telescope's optical system contracts as it goes in and out of the Earth's shadow; this is an intra-orbit variation occurring on the time scale of one orbit. The spatial magnitude of these variations can reach few microns and therefore grow the error in the focus estimates. On top of that, there also exists a slow oscillatory drift of the focus on the scale of several weeks, caused by factors like solar angle, target occultation and so on (Hershey & Taff, 1998; Heymans et al., 2005; Rhodes et al., 2006; Schrabback et al., 2007; Lallo et al., 2006; Makidon et al., 2006). The changes if focus are stochastic and can not be predicted beforehand, however they can be measured from the data itself after it has been acquired.

In the COSMOS Survey the observations were taken with the *F814W* filter, for which the width of PSF is as few as $FWHM \sim 0.1''$. This means that for the native pixel scale of the ACS WFC of $0.05''$/pixel the PSF is sampled with only 2 pixels. Such marginal sampling does not provide an opportunity to employ full capacity of the HST resolution. The best-available solution on how to improve the sampling of the HST imaging data employing drizzling techniques was described in the previous chapter. Therefore, in this chapter, we will assume that our data is properly sampled using the best-suitable data reduction parameters. We will discuss the issues of modeling the PSF, that would match data and provide an opportunity to de-blend brightness profiles of QSOs and galaxies in our samples.

Before discussing a particular case of creating a PSF for the ACS COSMOS Survey data, we would like to bring reader's attention to an example we already gave before in Section 3.2 on how the question of PSF creation was addressed in

the earlier studies, aimed at resolving host galaxies of AGNs on both low- and high-redshifts (Sánchez et al., 2004; Jahnke et al., 2004). Both of these studies were carried out within the GEMS Survey (Caldwell et al., 2008) , which was also observed with the ACS WFC employing two filters, F606W (V band) and F850LP (SDSS z band). Most of the GEMS survey was observed during a continuous interval of time in November 2002, except for 4 tiles observed a few months before and after the bulk of the data. Given that, one can assume that the focus value for the continuous observing run did not vary significantly, and neglect PSF time variability. The GEMS data were reduced with MultiDrizzle pipeline employing modified set of data-reduction parameters to increase the spatial resolution and improve the sampling of the data. To take into account spatial variability of the PSF profile due to the initial geometric distortion, the observer has to create a PSF for a given pixel position on the detector. In case of the high-redshift sample of AGNs from Jahnke et al. (2004) [1], the authors took advantage of the negligible PSF time-variability, employing all the non-saturated, de-blended stars in the field to create a catalog of their global pixel positions to be used to average empirical models of PSF. Thus, the density of stars in this joint catalog was sufficient to create a very local, high signal-to-noise PSF for any pixel position across the detector. A PSF model, averaged from 35 nearest stars located not further away than 35 pixels from the desired position, has sufficient locality to account for the spatial PSF variability. We have already described the detailed procedure of the empirical PSF creation in the Section 3.2 of this thesis.

Unfortunately in the COSMOS Survey we cannot assume negligible time-variability as it was made in Jahnke et al. (2004). The COSMOS Survey was observed during a much longer period of time and not in a continuous fashion. Therefore we could not adopt the strategy of assuming a constant focus distance and simply collecting all stars from all observations in order to use them to average PSFs for certain locations. Instead, we adopted the approach to model PSF as in the weak gravitational lensing studies made in the COSMOS field (Leauthaud et al., 2007, e.g). Due to a necessity to detected low-amplitude signal changes in weak lensing studies, that study requires precise knowledge of PSF in order to measure very small changes of galaxy shapes caused by intervening mass distribution. The best way to reach the required precision in a given situation is to use the Tiny Tim software[1]. It allows to model the PSF for any condition of any instrument of HST, taking into account focus distance and desired spatial location within the detector.

We followed the recipe given by Rhodes et al. (2007) which uses Tiny Tim for the weak lensing studies in the COSMOS field. The authors created an IDL wrapping code for Tiny Tim, which mimics the action of MultiDrizzle during the reduction of raw telescope data. We modified the code to be executed in Python and to produce a single PSF model instead of a field of position-dependent PSFs

[1] Although the study by Sánchez et al. (2004) applies same methods for PSF creation, we are going to refer here to Jahnke et al. (2004) due to the similarity of the redshift bins to our study

[1] `http://www.stsci.edu/hst/observatory/focus/TinyTim`

for a given focus distance. Given that our samples of quiescent galaxies and AGN hosts total to about $\sim$ 500 objects, such approach saves us a great deal of time. Using a set of carefully selected parameters, we created a sample of position and focus-dependent PSF models for F814W filter (tracing the central wavelength of this filter, 814 nm). The IDL pipeline[1] uses original TinyTim software by STScI to produce distorted models of the PSF, and then removes geometric distortion as it is done by MultiDrizzle.

8.1 Distortion removal effects and MultiDrizzle parameters for optimal image reduction

In Rhodes et al. (2007) the authors extensively tested the quality of the created models to make sure that their pipeline produces models that represent the properties of real data. They estimated optimal parameters for the data reduction (which we also utilized to reduce COSMOS data) in order to ensure that drizzling the data would improve the shape and stability of the brightness profiles of objects and minimize the PSF aliasing due to distortion removal. Primarily, in order to test how the distortion removal affects the data, they created a grid of identical stars with the same diffraction pattern and charge diffusion for every model, and then distorted this image. In and ideal case, after reducing this image with MultiDrizzle, the stars should appear identical. Therefore, changing the parameters for a MultiDrizzle data reduction, the authors could estimate the optimal values, when the stars in the grid show the least residuals to compare with their initial shapes. To measure shapes the Gaussian-weighted second-order moments were computed:

$$I_{uv} = \frac{\sum_i wIu_iv_i}{\sum_i wI},$$

where we sum over all i pixels. I is the pixel intensity, u,v are the coordinates and w is the width of the Gaussian weight function, determined by $w = max(2\sqrt{ab}, 6)$, where a,b are the semimajor and semiminor axes estimates, computed by SExtractor, and the minimum weight-function width of 6 was empirically determined for a given set of Sextractor parameters. These second-order moments were used to compute ellipticity e_i (e_1 determines the size of star's ellipticity and e_2 - it's direction) and size d for every object:

$$e_1 = \frac{I_{xx} - I_{yy}}{I_{xx} + I_{yy}} \tag{7}$$

$$e_2 = \frac{2I_{xy}}{I_{xx} + I_{yy}} \tag{8}$$

$$d = \sqrt{\frac{I_{xx} + I_{yy}}{2}} \tag{9}$$

[1] http://www.astro.caltech.edu/ rjm/acs/PSF/

Figure 39, taken from Rhodes et al. (2007), shows the measurements of ellipticity for 3 different types of images with geometrical distortion removed by MultiDrizzle employing different parameters for data reduction.

We already described the drizzling algorithm and the effects which appear after raw telescope data is drizzled (see Section 7). The main parameters that control the output data quality are *pixfrac*, *scale* and *kernel*. Thus, by altering these parameters one can significantly change the precision of measurements. For instance, if the size of the "input drop" during drizzling is kept at its original value ($pixfrac = 1$), it results in a degradation of the image resolution and significant correlations in noise. Additionally, pixelization effects i.e. how the value of the input pixel is redistributed between the output undistorted pixels, can cause the chaotic behavior in the final distribution of ellipticities as we see on the top two plots in Figure 39. Both pixelization and correlated noise are controlled by the *pixfrac* parameter. At the same time, reducing the default ACS WFC pixel-scale from 0.05″/pixel to 0.03″/pixel will introduce a significant improvement of PSF aliasing even for a single drizzled image shown on the bottom plot. The improvement is even more pronounced for an image that was combined from 4 dithered exposures. In general, the scatter in ellipticity drops proportionally to the square root of the number of dithers.

Figure 40 Rhodes et al. (2007) shows the RMS of ellipticity measured as a function of the *scale* parameter for two different kernels, used for drizzling-combining dithered images. By replacing the default square kernel with a gaussian kernel, user obtains a better result in RMS ellipticity, while stabilizing its general behaviour with decreasing values of the detector's pixel-scale. However, there is a drawback when using a gaussian kernel for the data reduction since it slightly increases the amount of correlated noise, which causes difficulties for the error estimation with GALFIT. This Figure also illustrates, that the scatter is even smaller for the values of around 0.02″/pixel, but it requires a significant increase of the storage space, therefore we adopted the value of 0.03″/pixel for reducing the COSMOS data.

8.2 Focus determination

We have already explained that temporal variability of the PSF during HST observations is caused by the changes of the focus distance of the telescope. These changes happen when the optical system of the telescope contracts due to temperature variations when telescope goes in and out of the Earth's shadow. The variability cycles are modulated by the target occultaion and solar angle during observations and can not be predicted beforehand. For the COSMOS data the focus values have been previously determined, so we used those for our PSF modeling.

One of the methods to measure the focus distance of the telescope at the moment of observations was described in the paper by Rhodes et al. (2007). The authors created grids of artificial stars for a continuous sequence of focus distance values in the range from -10 to +5 μm. The artificial stars they used were highly

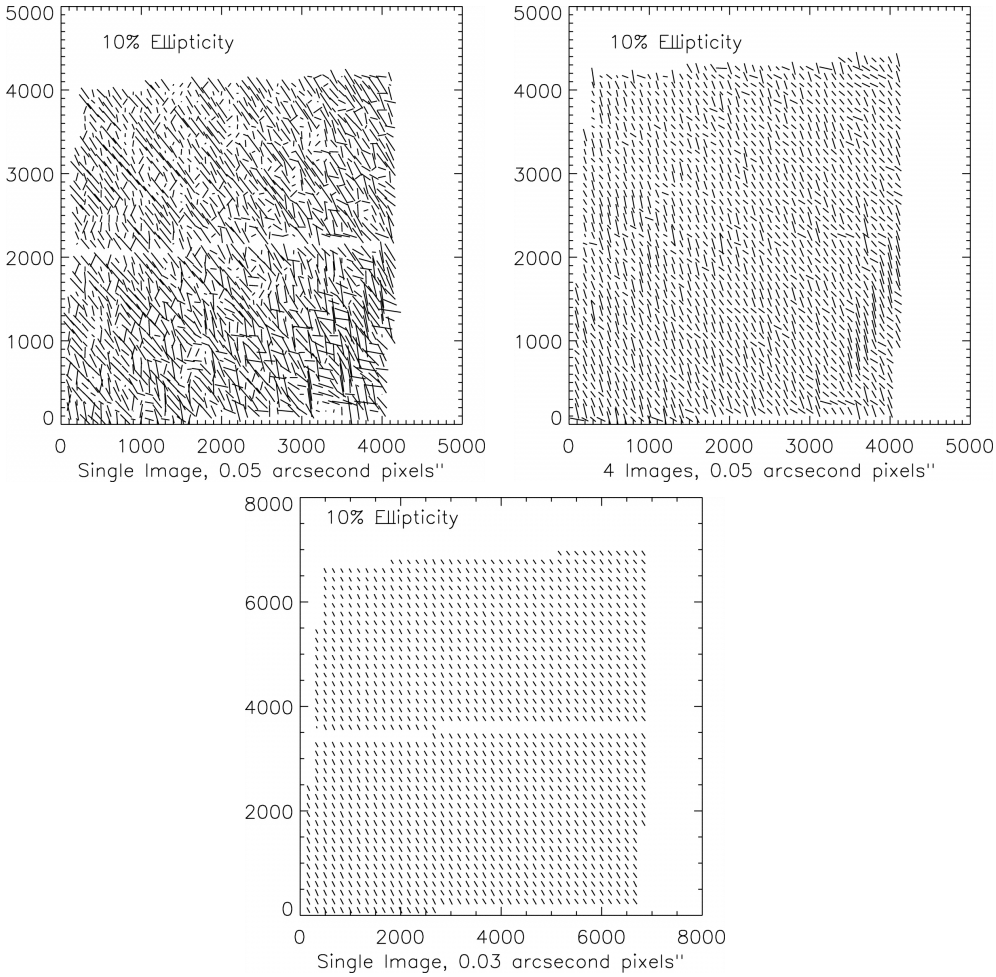

Figure 39: Aliasing of the modeled stars, which occurs when the geometricly distorted image is reduced with MultiDrizzle. On the top left: single drizzled image, reduced with the original pixel-scale of ACS WFC of 0.05″/pixel. The sizes of ticks represent the 10% ellipticity computed with formula (7) and orientation determined with the equation (9). On the top right: 4 artificially dithered images drizzled on the output image with the native pixel-scale of 0.05″/pixel and original value of *pixfrac=1*. The bottom image shows scatter in measured ellipticities of a single drizzled image, reduced with modified MultiDrizzle parameters: pixel-scale of 0.03″/pixel, *pixfrac=0.8* and a Gaussian kernel instead of the default Square kernel. These plots are taken from Rhodes et al. (2007).

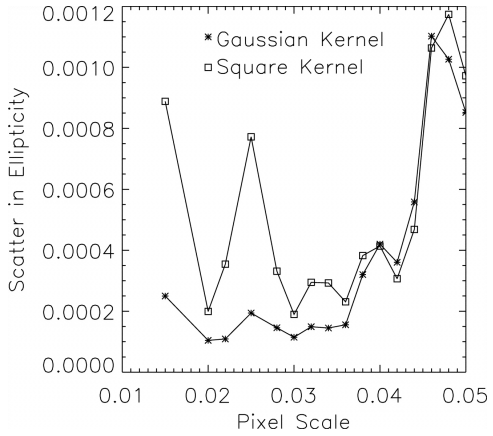

Figure 40: The RMS ellipticity measurements for the modeled stars, reduced with the MultiDrizzle, as a function of the pixel scale for two different re-sampling kernels, gaussian and square. Figure is taken from Rhodes et al. (2007).

oversampled, in order to avoid stochastic aliasing of the PSF during the geometric distortion removal, which occurs due to re-pixelization of the data. High oversampling in this case is equivalent to combining the output image using a large number of dithered exposures, thus, the stars in the grid appear as if they underwent perfect geometric distortion removal. The diffraction and charge diffusion of each star, then, depends only on the pixel position on the detector. The spacing between the artificial stars in the grid was made to be small enough, so that when matching a star to a model, one would not have to do an interpolation between adjacent models but one could simply take the nearest model without significantly increasing the error. The authors claim that the 30×30 star grid matches this requirement. Each grid was created for a focus value in the range from -10 to +5 μm with a step of 1 μm, which was sufficient to keep the ellipticity measurement error of the same star from two adjacent grids below the RMS noise in a real COSMOS image.

In each ACS COSMOS tile there are at least 10 non-saturated, well de-blended stars, that can be used for determining the focus distance of the telescope and controlling the PSF quality. These stars are compared to matching Tiny Tim models from each grid and then the focus distance value is determined from the χ^2 minimization for each pair of stars in the same tile:

$$\chi^2 = \sum_{\bigstar} \left((e_1^{\bigstar} - e_1^{TT})^2 + (e_2^{\bigstar} - e_2^{TT})^2 \right), \tag{10}$$

where $e_i^{\bigstar}$ is the ellipticity measurement for the real data and and e_i^{TT} is the ellipticities of the Tiny Tim models. Figure 41, taken from Rhodes et al. (2007), shows the determined focus values of each ACS tile observed for the COSMOS Survey. We used these estimates to produce PSF models for our samples of QSOs, galaxies and stars.

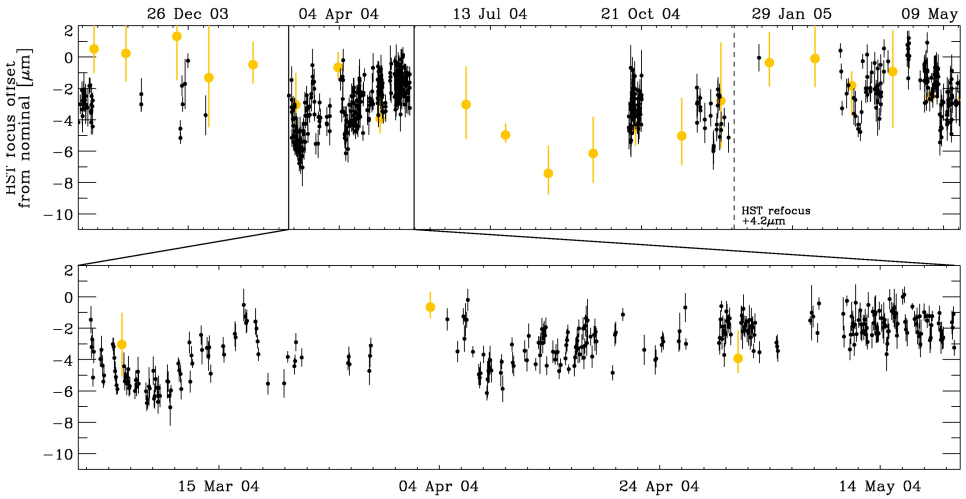

Figure 41: The focus measurements for COSMOS Survey from Rhodes et al. 2007 are shown with small black circles. The error-bars represent the variation of focus estimates of all stars used for these focus distance measurements. The yellow data points are the focus values from HRC focus monitoring program (Lallo et al., 2006). The error-bars represent the intraorbit focus variation.

8.3 The difference between modeled and real data

We created all the PSF models using the set of parameters described above. In order to check whether such models are realistic, we performed two kinds of tests to adjust the parameters of the IDL pipeline and confirm the immediate concordance between Tiny Tim model and real data. First, we compared images of individual stars in the COSMOS survey to corresponding Tiny Tim models, rescaling the flux of the PSF to match its star. The original settings of the pipeline include a double convolution of the PSF model with a square kernel of the width of the original pixels (taking into account oversampling). This convolution broadens the Tiny Tim model to match observations, without it the models would be too narrow. In attempt to replace this double convolution, we decided to apply a single convolution with a Gaussian kernel of the width of an original pixel. In this case the agreement between the FWHM of stars and PSF models is good and in general their profiles match well. Furthermore, we save time which was spent before for the additional convolution.

To compare the Tiny Tim PSF models to empirical, we created several position-dependent models from the stars in the GEMS survey, which we previously used for the low-level variable AGNs survey described in the Chapter I of this thesis. We combined the 35 nearest stars within 40 pixels for each pixel position, and compared the resulting PSFs to the Tiny Tim models. We adjusted the Tiny Tim pipeline to the GEMS observations, enabling the F606W filter (which was employed for the GEMS Survey), while the focus distance for the GEMS observations was determined from the field stars as described above to be -4 μm. Figure 42 shows an example of a pair of empirical and Tiny Tim PSF models and their surface

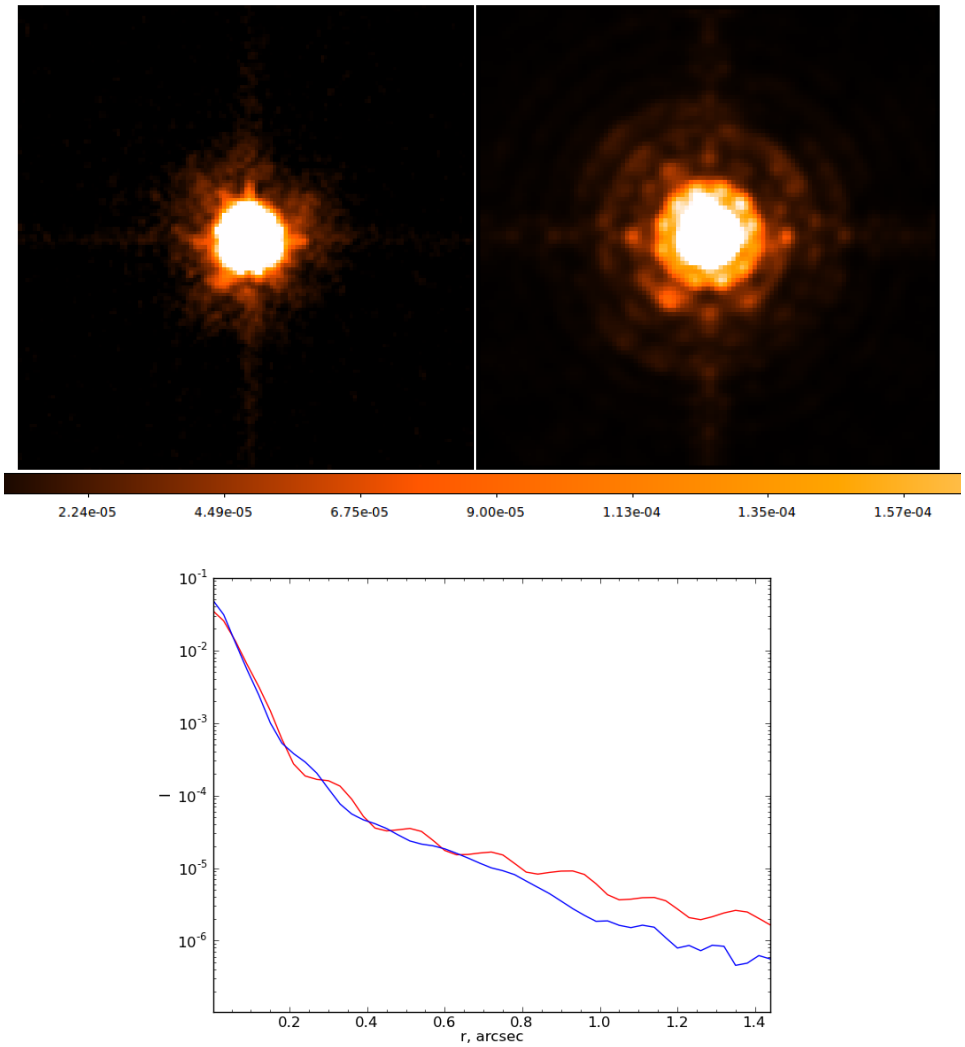

Figure 42: Top panel: empirical (on the left) and Tiny Tim (on the right) PSF models created for the pixel position of COMBO ID 784 from the GEMS Survey (Combo ID 784). The empirical PSF was averaged from the 35 nearest stars, providing very high signal-to-noise ratio, and the Tiny Tim model was created with the IDL wrapper by Rhodes et al. (2007), for the F606W filter and focus value of $-4\,\mu$m. On the bottom panel brightness profiles of both of these models are overplotted: the Tiny Tim profile is marked in red, and the empirical PSF made from stars is marked with a blue line.

brightness profiles. The Tiny Tim noise-free model on the top right shows a very pronounced diffraction pattern, which in the image of averaged stars is smeared out by noise in the individual images from which it was combined. Additionally, some of the smearing can be introduced by the charge diffusion (or CTE) in the CCD (discussed in Section 7.6).

Such mismatch can have important consequences when fitting brightness profiles of faint galaxies in presence of luminous AGNs. The residual flux in the central few pixels may be ascribed to an additional component, increasing the error of the fit or even causing spurious detection of a very compact host. However, this becomes an issue only for the brightest AGNs with $I \leq 19$. Meanwhile, Figure 42 also shows that there is a mismatch in the wings of the PSF profiles, at $r > 25 pix = 0.75\ arcsec$, however in terms of flux, this mismatch is so small, that it does not play a significant impact on the fitting process. To determine quantitatively what consequence such mismatches will cause on errors propagation, we created a set of simulations for QSOs and quiescent galaxies, which we will discuss in Section 10.

9 Methods

9.1 Simple PSF Subtraction

In the description of the sample we had shown that ~80% of our sample is dominated by Type I AGNs according to their hardness ratios: AGNs which show soft X-ray hardness ratios of $HR < -0.2$ are classified as Type I and those with $HR > -0.2$ as Type II, where H are the hard X-ray band counts and S the soft X-ray band counts. There are evidences that Type II AGNs at redshift $z > 2$ have enhanced star-formation activity with respect to Type I AGNs, therefore, we expect their rest-frame UV host-to-total ratios to approach the unity (Hatziminaoglou et al., 2010). As for the luminous Type 1 AGNs in our sample the contrast between nucleus and underlying host galaxy could make a detection very difficult or even impossible.

As a main tool to perform surface-brightness profiles decomposition we selected GALFIT. We already described in Section 3.1 of this thesis the main principles of how the GALFIT algorithm works. But before fitting the entire sample of ~350 AGNs, we performed a simple test to find and exclude objects, which did not show any signatures of host galaxies, and probably would not be resolved when fitting with GALFIT. We carried out simple PSF subtraction for the entire sample of AGNs of both types. This method had already proven itself to be useful to extract host galaxies magnitude in presence of bright AGN before (Sánchez et al., 2004; Jahnke et al., 2004), therefore we expected to obtain the first estimates of the number of AGNs with resolved hosts. The main idea of this method is that if one can create a PSF model that matches observations, the PSF can be then rescaled to the central flux of the object and subtracted from it. Then all the residual flux in the image, provided that the background is properly subtracted, can be assigned to a host galaxy. To determine the significance threshold, the residual flux is compared to residual flux of the PSF subtracted stars. If the residual flux of the hypothetical host after the PSF subtraction is larger than the bulk of the distribution of residual fluxes of stars, the host can be considered to be resolved.

In case of the COSMOS data, the simple PSF subtraction is complicated by the existence of a mismatch between the data and PSFs modeled with Tiny Tim. As we had shown in Figure 42, the PSFs, created with Tiny Tim IDL pipeline, have strong diffraction patterns, which in real data is smeared out by drizzling and noise. Because of this mismatch, after the PSF is subtracted from each star, the distribution of $\frac{F_{resid}}{F_{Total}}$ for the entire sample is significantly broadened. Figure 43 shows an example of residual image of a PSF-subtracted star and the brightness profiles of this star and its matching PSF in the right panel.

To scale PSF to the flux of star, we matched their central 2-pixels aperture fluxes. One can see that such scaling provides a good matching of profiles in the central $r < 6$ pixels area of the star. But at larger radii strong diffraction pattern in the Tiny Tim model causes large residuals between the data and the model, which results in a broad distribution of $\frac{F_{resid}}{F_{total}}$ for the entire sample of stars. Such

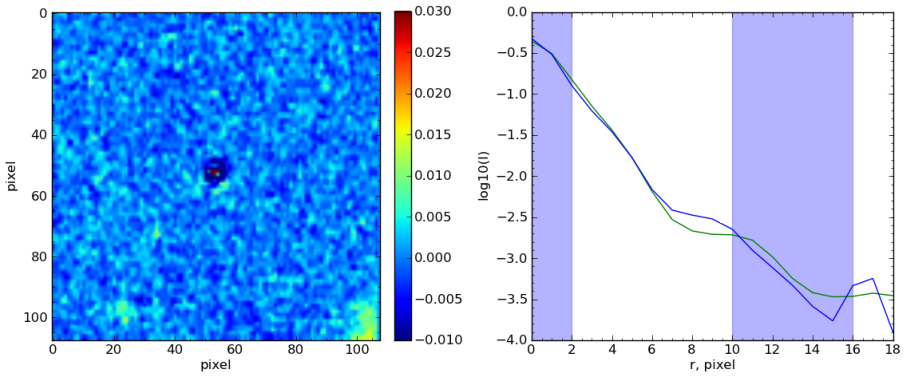

Figure 43: Left panel: residual image of a star after scaled PSF subtraction. Right panel: radial brightness profiles of this star (blue) and its scaled Tiny Tim PSF model (green).

distribution can not be used to isolate AGNs with resolved hosts from the pure point-sources. To minimize the residuals we compared flux only in the certain ranges, where the profiles are the most sensitive to the input from the extended galaxy component. These regions have to be large enough to not suffer from the variations of flux due to the diffraction spikes, and possibly avoid the regions of the largest mismatch between Tiny Tim models and COSMOS data. After a number of tests we adopted the regions, shown in color in the right-side plot in Figure 43. The area between $0 < r < 2$ pixels was used to match fluxes of objects and their PSFs, and has a zero residual flux by definition, and the $\frac{F_{resid}}{F_{total}}$ was measured in the $10 < r < 16$ pixels annulus. Figure 44 shows two pairs of brightness profiles: one for a star and its scaled PSF (blue and green curves), and another for a QSO with a detectable host galaxy and its scaled PSF (black and red curves). One can see that in the inner region ($r < 10$ pixels) of their radial brightness profiles, the star and the QSO are very similar. But at $r > 10$ pixels, the input from the host galaxy component stands out in the radial brightness profile. Therefore, if we were to use the ratios of residual-to-total fluxes in this annulus, it would provide a sensitive tool to isolate pure point-sources from those AGNs which have detectable hosts.

Figure 45 shows the distribution of $\frac{F_{resid}}{F_{total}}$ for samples of stars and AGNs. We adopted a threshold of $\frac{F_{resid}}{F_{Total}} \geq 0.01$ to select a sub-sample of AGNs to be modeled with GALFIT as the next step of the analysis. We also performed a visual check of the $\frac{F_{resid}}{F_{total}} \geq 0.01$ sub-sample, to make sure that all the AGNs in it indeed appear as point-sources. Indeed, a number of objects which showed residual fluxes after the PSF subtraction of less than 1% appeared to be possible to resolve with GALFIT.

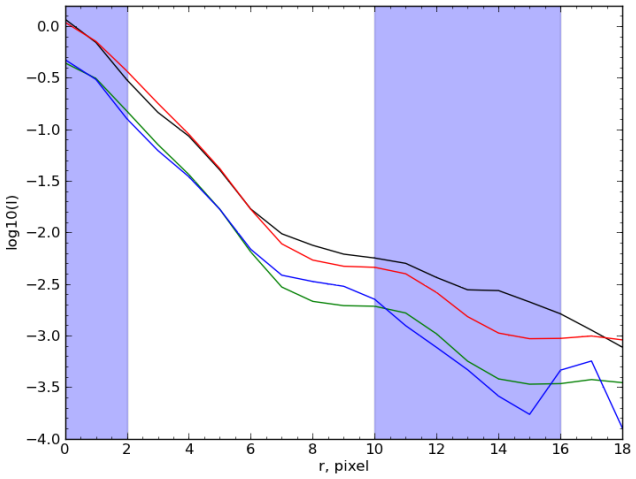

Figure 44: Pairs of radial brightness profiles for the star and its scaled PSF (blue and green curves) and for the QSO with a detectable host galaxy and its scaled PSF (black and red curves). The shaded area denotes regions where measurements of flux ratios were done.

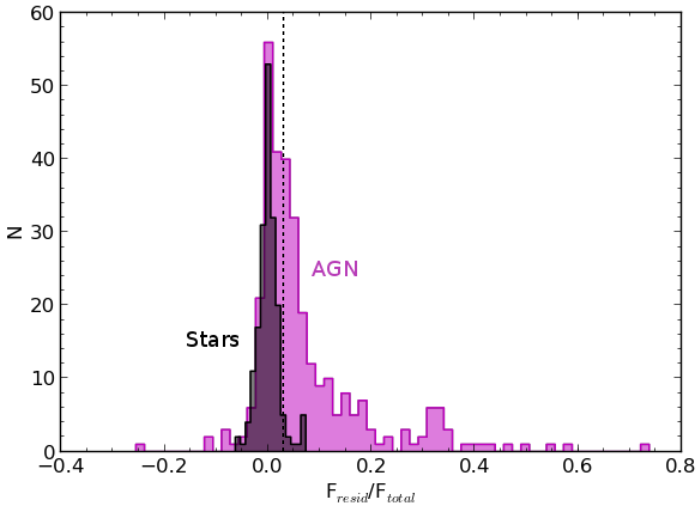

Figure 45: Distributions of $\frac{F_{resid}}{F_{total}}$ for stars (black histogram) and AGNs (in magenta). We adopted a threshold of $\frac{F_{resid}}{F_{total}} \geq 0.01$ to select QSO to be modeled with GALFIT to create a catalog of morphological properties of high-redshift AGN hosts.

9.2 GALFIT

In section 3.1 we described the main principles of the GALFIT algorithms, when we used it to fit GEMS data for the Low-level AGN variability survey. For COSMOS data we used a similar approach: we performed a 2-component *Sérsic+PSF* fitting, using SExtractor to compute the input guess parameters. The error maps were produced in the same way as in the GEMS survey, using the weight image output by the MultiDrizzle pipeline:

$$\sigma = \sqrt{var} = \sqrt{\frac{\frac{(sky + 75.)0.36^2 t}{W} + It}{t^2}}, \tag{11}$$

where *sky* is the sky level subtracted by the pipeline during data-reduction, *75* is an additional variance offset due to readout noise, *0.36* is the scaling factor caused by the change in the pixel-scale from 0.05″/pixel to 0.03″/pixel, t is the exposure time of the observations, which for all COSMOS tiles constitutes 2028 seconds, W is the weight image and I is the reduced science-image in the units of counts per seconds. This formula does not take into account two factors, which causes us to underestimate the noise by a small amount. Primarily, due to original geometric distortion of the data the pixel-scale of an image is not constant across the image (we discussed this in the data reduction section), which results in the change of a degree of the correlated noise in the image. Secondly, correlated noise in our data is slightly enhanced due to the Gaussian kernel used for the image re-sampling during drizzling. Because of these two reasons, the minimized χ^2 estimated by GALFIT will not converge to unity. To overcome this issue, we relied on the relative change in the minimized χ^2 for a different sets of input parameters in order to estimate the best-fit model.

Since the χ^2_ν depends on the number of degrees of freedom of a fit, we always used 2-components, attempting not to change the number of free parameters for a fit. We used a PSF+Sérsic profiles, allowing Sérsic index to have fixed values of $n = 1$ for an exponential disc and $n = 4$ for an early-type/spheroidal host galaxy. We kept this parameter fixed to increase the robustness of host galaxies' models in the presence of a bright point-source. The rest of the values were determined using SExtractor: the magnitude guess, determined by SExtractor, was split into estimates for components to have difference of $\Delta m = 2.5$ mag. The effective radius estimated by SExtractor is dominated by the presence of a bright point-source and does not serve as a good guess for GALFIT in general. It often requires manual interaction to improve the fit, same as the estimate of the minor-to-major axes ratio and the position angle. The output value of the minor-to-major axes ratio often serves as an indicator of the quality of the fit. Values outside of the $0.2 < b/a < 0.9$ range are often a sign that a fit has failed. We initially set the center of the components to the same pixel position, determined through the Gaussian fit of the AGN.

We implemented a Python code to run the first iteration of fitting with GALFIT:

this pipeline takes a list of objects that suitable for fitting (those which passed the $\frac{F_{resid}}{F_{Total}} \geq 0.01$ requirement) as input. Then, object by object, using SExtractor, the code creates a GALFIT input parameter file for two fixed values of Sérsic index and performs modeling. The resulting best-fit parameters for each component are written into two tables for each value of Sérsic index. After all models are created, the code initiates a function, which schematically visualizes the fits for each value of Sérsic index for each source, taking into account the estimated output magnitude, effective radius, minor-to-major axes ratio and the reduced χ^2_ν value of the fit. At this point the code requires manual interaction with user, who considers the shape and quality of the GALFIT residual image, using schematically shown model and the change in the χ^2 between the disc-dominated fit and the early-type fit. The user has to choose which profile suits better to the data (disc or early-type) or to flag the object if any of its output parameters raises doubts. Additionally at this step, the user has to assign a flag, which marks the environment of the considered object. The options are the following:

- '0' indicates the presence of a background object, without any signs of interaction with the investigated object. Such background objects usually were masked out;
- '1' indicates an event, which can be interpreted as a major merger, when the investigated AGN (or galaxy from the galaxies sample) is showing signs of strong interaction with a nearby object that has comparable brightness
- '1/2'(12) indicates an AGN/galaxy that shows signs of interactions, but the nearby companions are faint/invisible
- '2' is assigned to lopsided objects, where the detected host galaxy center is significantly offset from the estimated center of the point source
- '2/3' (23) is for questionable objects, which do not show strong lopsidedness, but neither have an entirely undisturbed profile
- '3' signifies undisturbed objects, which do not show any signs of interaction (this category is normally combined with '0')
- '4' is the label for unresolved objects, which does not show any indication of a detectable host

Our classification has a bias imposed by the human factor, however, the specifications of each category should provide a reliable estimate of the AGNs' and galaxies' environments.

As an output of this function, the user receives a table, which includes the choice of the best fitting model (an exponential disc or early-type galaxy profile), environment flags and with problematic cases marked out to be considered manually. Experience shows that many of the objects require this manual consideration

due to a very low contrast between the central point-source component and the host. For the most problematic cases, where the input parameters can not provide convergence of a model, the fit crashes. Sometimes, when an AGN (or a quiescent galaxy) is in a state of interaction, one needs to include additional components to fit the entire group in order to get a good estimate for the object of interest. Of course, there are objects which appear as pure point sources (even taking into account our primary selection cut with the Simple PSF subtraction method), and the fit does not converge. Such cases were marked as AGNs with unresolved hosts, and written in a separate table.

For the models, for which GALFIT failed to converge (or converged to unrealistic values) using the initial SExtractor guess parameters, we carried out the modeling manually. There are various options, which one can use to achieve a good result during fitting. Primarily, one can add more components in case of modeling a group of interacting galaxies. Furthermore, it may be useful to perform an additional GALFIT run, allowing the Sérsic index as a free parameter while fixing the magnitude or minor-to-major axes ratio. This way we can get an estimate of morphological type, when the residuals of GALFIT modeling are hard to judge (e.g. see Figure 46). In some few cases if GALFIT converged to an obviously false models, we kept minor-to-merger axes or position angle parameters fixed.

As the final step, we used GALFIT to attempt to fit those AGNs, which were suggested as unresolved by simple PSF subtraction to improve completeness of the AGN host sample. Most of these objects, though, did not reveal any host galaxy signatures.

Similarly, fitting quiescent galaxies is not an easy task, however the types of stumbling block are different from the ones arising when fitting QSOs. Since we selected objects in a way to match their rest-frame UV brightness and redshifts to the brightness and redshifts of the AGN host galaxies, many of these galaxies are undergoing strong interactions. A large number of sources in the sample are hard to de-blend, and they require a careful placement of additional components or, sometimes, masking in order to separate fluxes of distinct galaxies. A few of the quiescent galaxies required an additional point-source (or a Sérsic) component for a more adequate fit, which might have been due to the presence of an unresolved galactic bulge.

At this point, we would like to discuss few objects in the AGN sample, for which the host galaxies are estimated to be extremely compact and have very high surface brightness in comparison to the rest of the detected galaxies in the sample. These galaxies have absolute magnitudes $M_{UV} < -22.5$ and $R_e < 4$ pixel~ 0.8 kpc, of both early-type and late-type morphologies. They may either be very bright compact star-forming bulges or, more likely, fake detections due to the PSF mismatch. We checked these objects manually. Most of them are bright ($m_{814} < 20.5$), spectroscopically confirmed broad-line Type 1 AGNs except for 2 objects. Taking into account that the width of the HST/ACS PSF is $FWHM \sim 4$ pixels, it is highly unlikely that these fits are just. The visual check revealed very strong modeling residuals which resemble the diffraction spikes in the Tiny Tim PSF model, for the

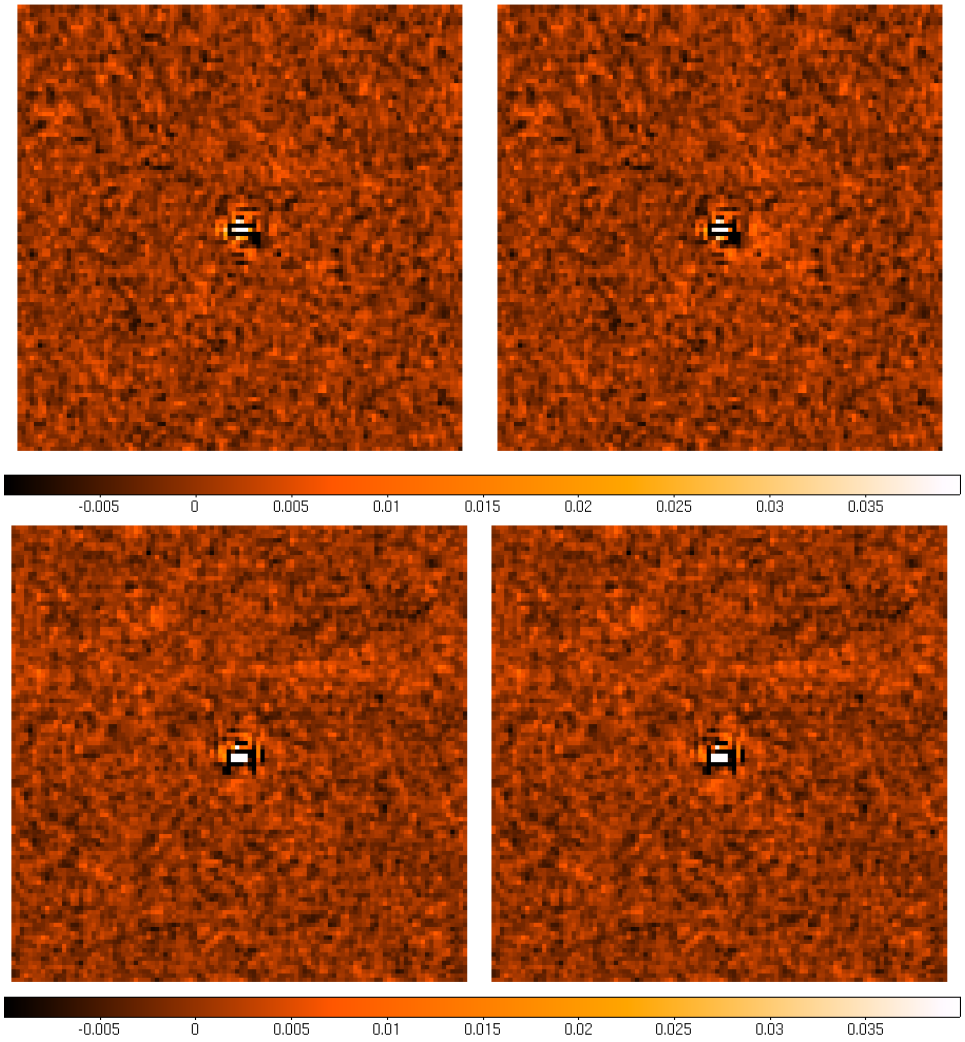

Figure 46: The residual images of GALFIT modeling for two objects with $n = 1$ model on the left and $n = 4$ model on the right. These two galaxies give an example, how difficult is can be to judge the model by eye. Top panels: both models converge, and return good estimates of fitting parameters with a nearly same value of χ^2. The model with a Sérsic index as a free parameter, however, is suggesting a disc-dominated profile. The residual pattern one can see in this image results from drizzling, using the gaussian kernel, and indicates a good match between the data and its Tiny Tim PSF model. Bottom panel: another QSO fitted with an exponential disc $n = 1$ and early-type $n = 4$ profiles. Both models converge with the same value of χ^2 and look very similar without giving any suggestions on a better suited model. The effective radii estimates, however, are significantly different. When n is left as a free parameter model converges to a disc-dominated morphology. The few bright pixels in the central area of a residual indicate a mismatch with the used PSF model.

reference see Figure 42. We give an example of one of these objects in Figure 47. This object, however, shows a distinct tidal tail. It extends by approximately 20 pixels towards the south-west direction of the image, which, on the redshift of this QSO of $z_{spec} = 1.96$, corresponds to ~ 5 kpc. Because of the tidal feature visible, this AGN is included into the sample, but it is excluded from the analysis, as well as the other ultra-compact high surface-brightness objects.

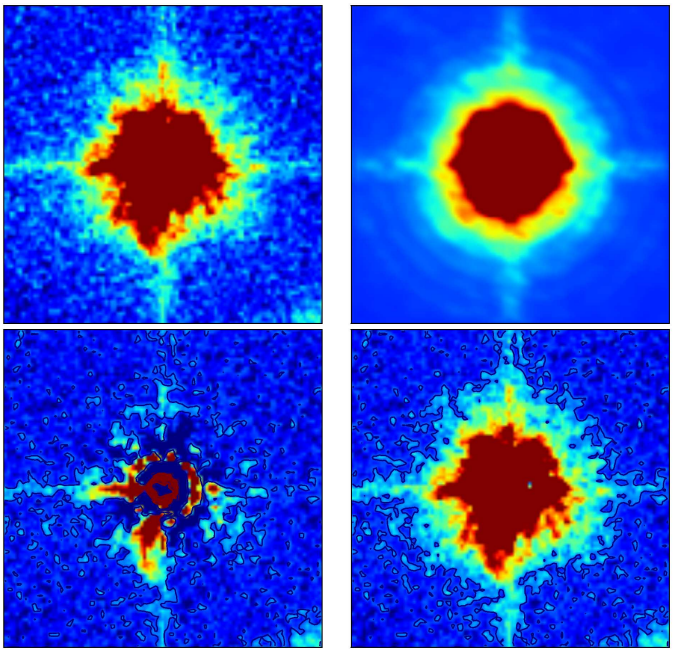

Figure 47: GALFIT fitting results of XMM ID 10889. The top left plot is the original image before being fitted by GALFIT; top right plot is the best-fitting PSF+Sérsic model; in the bottom left panel is the residual from original image minus its model; the bottom right plot is the host galaxy remaining after the *PSF* has been subtracted from the original image. GALFIT residual image indicates a strong mismatch with the PSF model, used to fit this image. The residual image reveals a tidal tale, which extends from the center of the image in the south-west direction. This object lies at a redshift of $z = 1.96$, which corresponds to the scale of 8.27 kpc/″. Therefore, at the pixel-scale of 0.03 ″/pixel the detected tidal tail extends by ~ 5 kpc.

10 Simulations

To verify how well GALFIT can estimate the parameters of AGN host galaxies, we created a sample of mock quasars. As the input parameters for mock AGN observations we used values recovered by GALFIT for our AGN sample summarized in Appendix A. For each set of parameters m_{host}, m_{PSF}, r_e, and b/a we made two realizations with two different values of Sérsic index n in the ranges $0.5 \leq n < 2.5$ for a disc-dominated morphology and $2.5 \leq n \leq 5$ for an early-type morphology.

The most crucial element in these simulations is the input PSF that has to be provided to GALFIT. The choice of the input PSF determines how well simulations will reproduce the real data, since every mock galaxy created with GALFIT is convolved with this PSF. We already discussed before how we created PSF models with the Tiny Tim IDL Pipeline and showed that they do not entirely match observations (see Section 8). Therefore, we could not use these models since we wanted to examine how does the mismatch between the model and the observed data affects the parameters estimation by GALFIT.

To make the simulations as realistic as possible, we used the empirical PSF created from the field stars. We adopted the approach we described previously in Section 3.2, averaging the entire sample of stars, independently of their pixel-positions and focus values. The biggest drawback of this step is that we neglected both spatial and temporal variability of the ACS PSF. However, we also combined individual Tiny Tim PSFs of each star, that contributed to the empirical PSF, into a composite Tiny Tim PSF, to smooth out the effects of spatial and temporal variability.

Adding realistic noise, that would reproduce the properties of noise in the ACS drizzled data to the simulated images is a non-trivial task. We already explained before in Section 7.2 that drizzling narrows down the original distribution of noise in the images and introduces correlated noise between adjacent pixels. This happens because (1) output pixels are combined from several input pixels and (2) each input pixel may contribute to more than one output pixel. In this case the output values of pixels are weighted proportionally to the overlapping area between input pixels. To get a realistic noise distribution, we added noise to the images in three steps, as described in Sánchez et al. (2004). First, we assumed a Poisson noise distribution for every noise-free image, adding mean sky background level, computed from sky levels estimated by MultiDrizzle during data reduction, and readout noise just like for real data. Then, to narrow down the resulting noise distribution and mimic correlated noise, we smoothed the images with a median filter, using a 3×3 pixel box. Finally, we combined this smoothed image with the original pure Poisson noise one, assuming different relative weights in order to reproduce the properties of noise in the real data. The combination of 80% of the original map and 20% of smoothed seems to recover the *RMS* of the real data well.

The resulting mock galaxies sample consists of 340 simulated AGNs, which we pass through the same fitting procedure as the real AGNs. We fitted all the objects two times with two fixed values of Sérsic index (n=1 and n=4) for a "*Sérsic+PSF*"

model to create a final catalog of parameters. We compared these estimated parameters with the input ones that we used to create models to see how well they can be recovered and compute systematic fitting errors.

Figure 48 shows the relation between the input and output simulated host galaxies' magnitudes. Red and blue circles show galaxies which were fit with the Sérsic profiles, which match the ranges for input Sérsic index for early- and late-type morphologies correspondingly. We also plotted results for mock AGN hosts, fitted with obviously wrong profiles: blue dots mark early-type galaxies fitted with exponential discs and red dots are late-type galaxies fitted with de Vacouleurs n=4 profile. According to this plot we can conclude that the wrong choice of the Sérsic index does not significantly increase error of the host galaxies magnitude estimates at least for the fainter galaxies.

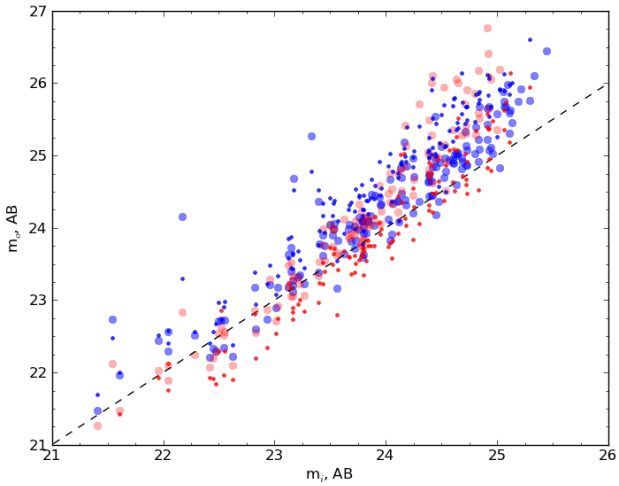

Figure 48: Comparison of the input host galaxies magnitudes to the magnitudes recovered by GALFIT. Red circles mark early-type morphologies, blue circles disc galaxies; blue dots stand for early-type galaxies fitted with the $n = 1$ Sérsic profile, and red dots for late-type galaxies fitted with de Vacouleurs profile. Dashed line marks 1:1 relation.

The distribution of recovered host galaxies magnitudes in Figure 48 shows that for host galaxies of up to an intrinsic magnitude of $m_{host} \sim 24.5$, even in the presence of a bright PSF component and for host-to-total ratios spanning from as few as 2% and up to 80%, GALFIT returns very good estimates of magnitude with an error less than < 0.3 mag. Only beyond $m_{host} \sim 25$ recovered host galaxies magnitudes become systematically underestimated. These magnitudes correspond to very weak signal-to-noise, that can be computed as

$$\frac{S}{N} = \frac{\langle \rho \rangle}{\sqrt{\langle \rho \rangle + \langle \rho_{sky} \rangle + \sigma_{sky}}}, \tag{12}$$

where ρ is host galaxy's counts within the area encircled by effective radius, ρ_{sky}

is mean background counts per pixel, and σ_{sky} is the RMS of the background estimate. In this case the limit $m_{host} \sim 24.5$ corresponds to the signal-to-noise $\frac{S}{N} \sim 2$, at which estimation of magnitudes (and other parameters of a galaxies) becomes troublesome. Taking into account these measurements we suggest that all the host galaxies estimates that are fainter than $m_{814} > 25$ should be considered cautiously.

In a similar manner, to check the quality of the effective radius estimation, we plot the relation between the input (value that was used to create a model) and the output estimated by GALFIT values of the effective radii of mock AGN hosts. Figure 49 shows this relation for early-type galaxies, shown in red, and disc-dominated galaxies, shown in blue. Dashed black line marks 1:1 relation. Dotted lines mark the linear fit to the data points of each morphological type. Since our simulations are realistic and the fitting process was same as for the real data, we can use these linear fits to correct the effective radii estimates for our real AGN hosts. The effective radii corrections, however have a large error due to the large scatter of recovered values, especially for early-type galaxies.

In Figure 50 we plot the degree of size over/underestimation $\frac{R_{e,o}}{R_{e,i}}$ as a function of the host galaxy output effective radius and magnitude. The top panel shows that all galaxies, that are intrinsically larger than 8 pixels (0.24″) independently of morphological type, have their size underestimated. As for the galaxies that are smaller than this limit, their size can be both under- or overestimated. If underestimation appears natural in case of noisy data, the reason why small intrinsic radii may get overestimated is not clear. Overestimation may be due to several reasons, like PSF mismatch, that becomes stronger for brighter objects, low signal-to-noise of objects or low host-to total ratio. The bottom panel of Figure 50 shows ratio of output to input effective radii as a function of host galaxy magnitude. One can see there is a slight correlation visible, however there are no implications that the overestimation of galaxy's size depends on the host galaxy's magnitude. If we make the same plot for the host-to-total ratios, shown in Figure 51, it shows that for the smaller host-to-total ratios (lower contrast), the probability of overestimation grows. We can conclude that using our simulations, we can accurately correct the effective radii measurements done with GALFIT for all AGN hosts larger than $r_e > 8$ *pix* and less accurately, for AGN hosts that have host-to-total ratios of less that 20% and $m_{tot} > 21.2$.

To estimate how strong the impact of the PSF mismatch is for the recovery of AGN host galaxies' parameters, we also created a simulated sample of non-AGN galaxies, using the same set of parameters, but excluding a PSF component in the "Sérsic+PSF" model. We performed the automated run for the fitting of these quiescent galaxies. The results are shown in Figures 52, 53 and 54.

Figure 52 shows the relation between input and estimated magnitudes of simulated galaxies with a single Sérsic component. Along the entire magnitude range the recovered magnitudes are in very good agreement with the input values, showing a very small scatter of 0.15 *mag*. Thus, the absence of a PSF component reduces the scatter by a factor of 2-2.5 depending on the $\frac{S}{N}$.

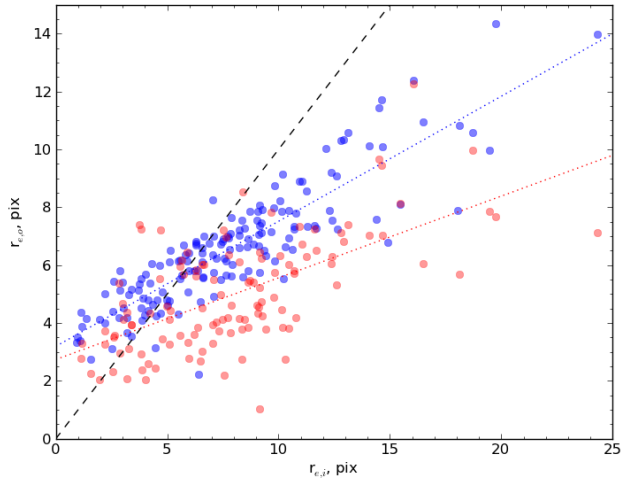

Figure 49: The relation between input and estimated effective radii for early-type morphologies (red circles) and late-type morphologies (blue circles) in units of pixels. The black dashed line marks $r_{e,o} = r_{e,i}$, the dotted lines of the corresponding colors are the linear fits to the data-points.

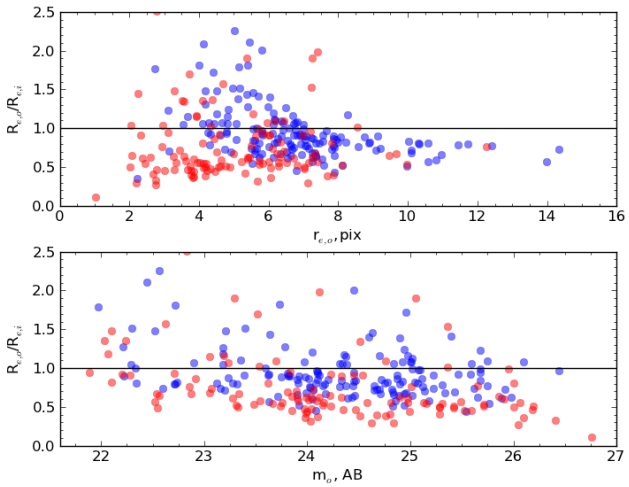

Figure 50: **Top panel**: the ratio of input and output effective radii as a function of output radii values, measured in pixels. **Bottom panel**: the ratio of input and output effective radii as a function of estimated host galaxy magnitude. Color code represents the morphological type of the host galaxies, blue for late-type galaxies and red for early-type/spheroidal.

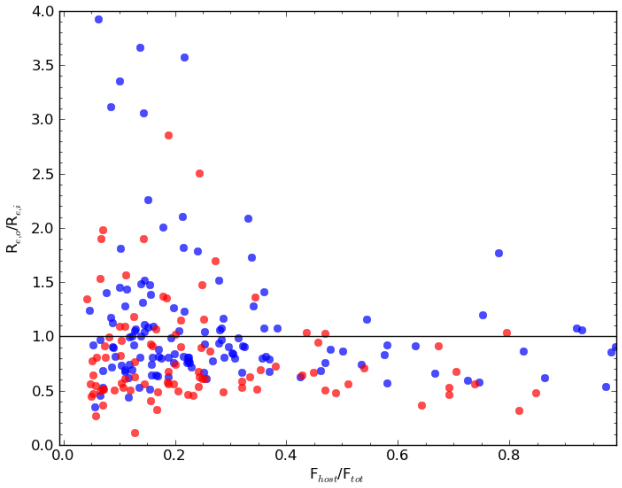

Figure 51: The ratio of input and output effective radii as a function of host-to-total ratio of mock AGN hosts. Color code represents the morphological type of the host galaxies, blue for late-type galaxies and red for early-type/spheroidal.

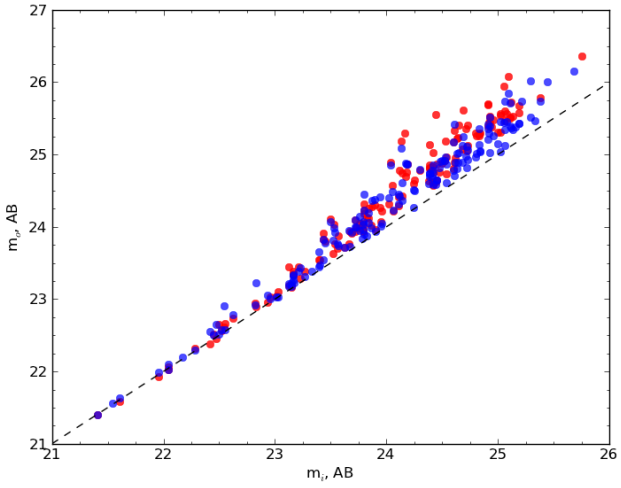

Figure 52: Relation between input and estimated magnitude of the simulated non-AGN galaxies. Red circles are early-type galaxies, blue circles are late-type galaxies. The dashed line marks the 1:1 relation.

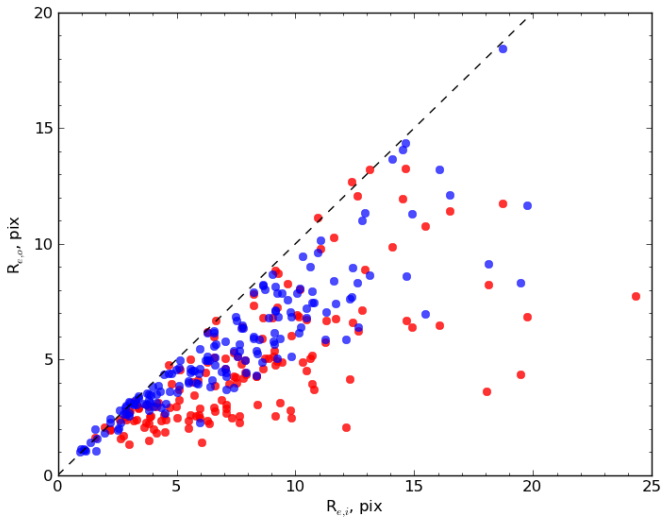

Figure 53: The relation between input and estimated effective radii for early-type morphologies (red circles) and late-type morphologies (blue circles) in the units of pixels. The dashed black line is the 1:1 relation.

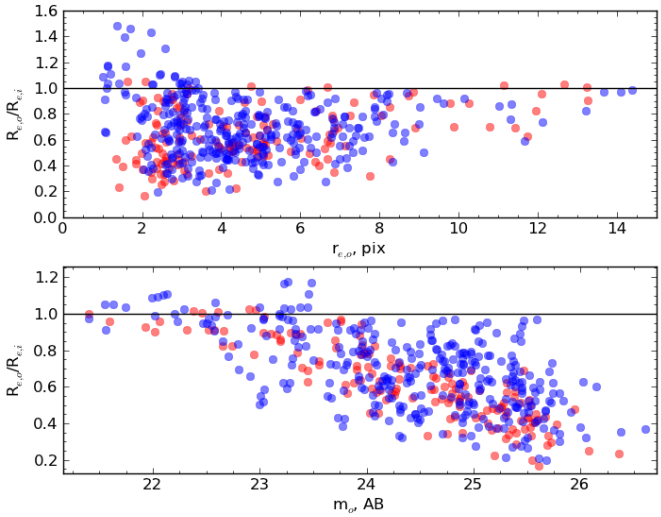

Figure 54: **Top panel**: The ratio of input to estimated effective radii for early-type morphologies (red circles) and late-type morphologies (blue circles) as a function of the output effective radii in pixels. **Bottom panel**: Same ratio as a function of galaxies' magnitude.

For the estimation of effective radii of the quiescent mock galaxies, the absence of a bright PSF component plays a very important role. Figure 53 shows the relation between the estimated and input values effective radii of early and late-type non-AGN galaxies. Same as for the AGN hosts, the quality of the effective radii estimation for the disc-dominated galaxies is slightly better, showing less scatter from the input values than early-type galaxies.

Figure 54 has two panels that show the ratios of output to input effective radii as a function of estimated effective radii and brightness of galaxies. In the top panel one can see that almost over the entire range of effective radii, the recovered values are underestimated, with the error of estimate going down towards larger recovered sizes. There are only few objects with magnitudes $m < 23.5\ AB$, which have their effective radii overestimated. Their effective radii estimates R_e <3 pixels (0.09″), however, point out that they are below the PSF resolution limit. The bottom panel of Figure 54 shows a correlation between the degree of underestimation of the effective radii and the galaxies' brightness, which happens naturally due to decreasing signal-to-noise. More information on the quality of the GALFIT parameters estimation for a single Sérsic profile mock galaxies can be found in Häussler et al. (2007).

11 Results and Discussion

The COSMOS survey is a unique field given its size and diapason of wavelength coverage. It provides comprehensive ground- and space-based observations of ~ 2 deg^2 of the sky over the entire spectral range of X-ray, UV, optical, IR and radio bands. These allow to construct statistically large samples of extragalactic sources, with SEDs and high quality photometric redshifts available for most objects. COSMOS field has complete coverage with Advance Camera for Surveys of HST in one photometric F814W band. Employing these high-resolution single-band photometric data, we study the morphological properties of galaxies harbouring AGNs in comparison to normal non-AGN galaxies and constrain the evolution of their properties with redshift. We will also compare the fraction AGN hosts and quiescent galaxies with disturbed morphologies to make inferences about mechanisms for fueling star-formation and AGN activity in these galaxies.

Resolving high-redshift Type I AGN hosts is a very challenging task. Primarily, due to the contrast between the central point-source and underlying host galaxy. In our particular case, this task is even more complicated due to the fact that the F814W filter corresponds to the rest-frame UV emission of the hosts at the redshift range under study. At such rest-frame band, the AGN emission significantly exceeds the stellar flux of the host. Furthermore, the UV emission originates from young stars in star-forming regions of the host galaxies, the distribution of which makes galaxies appear clumpy an irregular.

In the previous section we described how we visually classified both samples of galaxies, assigning them to five classes according to the disturbance signatures they have. HST provides an excellent spatial resolution, that allows us to separate fluxes of members in close groups, which would not be possible to resolve even with the biggest ground-based telescopes (at least not without adaptive optics). It also has great sensitivity, which allows to identify faint tidal features even in the presence of bright central nucleus. We classified the AGN sample, assigning a flag for the class of disturbance to each object in the sample. Before performing the classification, we subtracted the central point source from each object. This allows to de-blend companions and often helps to identify faint low signal-to-noise tidal features, which are not visible otherwise. For our samples we split galaxies into early-types and late-types using the best-fit Sérsic indices (n=1 for late-type galaxies and n=4 for early-type). We did not perform a morphological classification by eye, since very few of the objects have any signatures that may allow to assign them uniquely to a particular morphological type. It is known that the Hubble sequence holds at least until redshift $z \sim 1$, with stellar mass-density dominated by red, evolved galaxies (Bell et al., 2004). But already at redshifts $z \geq 1.5$ the Hubble sequence is no longer applicable (van den Bergh, 2001; van den Bergh et al., 2002). Galaxies at these redshifts are reported to be dominated by young stellar populations, and appear highly irregular and clumpy or exceptionally compact (Giavalisco et al., 1996; Lowenthal et al., 1997; Daddi et al., 2004). These irregularities become even more pronounced if observed in the rest-frame UV emission

at redshifts $z \geq 1.5$. Visual inspection may reveal features that favour a certain morphological type over another, i.e. a presence of a tidal tail should indicate a later type morphology for a galaxy on a low redshift, but this is not necessarily true for galaxies at high redshift. Such objects should be treated with care: fitting the entire brightness profile of a spheroid-dominated (Sérsic index $n > 2.5$) galaxy including a tidal tail, would result in estimates of n lower than the intrinsic value for a pure spheroid.

Further on, if not stated otherwise, we will refer to all those hosts in our sample that are flagged '1' and '1/2' as heavily disturbed galaxies as described in Section 9. To this group belong those objects, that show clearly pronounced signatures of interaction: if two or more galaxies of similar brightness are interacting they are categorized as MM, and if a galaxy shows tidal tails around a much fainter companion, or no companion is visible, such an object was categorized as a minor merger. The lopsided galaxies, flagged '2', show large offsets between the centers of their individual fitted components and are considered as galaxies with post-merger signatures. In the same way, a classification was made for the quiescent galaxies sample.

Single-band data cannot provide a way to constrain star-formation histories of galaxies. However, if we adopt the "continuous star-formation" approximation we can estimate the current star formation rates in each galaxy. The approximation implies that the UV emission in a galaxy origins from star-formation that is constant over the last < 100 Myr. Then, if we also assume that the amount of dust in these galaxies is negligible, we can expect the blue part of their SEDs to be essentially flat (in terms of f_ν) in the far UV range (1500-2800 Å), which results in a very blue colour as shown by Kennicutt (1998). The rest-frame UV flux in such objects can be converted to star-formation rates using

$$SFR\,(M_\odot\,yr^{-1}) = 1.4 \times 10^{-28} L_\nu(ergs\ s^{-1}\ Hz^{-1}),$$

which is a conversion of the Madau et al. (1998) calibration to a Salpeter (1955) Initial Mass Function with limits of 0.1 to 100 $M_\odot$. K-corrections for such star-forming, unabsorbed galaxies will be negligible, so we can convert observed apparent magnitudes into rest-frame UV luminosities and star-formation rates. We reckon that in our redshift range such an approximation is fair and provides good estimates of absolute luminosities for the star-forming AGN hosts and non-AGN galaxies. Therefore we apply these assumptions to study properties of our sample of AGN hosts.

The surface-brightness profiles of high-redshift QSOs are hard to fit not only because of contrast issues. Very important is also the noise, which impedes the recovery of a brightness profile's intrinsic parameters, especially for the faintest objects, where it causes a systematic underestimation. The PSF is the crucial factor in a successful separation of the central nuclear point source flux from the underlying host galaxy, therefore any mismatch between the modeled PSF and the real data, as described in the Section 8, leads to an increase of fitting error with

decreasing host-to-nucleus ratio. However, since the difference between the real and modeled PSF does not change significantly for different galaxies, a successful detection of the host galaxy will depend mainly on a good estimation of the guess input parameters for fitting and, in some cases, a manual inclusion of additional components to de-blend interacting objects. In this Chapter we are going to show the results of our very careful 2-dimensional modeling of the brightness profiles in our samples of AGN host and quiescent galaxies. We compiled a large catalog of resolved AGN host galaxies, that provides good statistics over the entire redshift range of the sample, and allows to compare recovered properties of AGN hosts to predictions from theoretical studies and results of other surveys. We also created a sample of matching quiescent galaxies, which were selected to have the same rest-frame UV luminosity and redshift as each corresponding AGN host. We will also compare the environments in which these galaxies reside. Our Survey is at least a factor of 3 larger than any existing survey of high-redshift AGN hosts compiled until now, so it can provide much better statistics than any other sample to date. Besides being able to compare morphological properties and constrain their evolution with redshift, we also analyzed environments in which AGNs and non-AGN galaxies reside: depending on how common interactions are for hosts in our sample in comparison to normal non-AGN galaxies, we can make conclusions about which processes are more likely to trigger AGN activity and star-formation in these of objects. Table 5 in Appendix A summarizes the final catalog of resolved AGN host galaxies and properties of their AGN taken from Brusa et al. (2010).

We had already described the methods we applied to all objects in our samples to model their brightness profiles and characterize their morphology. After we performed the fitting of the AGN sample we found that 265 out of 350 objects can be resolved into a sum of a point-source + Sérsic profile for a disc (n=1) or early-type/spheroidal (n=4) galaxy model. This constitutes by far the biggest sample of high-redshift ($z > 1.5$) AGN host galaxies to date: so far the largest study of X-ray selected AGN hosts on the redshift $z > 1.5$ had 72 objects, observed within the CANDELS Survey (Kocevski et al., 2012). A number of studies of X-ray selected high-redshift AGN host galaxies have even smaller samples of about 30 objects (Schawinski et al., 2011, 2012; Jahnke et al., 2004; Kotilainen et al., 2009). Most of the studies performed in the last 10 years use GALFIT as a standard tool to analyze brightness profiles, the same as we used for our survey. We have taken great care to ensure that each GALFIT input parameter set converges to the model that is less affected by degeneracies caused by noise, low host-to-nucleus ratio and PSF mismatch. 2-D modeling provides a good separation between the fluxes of different components, allowing us to recover parameters of host galaxies in 76 % of the cases for the sample of X-ray selected AGNs suitable for analysis. The host-to-total ratios for the resolved galaxies span a large range starting from as few as 3 % for Type I AGN to about 80% for several Type II AGN in our sample. The lower limit of the resolved host flux shows that GALFIT provides a high sensitivity to resolve hosts in the presence of a bright central point source even if there is a slight mismatch between real and modeled PSFs, which for this study were created

with Tiny Tim individually for each galaxy. We would like to emphasize again that our data set is observed in a band that corresponds to the rest-frame UV emission, which further on aggravates the difficulty of detecting a host, due to the clumpiness and irregularity of high-redshift star-forming galaxies.

11.1 The Size-Luminosity relation

One of the approaches to constrain galaxy evolution is to study fundamental properties of galaxies such as luminosity/absolute magnitude, size and mass, and the scaling relations between them. By comparing scaling relations at different epochs, they can be used to constrain the evolution of galaxies. In the case of AGN hosts this will help to better understand the connection between AGN activity/SMBH growth and growth of the stellar component of its mother galaxy, and also may help to reveal triggers of nuclear activity in a galaxy.

One of such scaling relations, namely the size-absolute magnitude relation of the recovered AGN host galaxies is shown on the top two plots of Figure 55 separately for early-type (on the top panels of each plot) and disc-dominated morphologies (on the bottom panels of each plot). Absolute magnitudes are computed using the formula $M = m - 5log(d_{L,Mpc}) - 25$, where d_L is the distance luminosity in units of Mpc and m is apparent rest-frame monochromatic magnitude. As we explained before, k-correction values are small for star-forming galaxies and we neglect them for the conversions of apparent magnitudes into rest-frame. For each successful run GALFIT returns a range of best-fit parameters, among which are the magnitude of the Sérsic component and its estimated effective radius. The estimate of the host galaxy magnitude is converted into rest-frame UV luminosity/absolute magnitude using the standard conversion:

$$f_{rf} = 10^{\frac{ZP-m_{obs}}{2.5}} \times PHOTFLAM \times (1+z)$$

$$\lambda_{rf} = \frac{\lambda_{f814w}}{(1+z)}$$

$$\nu_{rf} = \frac{c}{\lambda_{rf} \times 10^{-8}}$$

$$f_{\nu,UV} = f_{rf} \times \frac{\lambda_{rf}}{\nu_{rf}} =$$

$$= 10^{\frac{ZP-m_{obs}}{2.5}} \times PHOTFLAM \times (1+z) \left(\frac{\lambda_{f814w}}{(1+z)}\right)^2 \frac{10^{-8}}{c}$$

$$L_{\nu,UV} = 4\pi d^2 f_{\nu,UV},$$

where m_{obs} is the aparent magnitude estimated by GALFIT, PHOTFLAM is the ACS inverse sensitivity in units of $erg\ cm^{-2}\ s^{-1}$ Å^{-1}, z is redshift of the object, λ_{f814w} is the central wavelength of the F814W filter, c is the speed of light in units of $cm\ s^{-1}$, ZP is the F814W filter zero-point ($ZP = 25.959$ in AB magnitude), and

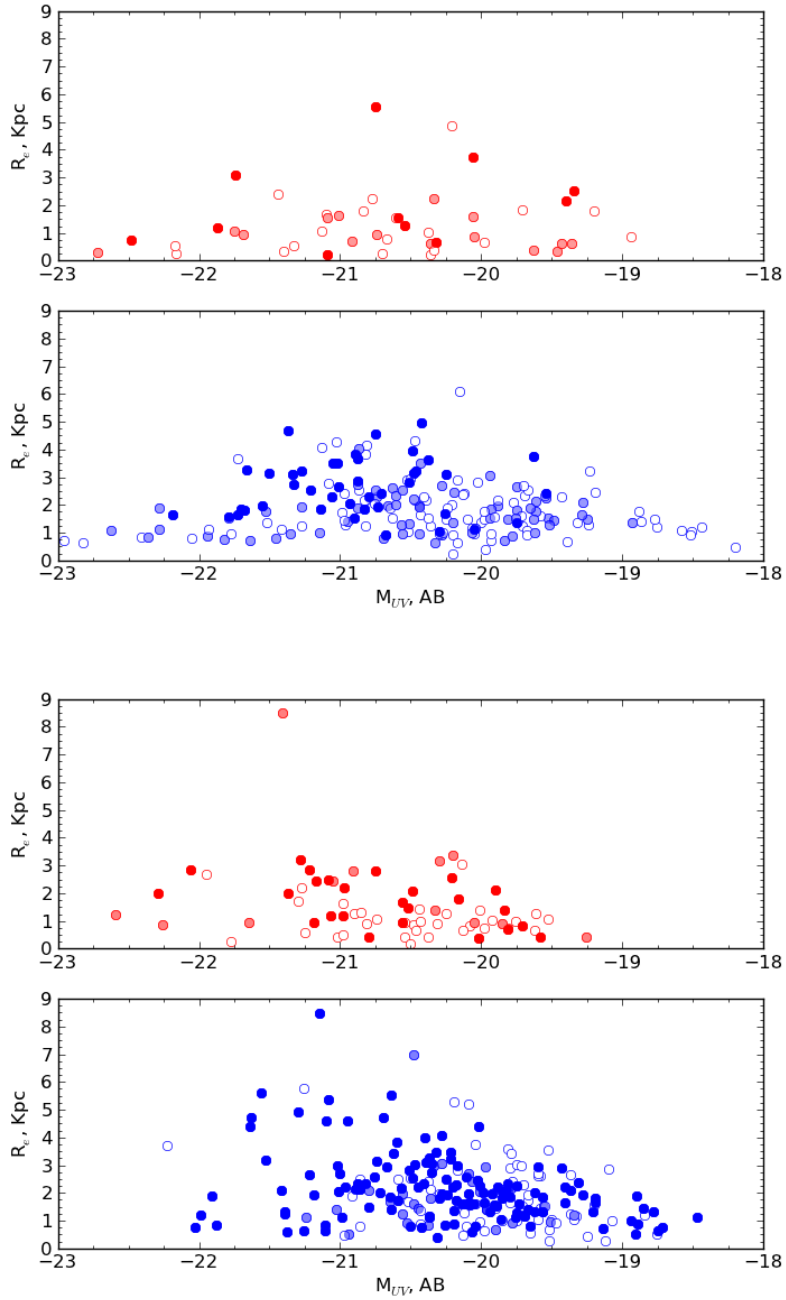

Figure 55: **Top two panels**: The estimated host galaxies radii and absolute magnitudes: early-type/ spheroidal morphologies on the top, disc galaxies on the bottom. **Bottom two panels**: Results of fitting of the quiescent galaxies sample for early-type/ spheroidal morphologies on the top, disc galaxies on the bottom. On all plots, the gradient of shading represents the stage of disturbance: filled circles are objects from the class of heavily interacting and very disturbed objects, flags '1' and '1/2'; shaded circles show galaxies classified as lopsided, flag 2; empty circles are galaxies which do not show any signatures of interactions, flags '0','2/3','3'.

d is the luminosity distance in the units of *cm*. The effective radius of a galaxy is defined as the radius encircling half of the galaxy's total flux. To convert this value into physical units, we used the drizzled ACS data pixel-scale (0.03"/pix) and the Standard ΛCDM Cosmology, assuming $H_0 = 70 kms^{-1} Mpc^{-1}$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, to estimate the angular size distance at the redshift of each AGN. In the following, we will use galaxy size and effective radius synonymously.

One of the questions that appears relating to observations of high-redshift galaxies is whether the estimated morphological type and also size of galaxies change when observed in different bands. The rest-frame morphology change is accounted for by the morphological k-correction and is known to exist both for local and high-redshifts galaxies (Windhorst et al., 2002; Papovich et al., 2003; Cassata et al., 2005). This happens because young and hot stars, which dominate the emission at the rest-frame UV tend to concentrate in the galactic disc. The change of the rest-frame morphology may lead not only to a wrong estimation of the morphological type of a galaxy. If a galaxy is observed in a band that corresponds to UV emission, as in the case of our survey, the galaxy may have clumps of star-formation distributed irregularly in the galactic plane, that can be falsely identified during deblending as nearby objects, therefore placing an additional error on the effective radii estimation and to the disturbance rate of the galaxies' environment. Some studies carried out on star-forming galaxies using observations at several wavelength ranges suggested that for continuously star-forming galaxies the morphological k-correction is not significant. For instance, comparing the morphological properties of non-AGN star-forming galaxies at redshifts $z > 1$, Papovich et al. (2005) claim they do not find significant changes in the rest-frame morphology between observations of a galaxy at the rest-frame UV and B bands. Thus, under the assumption that our galaxies are forming stars continuously over > 100 Myr, we can expect that the separation of our sample into early and late-type morphological types is not a consequence of the observations conducted in the rest-frame UV range and also that the estimates of effective radii by GALFIT are not biased.

Figure 55 clearly shows that the size-luminosity relation for AGN host galaxies is similar to the one of non-AGN galaxies, which is shown on the two bottom plots. Moreover, for both kinds of galaxies, the size-absolute magnitude relations for different morphological types show essentially the same slope and scatter. A similar result is shown by Trujillo et al. (2006), where the authors show that at redshift $z > 1.4$, the size-luminosity relation for galaxies with low Sérsic indices looks alike to the one with high Sérsic indices, however the latter one has larger uncertainties. Thus, further on we do not split the size-luminosity relation into different morphological types for convenience.

Figure 55 shows the size-luminosity relations both for the AGN host galaxies and the control quiescent galaxies sample. To find if there is any evolution in these relations with redshift, we split both distributions into three redshift bins: $z < 1.8$, $1.8 \leq z < 2.3$ and $z \geq 2.3$. The result is shown in Figure 56, where AGN hosts are depicted in the top panel and quiescent galaxies in the bottom panel. The galaxies from the lowest redshift bin, shown with blue triangles, clearly have a

larger average size for a given luminosity than galaxies at higher redshifts. The two high-redshift bins in both samples, however, show a large overlap between them. The 2-dimensional KS-test returns 7% probability for the null-hypothesis rejection, implying that the samples are not significantly different. The excess of large galaxies in the high-redshift bin can be explained if we consider the results obtained from our mock QSO simulations. They showed that for a given quality of the PSF model, created with Tiny Tim, GALFIT returns biased measurements when recovering small values of input effective radii in the presence of a bright nucleus. As we have shown above, computed effective radii, that are smaller then $r_e \sim 8\ pix$, are subject to an overestimation. Recovering of the small effective radii may as well be limited by the PSF resolution ($FWHM_{ACS} \sim 4$ pixels)

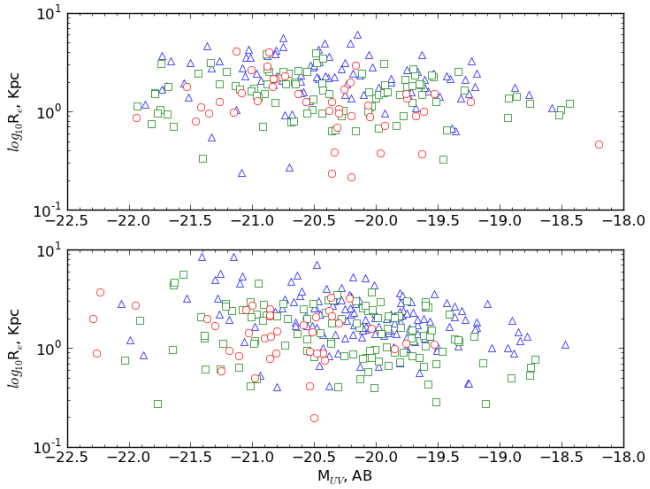

Figure 56: Size-luminosity relation evolution for AGN host galaxies (at the top) and quiescent galaxies (at the bottom). Blue triangles mark the lowest redshift bin $z < 1.8$, green squares the medium redshift bin, $1.8 \leq z < 2.3$, and red circles are galaxies at the highest redshifts $z \geq 2.3$.

The evolution of the size-magnitude relation can be interpreted in two ways: either a galaxy of a given size at higher redshift has a higher surface brightness than a galaxy of the same size at a lower redshift, or a galaxy of a given brightness at a high redshift is more compact than a galaxy of the same brightness at a lower redshift. To study the size-magnitude relation evolution in greater detail, we calculated apparent surface brightness values of each object. The surface brightness was computed as

$$\mu = m_h + 2.5log(2\pi r_e^2 b/a),$$

where m_h is the apparent host galaxy magnitude, estimated by GALFIT, r_e is the apparent effective radius, converted into *arcsec* using the ACS pixel-scale (not corrected for underestimation), and b/a is the minor-to-major axis ratio estimated by GALFIT. We plot estimated values as a function of redshift in Figure 57. This plot

shows a systematic increase of surface brightness towards higher redshift at least until redshift $z \sim 3$ for AGN hosts and over the entire redshift range for quiescent galaxies. Thus, for both samples, a galaxy at redshift $z = 3$ on average has a surface brightness ~ 0.5 mag/arcsec^2 higher than at redshift $z = 1.5$.

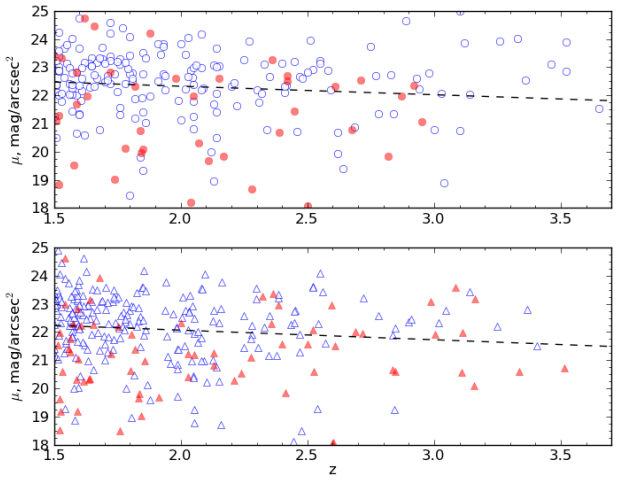

Figure 57: Estimated surface-brightness of AGN host galaxies (top plot) and quiescent galaxies (bottom plot) as a function of redshift. Colors depict morphological class:disc-dominated galaxies (blue color on both plots) or early-type galaxies (red color on both plots). Dashed black lines are linear fits to both samples (the AGN sample, however, is fitted only at $1.5 < z < 3.0$).

11.2 The galaxy size evolution

The ΛCDM Cosmological model of the Universe is currently accepted as the concordance Standard Model of the Universe, since it is the simplest model that can explain:

- the structure of Cosmic Microwave Background (CMB) with WMAP (Hinshaw et al., 2009; Reichardt et al., 2009; Brown et al., 2009; Komatsu et al., 2011)
- the Large Scale Structure of the Universe (Reid et al., 2010; Percival et al., 2010)
- the synthesis of the primordial elements, such as Hydrogen, Deuterium, Helium and Lithium (Burles & Tytler, 1998)
- the supernova data (Hicken et al., 2009; Kessler et al., 2009)
- strong and weak gravitational lensing effects Suyu et al. (2010); Schrabback et al. (2010)

In this model, a flat universe is dominated by Cold Dark Matter (CDM) and Dark Energy. The latter is associated with a cosmological constant Λ in the Einstein Field Equations, the factor that determines the current expansion of the Universe. The constant, denoted as Ω_Λ, is associated with the fraction of the Dark Energy of the total mass-energy of the Universe. The fraction of CDM together with the fraction of baryonic matter constitute the Ω_M constant. Thus, these constants determine the concordance model of the Universe with a flat metric of the spacetime:

$$\Omega_\Lambda + \Omega_M = \Omega_{tot} = 1.$$

The combined studies of the CMB anisotropy by WMAP, Large-scale structure and Type Ia Supernovae revealed best-fit values for cosmological parameters of about $\Omega_\Lambda = 0.7$, Ω_M=0.3, fraction of baryonic matter of Ω_b=0.04, and $H_0 = 70$ km s^{-1} Mpc^{-1}.

One of the most important consequences of the ΛCDM cosmological model is that it predicts the "bottom-up" hierarchical formation of the structures in the Universe. In this scenario, in the early Universe, the upper limit on the mass of a stable overdense region is rather small, thus small structures form first, and then they build up larger structures (for a review see Peacock (2003)). Consequently, low-mass galaxies form first in the early universe and further on they grow mass via merging and accretion.

According to the hierarchical scenario the size of the baryonic disc of a galaxy should constitute roughly a fixed fraction of the size of the dark matter halo in which it resides (Mao et al., 1998; Mo et al., 1998; Somerville et al., 2001; Bouwens & Silk, 2002). The virial radius of a dark matter halo scales with redshift and virial mass or circular velocity as

$$r_{vir}(z) = \left(\frac{GM_{vir}}{100H^2(z)}\right)^{\frac{1}{3}} = \frac{V_{vir}}{10H(z)}$$

where $H(z)$ is the Hubble parameter at redshift z:

$$H(z) = H_0[\Omega_M(1+z)^3 + \Omega_\Lambda]^{1/2}$$

(Carroll et al., 1992). If we assume that effective radius scales with virial radius and take into account that observations should better track the evolution at fixed luminosity rather than circular velocity or virial mass, we can expect the effective radius to scale somewhere in between these two functional forms shown above.

Figure 58 shows the evolution of effective radii with redshift for our samples of disc-dominated AGN hosts (blue triangles) and quiescent galaxies (red circles). 2-D KS statistics indicate that these two distributions are drawn from the same mother sample, therefore we treat them together. If we fit this joint distribution with a power law $r_e \propto (1+z)^{-\alpha}$, we get the best-fit value of $\alpha = 1.06 \pm 0.3$, shown as a green solid line. For the matter-dominated Universe, the Hubble parameter at redshift z can be approximated as $H(z) \sim (1+z)^{\frac{3}{2}}$, in that case our solution

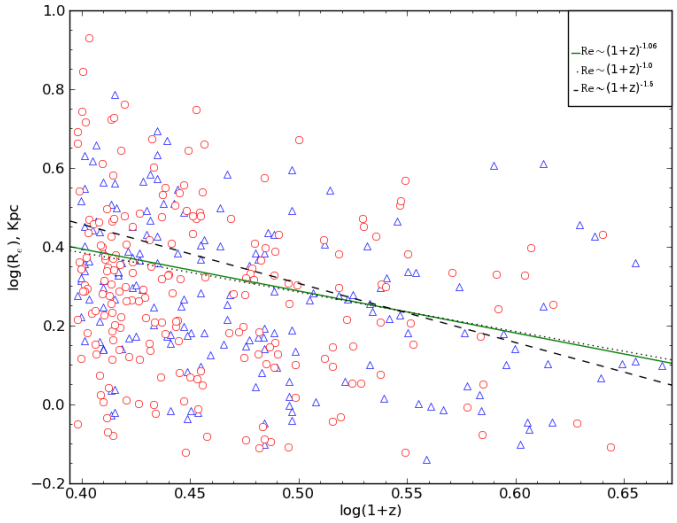

Figure 58: Evolution of the effective radii with redshift for AGN hosts (blue triangles) and quiescent galaxies (red circles). The green solid line marks the $(1+z)^{-1.06}$ power-law fit to the distribution of disc-dominated galaxies. The dashed line shows the trend expected in the hierarchical model of the Universe if galaxy sizes evolve as $H(z)^{-1}$ with fixed circular velocity, and the dotted line shows a scaling as $H(z)^{-2/3}$ with fixed virial mass. The curves are normalized to the mean size of a galaxy in our sample at redshift $z = 2.5$, which is approximately 1.7 kpc.

$r_e \sim (1+z)^{-1.06\pm0.3}$ is consistent with expectations for effective radii scaling at a fixed virial mass as $r_e \sim (1+z)^{-1.0}$.

The mean physical value of effective radii drops by $\sim 53\%$ between redshifts $z = 1.5$ and $z = 3.5$ (~ 2.5 kpc at $z = 1.5$ and ~ 1.3 kpc at $z = 3.5$) for our samples. In other studies that were done to calculate the evolution of effective radii (both AGN host and non-AGN galaxies), the results are in agreement with our measurements within the error bars (Schawinski et al., 2011, 2012; Papovich et al., 2005; Trujillo & Aguerri, 2004; Ferguson et al., 2004). For instance Papovich et al. (2005) have shown that the effective radii of two samples of galaxies, observed in the rest-frame B-band in two redshift bins $0.7 \leq z \leq 1.4$ and $1.9 \leq z \leq 3$ change by about 40%, with a mean size of ~2.3 kpc ($\sim 0.28''$) at $z \sim 2.3$. However, such comparisons have to be treated cautiously due to different selection techniques.

We have already shown the results of mock AGN and quiescent galaxies simulations we ran to check the GALFIT efficiency at recovering parameters of faint host galaxies of high-redshift AGNs. If we take into account the correction to the effective radii from the results of these simulations as shown in Figure 49, the power law index of the fit to the evolution of effective radii with redshift shown in Figure 58 will slightly change, as $\alpha = 1.52 \pm 0.2$, which is consistent with literature (for example Ferguson et al. (2004)). We plot the corrected effective radii evolution in Figure 59.

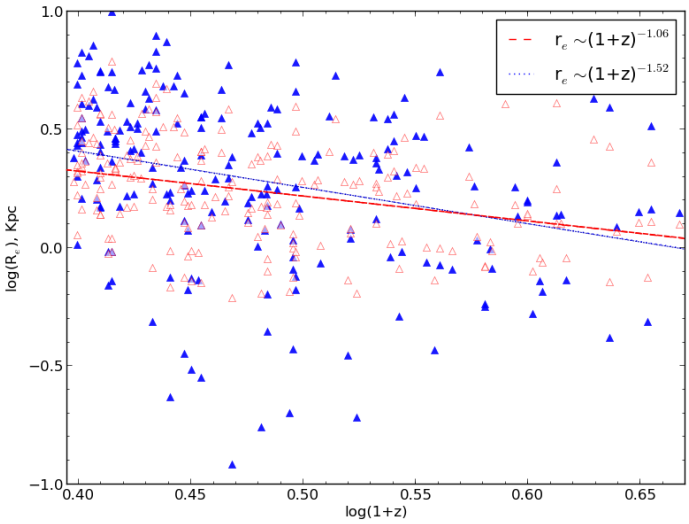

Figure 59: Corrected effective radii of AGN host galaxies as a function of redshift. Empty red triangles mark the old uncorrected values of AGN disc-dominated hosts and solid blue are the corrected ones. The red dashed line is a power-law fit to the old uncorrected distrubution, the dotted blue is the new power-law fit for the corrected values of disc-dominated AGN hosts, excluding few high-redshift large radii outliers.

11.3 Morphological properties of AGN host galaxies in comparison to quiescent galaxies

According to the current paradigm of hierarchical structure formation, Major Mergers (MM) are crucial in the growth of the present-day galaxies: numerical simulations have shown that classical bulges and early-type galaxies form as a result of MMs (Kawata & Gibson, 2003; Meza et al., 2003; Kobayashi, 2004; Naab et al., 2006; Hopkins et al., 2010). These kind of major gravitational interactions offer a natural and appealing explanation for the loss of angular momentum and transport of gas from galactic outskirts towards the center to supply QSO activity and nuclear star-formation. In this case, simplistic models used in simulations expect the majority of AGN host galaxies to be either in a heavily interacting state in case the AGN phase starts simultaneously with a merger, or to form an early-type galaxy or a significant spheroidal component, showing characteristic post-merger signatures in case the AGN phase is delayed with respect to MM event. However, depending on the initial conditions, some simulations have shown that the stellar discs can survive the collision (Barnes & Hernquist, 1996; Springel & Hernquist, 2005; Hopkins et al., 2009) or the merger remnant may re-grow a disc (Hopkins et al., 2009; Doménech-Moral et al., 2012; Bundy et al., 2010). It is still under debate whether in that case any significant BH fueling occurs and the galaxy goes through an AGN phase (Hopkins & Hernquist, 2009). Therefore, to draw conclusions regarding on processes are more likely to trigger AGN activity, simulations should be compared to observations of large samples of AGN host galaxies of a wide range

of luminosities and at different redshifts.

Several studies of moderate luminosity AGNs of $L_x \sim 10^{42} - 10^{44}$ erg s^{-1} at redshift $z > 1.5$ found a large fraction of their host galaxies to have disc-dominated morphologies, often hosting a significant spheroidal component (Kocevski et al., 2012; Schawinski et al., 2011, 2012). Visual classifications of these samples showed that a large number of galaxies show signatures of strong ongoing interactions. However, it appeared that the fraction of interacting objects in AGN host samples is not larger than in non-AGN samples. At lower redshifts $0.3 < z < 1.0$ Cisternas et al. (2011) find a similar prevalence of galaxies with late-type morphology over early-type galaxies, with a fraction of Major Mergers among AGN hosts even smaller than in the comparison sample. The authors interpret this as an argument that challenges MM as the main driver of AGN activity and BH growth of medium luminosity QSOs. Due to the lack of post-merger signatures in this sample of AGN host galaxies, it is more likely that other processes, like minor mergers and secular evolution, may be more responsible for transporting gas and triggering AGNs.

As we said before our sample of AGN host galaxies is significantly larger than any sample at similar redshifts that has ever been compiled until now. Therefore it can provide better statistics. To quantify galaxy morphology we used GALFIT estimates of the Sérsic index, because at the redshifts of our sample visual classification is highly ambiguous, due to the uncertainty of a Hubble classification above redshift $z \sim 1$ and low signal-to-noise of some PSF-subtracted galaxies. In the top panel of Figure 55 it is shown that in the sample of AGN hosts the disc-dominated morphologies significantly dominate over early-type galaxies, with a ratio of 79% to 21%. This result agrees with the above mentioned studies by Kocevski et al. (2012); Schawinski et al. (2011, 2012) and, as well, with the ratio between late-type and early-type host galaxies in the low-redshift study of Cisternas et al. (2011). Unlike the low-redshift regime, at redshift $z > 1.5$ early-type/spheroidal galaxies show a lot of recent star-formation and a similar rest-frame UV luminosity as late-type galaxies.

Two important questions are whether the prevalence of the late-type morphologies among AGN host galaxies is a feature unique to AGN hosts and if the ongoing star-formation in the early-type galaxies is connected to the AGN activity. We compared the results of the morphological classification of AGN hosts to that of the quiescent comparison sample. In the bottom panel of Figure 55 we show the results of single -érsic profile fits for our sample of quiescent galaxies. The prevalence of late-type morphologies over early-types, same as for AGN host galaxies, is obvious, with almost the same fractions of 76% of late-types galaxies and 24% of early-types. Also, the distribution of estimated UV luminosities of early-type quiescent galaxies is statistically identical to the one of AGN hosts. We show the normalized distribution of star-formation rates (SFR), estimated from the rest-frame UV luminosity as in Kennicutt (1998), in Figure 60. Thus we can conclude that, independently of whether a galaxy hosts an AGN or not, the late-type morphology is common among high-redshift star-forming galaxies and is observed in

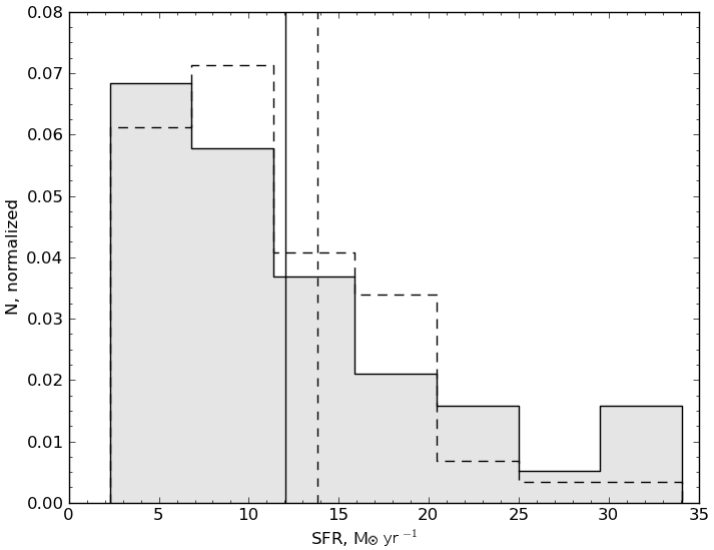

Figure 60: Normalized distributions of star-formation rates in early-type/spheroidal galaxies in the AGN host (grey histogram) and non-AGN galaxies (dashed-line histogram). Vertical solid and dashed lines show mean SFR value of both distributions.

~75% of the galaxies. The remaining ~25% show early-type/spheroidal morphology and a lot of ongoing star-formation activity, contrary to low-redshift early-type galaxies.

The observed morphology distribution places another question mark to the question of what triggers AGN activity, since it is highly unlikely that the majority of disc-dominated galaxies could have recently gone through major interactions and preserved their stellar discs. However, an alternative to MM mechanisms, like minor mergers and secular evolution (which operates on larger time-scales), provides an efficient way to transport gas to form stars and fuel AGN and seems very appealing and consistent with the observations.

11.4 Environments of AGN host and quiescent galaxies

In an attempt to study possible AGN triggering mechanisms, we investigated the environments in which AGN galaxies reside in comparison to the environments of those that do not show AGN activity. For instance, if AGN hosts reveal high disturbance rates in a comparison to non-AGN galaxies, with strong signatures of ongoing major interactions, it suggests that MMs play a significant role in transporting gas to the hosts' centers, that fuels instant AGN accretion. However, simulations of gas-rich mergers imply several different scenarios, which involve a delay between the merger and the start of AGN activity. In one scenario, the peak of star-formation and AGN activity happens on later stages of a merger, not long before the coalescence, while at the first passage only a modest star-burst occurs (Springel et al., 2005; Di Matteo et al., 2005). Other ad hoc models, that include obscuration,

show that at early stages of a merger the gas, that was driven to the central region of a galaxy during the interaction, is optically thick, with column densities high enough to completely obscure the X-ray emission of the AGN, as is observed in the local universe. In this case, the AGN may remain obscured until the latest stage of a merger. This time may be enough for the most prominent post-merger features to disappear, so that the host galaxy only shows faint post-merger signatures or even a relaxed undisturbed state (Hopkins et al., 2007, 2005; Lotz et al., 2010).

MM were initially associated with powerful nuclear star-bursts (Gunn, 1979), and the connection to QSOs was made later on. Observational studies of luminous quasars with close companions, showed signatures of recent interactions whenever it was possible to resolve the hosts (Stockton, 1982; Heckman et al., 1984; Gehren et al., 1984; Hutchings et al., 1984, 1988). These observations lead to a model, in which the AGN is triggered by an interaction with a massive companion. In 1983 with the launch of the Infrared Astronomy Satellite (IRAS), Ultra Luminous Infrared Galaxies (ULIRG) were discovered, this strengthened the QSO-merger connection scenario. More than 95% of ULIRGs are observed to be in a merging state, and many of them contain AGNs. After this discovery ULIRGs were interpreted as an early stage of a QSO evolution chain (Sanders et al., 1988a,b; Sanders & Mirabel, 1996; Surace et al., 1998; Surace & Sanders, 1999). Various simulations of MM are capable of reproducing different aspects of the observations, for instance SMBH fueling provided by the MM process, as observed in ULIRGs (Barnes & Hernquist, 1996, 1991; Mihos & Hernquist, 1996; Springel, 2000; Barnes, 2004). It is harder to reproduce the quantitative results of the observations with simulations, for example the ratio between MM and stochastically fueled AGNs. A model by Hopkins & Hernquist (2006) suggests that the most luminous AGNs $L_x > 10^{43}$ erg s^{-1} at low redshift should be triggered by MMs, as it is observed for the majority of ULIRGs. On the other hand, lower-luminosity AGNs should be fueled by stochastic accretion of gas transported within a galaxy by means of minor mergers and/or secular evolution. As a consequence, in this luminosity- and redshift dependent model, galaxies that are hosting more luminous AGNs should have a spheroidal component, which results from MMs. However, this model also predicts that at the peak of AGN accretion around redshift z∼2 the majority of AGN population should be already at luminosity $L_x \sim 10^{43}$ erg s^{-1}. This happens because in this model merger-driven accretion is coupled to cosmological galaxy merger rate that rapidly grows with redshift (Conselice et al., 2003), whereas secular evolution is coupled with the mass function and gas fraction in late-type galaxies, that evolves with redshift much slower than the cosmological galaxy merger rate. Thus, this model predicts an equal number of AGNs fed by MM and stochastically driven accretion at redshift $z \sim 2$ for AGNs with luminosity $L_X \sim 10^{42}$ erg s^{-1}. This result of simulations contradicts the observations, which find a significantly higher fraction of late-type disc-dominated host galaxies of moderate luminosity AGNs with $L_x \sim 10^{43} - 10^{44}$ erg s^{-1}, which means that stochastic BH fueling plays a more important role for moderate-luminosity AGNs (Schawinski et al., 2011, 2012; Kocevski et al., 2012).

As we just discussed, the Hopkins & Hernquist (2009) model predicts an even number of AGNs fueled by MMs and stochastic accretion at redshift $z \sim 2$ around luminosity $L_X \sim 10^{42}$ erg s^{-1}. Even though a comparison with observations previously showed this model to significantly underestimate the role of stochastic AGN fueling, one can expect that: i) the higher an AGN's luminosity, the higher the probability that its host recently experienced a MM, and ii) if it did, it should show signatures of a recent interaction. As we have shown, galaxies with late-type morphologies significantly dominate over early-types in our sample of AGN hosts, which makes it improbable that most of these galaxies experienced a recent MM. Nevertheless, we performed a classification of the disturbance of the AGN hosts (and the quiescent galaxies) to check if there is any correlation between the MM fraction and AGN luminosity. Since our AGN sample spans the range from medium to high luminosity $L_x \sim 10^{43} - 10^{45}$ erg s^{-1}, we could naively expect, that galaxies showing disturbed morphologies will have systematically more luminous AGNs. We plot the distribution of X-ray luminosities of our AGN sample in Figure 61. The top panel shows the distributions of X-ray luminosities of 3 classes of disturbed galaxies, with major interactions shown in blue, minor interactions/tidal signatures shown with a solid black line histogram, and lopsided galaxies shown with a dashed black histogram. On the bottom panel we show the distributions of luminosities for undisturbed objects: galaxies with slightly irregular profiles are shown with a dashed line and those without and signatures of disturbance are shown with a solid line. This plot shows no obvious trend for disturbed galaxies to have higher AGN X-ray luminosities.

We have discussed the remarkable similarity of the properties of AGN host and non-AGN galaxies. We found that three quarters of galaxies in both samples have disc-dominated morphologies, which indicates that Major Mergers do not play an important role in the SMBH fueling for a given X-ray luminosity range as it was previously expected. Another direct and straightforward test to check if AGN host galaxies are triggered by MM is to compare the disturbance rates among the AGN hosts and non-AGN galaxies. In Figure 55 we showed the classification of environments of galaxies according to the scheme described in the Section 9: heavily disturbed galaxies of both morphological types are marked with solid circles (of red colour for early-type and blue colour for late-type galaxies). To this group we assigned objects from two classes of disturbance: major mergers and minor mergers, since sometimes it is hard to place a distinct cut between them. Objects from these classes show strong signatures of interactions and differ only by the luminosity of the AGN companions. It should be kept in mind, that these two classes represent different SMBH fueling mechanisms, but it is easy to falsely classify them, therefore the estimate of the MM fraction is very rough. Another source of error is introduced by the high clumpiness and irregularity of the galaxies' appearance in the rest-frame UV: it is possible that an interacting group we identified in reality belongs to the same galaxy that has several prominent star-forming regions. In Figure 62 we show examples of galaxies (AGN hosts on the left and quiescent galaxies on the right) classified as heavily disturbed. Further on, the shaded circles

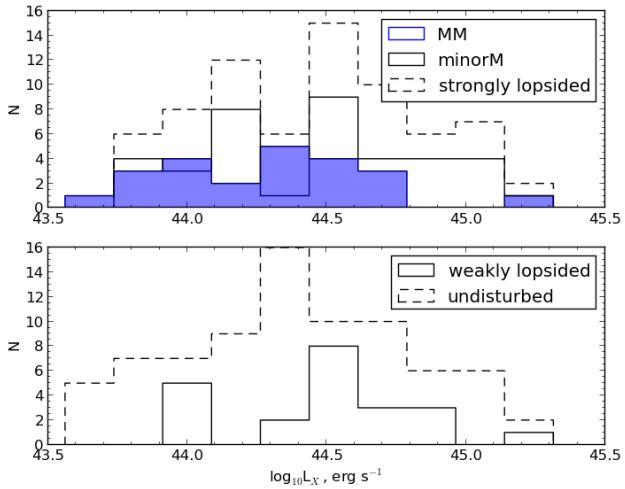

Figure 61: Distributions of X-ray luminosities for AGN of different class of disturbance. In the top panel we show the class of disturbed galaxies: shaded blue are MM, the solid black line depicts minor mergers and the dashed line represents lopsided hosts. The bottom panel shows undisturbed morphologies: the solid line shows galaxies with slightly irregular profiles, and the dashed line represents entirely undisturbed morphologies. Median of each distribution is shown with corresponding line style.

in Figure 55 represent the class of strongly lopsided galaxies, which we consider as a post-merger signature. The components of lopsided galaxies are strongly misaligned with each other and often reveal clumpy knots of star-formation, but only weak or no signs of tidal tails. Galaxies, which have less strong lopsidedness, which is quantified as the distance between the centers of the fitted components $d_{psf-sersic} < 2.5$ pixel ($0.075''$) (for the AGN sample), or slightly irregular brightness profiles (for the quiescent sample) were plotted together with undisturbed objects marked with open circles.

The final results of the disturbance classification are shown in Table 4. The comparison shows that the fraction of disturbed objects in the sample of quiescent galaxies is roughly two times that found in the AGN host sample. However, this maybe explained by the difficulty of defining galaxies as lopsided since we performed one-component fits. Still, the fraction of undisturbed galaxies points out that quiescent galaxies at these redshifts are more often found in a state of interaction. Also, the fraction of MM among quiescent galaxies is higher than for the AGN host. However there is a possibility, that there might be a hidden AGN, buried by surrounding gas and dust that is brought to the galactic center during an interaction. For instance, if the column densities for gas-rich MM are as high as $N_H > 10^{24}$ cm^{-2} the X-ray emission of up to $\sim$ 10 keV of such AGN will be completely obscured. Therefore these galaxies may not be classified as AGN. Nevertheless, the ubiquity of disturbed galaxies in the quiescent galaxy sample in

a comparison to AGN host sample once again implies that MM do not play an important role in fueling AGN activity, at least for the medium AGN luminosity range at redshifts $1.5 < z < 3.5$.

Morphological type	Disturbed	Lopsided	Undisturbed	MM fraction
AGN host galaxies				
Early type	29 %	27 %	44 %	8 %
Late type	21 %	32 %	47 %	9 %
Total	26 %	32 %	42 %	9 %
Quiescent galaxies				
Early type	37 %	17 %	46 %	13 %
Late type	57 %	14 %	29 %	22 %
Total	53 %	14 %	33 %	15 %

Table 4: Fractions of the disturbed and undisturbed objects among AGN host and quiescent galaxies of different morphological types.

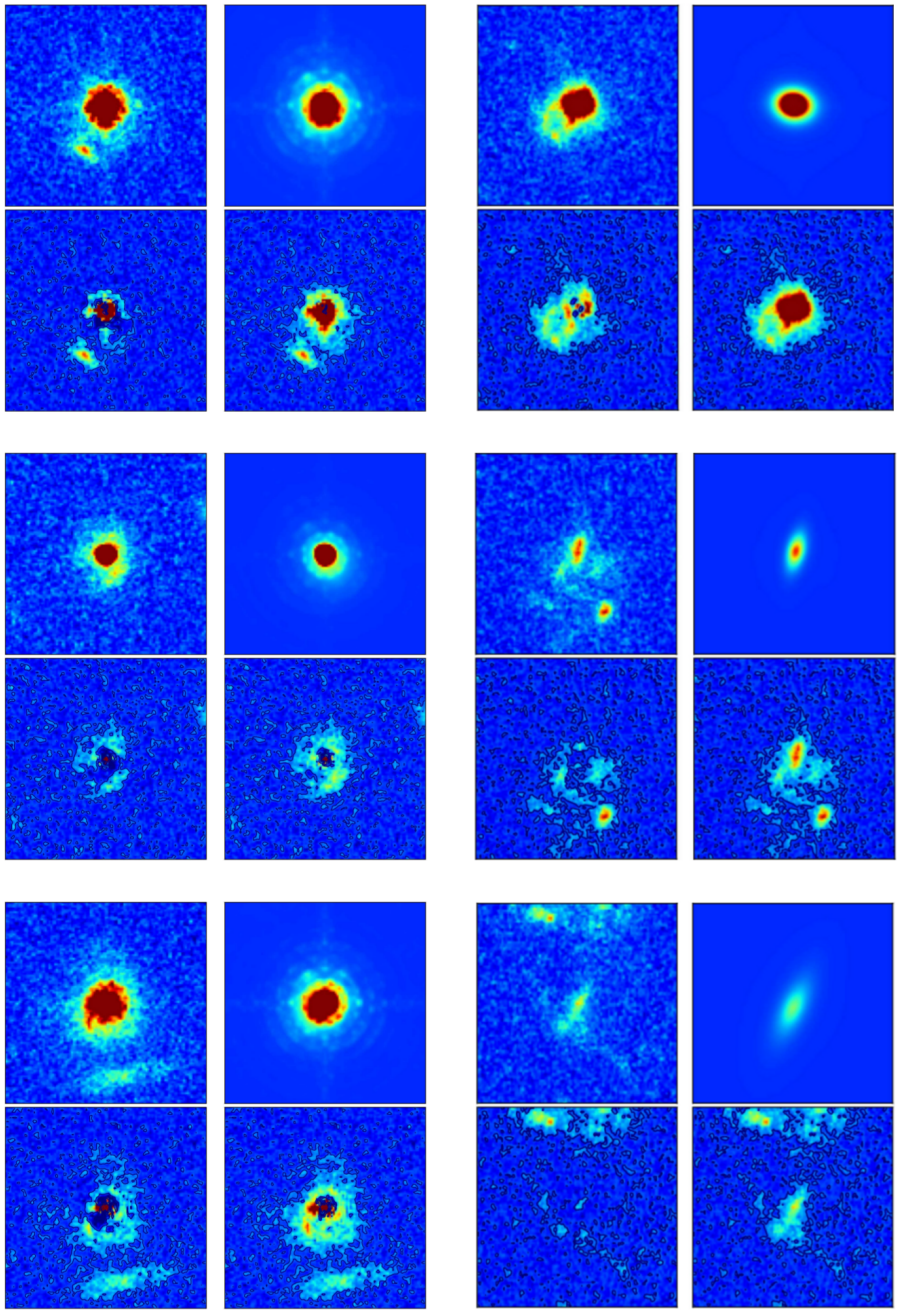

Figure 62: Examples of disturbed galaxies in the AGN host (on the left) and non-AGN galaxies (on the right-hand side). Each object panel has 4 thumbnails: top left is the cutout from the ACS tile, top right is the GALFIT model, bottom left is the residual of the image, and bottom right is the PSF-subtracted image, which essentially is the host galaxy. For the quiescent galaxies on the right, the bottom right plot is the original image, since we applied a one-component fit.

12 Conclusions and future perspectives

In this thesis we described the results of a survey aimed at resolving and studying properties of the host galaxies of high-redshift X-ray selected luminous QSOs from the COSMOS Survey. Most of the objects in our sample are unobscured AGNs, with either a spectroscopic or high-quality photometric redshift available. Our survey covers the redshift range $1.5 < z < 3.7$ covered by our survey, which corresponds to the epoch of the peak of AGN and star-forming activity. Almost all the studies of high-redshift AGN hosts conducted until now, were done within surveys that covered relatively small areas on the sky with multi-band photometry, and only a small number of sources available for the analysis. However, multi-wavelength observations allow to access the rest-frame optical range for high-redshift objects, which provides higher flux ratio between the host and nucleus in the rest-frame UV. The COSMOS Survey on the one has a large area and large number of sources available for the analysis, but on the other hand a single HST filter coverage, that corresponds to the rest-frame UV at redshifts $z > 1.5$, which make it somewhat challenging to study the hosts of luminous AGNs.

We attempted to resolve host galaxies of a large sample of X-ray selected QSOs with $43.5 < logL_X < 45.5$ from the XMM COSMOS Survey. At the redshifts of our sample, galaxies no longer obey the Hubble sequence and instead often have a very irregular structure with bright knots of star-formation. Despite all these complications, we have shown that we could recover host galaxy's properties of $\sim 76\%$ of these X-ray selected QSOs using 2-D surface brightness profile fitting with GALFIT. We assumed a Sérsic+PSF model for every resolved AGN, allowing the Sérsic index to take values of n=1 for late-type disc morphologies and n=4 for early-type spheroidals. The final sample of hosts consists of 265 objects, all of which have measurements of apparent magnitudes and radii available, which we converted to absolute units using the Standard ΛCDM Cosmology. We did not apply k-corrections for these galaxies, because its value is negligible for actively star-forming galaxies. Additionally, we performed a visual classification of the degree of disturbance of these host galaxies, to test the recent finding that medium luminosity AGN are, probably, not triggered by Major Mergers, but rather secular processes.

To compare AGN hosts to non-AGN galaxies we have also created a sample of quiescent galaxies, matched to the hosts in absolute magnitude-redshift space. We performed a morphological analysis of the quiescent galaxies' properties the same way as for the AGN hosts, except for using a single Sérsic component fit. At the considered redshifts, UV-bright galaxies are frequently found in an interacting state, therefore fitting these objects often involved the manual inclusion of additional components to fit neighboring galaxies, which are often impossible to mask out. However, the accurate knowledge of initial guesses for the fitting parameters, that should be provided to GALFIT, is not as crucial as for AGN host galaxies, due to the absence of a PSF component in the fits.

After the fitting of galaxies in both samples, we compared their properties and

environments. These are the main achievements of this study:

1. We performed a data reduction of the HST/ACS COSMOS data, offering a better better sampling and stability than publicly available data in the MAST archive[1].

2. The resulting absolute size-magnitude relations of AGN hosts and non-AGN galaxies appear to be very similar over the entire redshift range, independently of morphological type, except for a number of very compact bright galaxies among AGN hosts, which, however, may be false detections. If we compare our results to apparent size-magnitude relations from other studies of galaxies of brighter apparent magnitudes they are fully consistent. Taking into account, that the observations are done at different rest-frame wavelengths, it suggests that the evolution of sizes of galaxies may not be changing significantly when observed at different bands.

3. We showed that galaxy sizes evolve as $\sim (1+z)^{-1.3}$. This scaling corresponds to predictions of the hierarchical model for scaling with the size of the dark matter halo, in which a galaxy resides: under the assumption, that the radius of the baryonic disc is a fixed fraction of the size of its dark matter halo, the baryonic disc radius is expected to scale at a fixed circular velocity as $\sim H(z)^{-1}$ or at a fixed virial mass as $\sim H(z)^{-\frac{2}{3}}$, where H(z) is the Hubble parameter $H(z) = H_0(\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda)^{\frac{1}{2}}$. The comparison of absolute radii to other studies is more complicated because of their different luminosity selection. Still, the absolute values we find are within the error bars of measurements from other studies. If we apply correction to the host galaxies radii implied by mock QSOs simulations, the scaling relation is measured to be in an even better agreement with theoretical prediction, scaling at a fixed circular velocity as $\sim (1+z)^{-1.5}$.

4. Allowing the Sérsic index to take values of n=1 for a late-type morphology and n=4 for an early-type/spheroidal we were able to determine the prevalent morphological type of the AGN host galaxies. According to numerical simulations, Major Mergers are responsible for the growth of a fraction of present day galaxies, delivering gas for star formation and AGN fueling from the outskirts of a galaxy to its center. But simulations also show, that after galaxies experience a Major Merger, they form an early-type galaxy or at least a significant bulge component. Our results are showing a significant prevalence of disc-dominated morphologies over early-type galaxies, which is suggesting that these galaxies probably did not recently go through a Major Merger stage. The prevalence of disc-dominated morphologies was also previously reported both for low-redshift and high-redshift host galaxies of medium-luminosity AGNs, therefore our results, which provide much better

[1] at the moment when this project was initiated

statistics due to the large number of objects, confirm and strengthen these findings at redshifts $1.5 < z < 3.7$.

5. To investigate further the possibility that AGN host galaxies from our sample have recently gone through a Major Merger phase, we categorized them according to the disturbance signatures they show and compared them to the quiescent galaxies. We found a significant number of objects in the AGN hosts sample, which are showing signatures of major interactions, as well as a large number of objects that have post-merger signatures, like tidal tails or highly off-centered components. However, quiescent galaxies show an even higher fraction of strongly interacting, irregular objects, than AGN hosts. There are two possible explanations for this result. Simulations show that at the first stage of an interaction, no AGN activity should occur. Furthermore, in gas-rich mergers the infall of gas to the center of a galaxy can occur in amounts large enough to hide even the hardest X-ray emission of AGN. In the second scenario simulations predict a delay between the interaction and the start of star formation, and the triggering of AGN activity. If this delay is larger than the relaxation time of a galaxy that went through a merger/interaction, most of the morphological signatures may have already disappeared. It is, however, very unlikely that Major Mergers are the preferred gas transportation mechanism, since we find the majority of AGN hosts to show a disc-dominated morphology.

The existence of the scaling relationships between the mass of the Super-massive Black Holes in the centers of galaxies and various properties of stellar bulges these BH reside in (such as M_{BH}-L_{bulge}, M_{BH}-M_{bulge}, M_{BH}-σ) is usually considered strong evidence for a co-evolution between BH and its mother galaxy. Simulations show that when a pair of galaxies are going through a gas-rich Major Merger, the gas, which is transferred to the center of a galaxy, feeds a nuclear starburst and AGN accretion. When the AGN reaches its peak accretion rate, the quasar might expel the remaining gas by mechanical (or even radiative feedback), shutting down both star-formation and AGN activity. Such a scenario of self-regulated BH growth explains the correlation between the black hole growth and stellar bulge properties. However, the mechanism of feedback is not clear: so far, radiation, outflows, winds, jets have been proposed, which separately or combined may provide the desired effect in simulations. But observationally, feedback is hard to detect. One of the few observational proofs was recently reported by Cano-Díaz et al. (2012), where they report on the detection of a strong, redshift z=2.4 AGN outflow, with the star-formation in the host galaxy being shut down in the region to which the outflow extends. For a general review of radiative and kinematic quasar feedback see Fabian (2012). In this context, our sample of star-forming high-redshift AGN hosts, which covers a large redshift and luminosity ranges, provides a very good base to conduct studies of radiative feedback using AO-assisted IFU observations.

A Appendix

XMM ID	RA (J2000)	Dec (J2000)	z	m_{host}, AB	R_e, arc-sec	n	b/a	m_{PSF}	Flag[1]	HR	$L_{0.5-2keV}$	$L_{2-10keV}$
00103	150.11695	1.92988	1.52	23.67	0.1530	4	0.76	23.90	12	-0.50	6.12e+43	9.98e+43
00110	150.09154	2.39908	2.47	25.39	0.2625	1	0.19	21.99	3	-0.61	3.53e+44	4.18e+44
00110	150.09154	2.39908	2.47	24.53	0.3207	1	0.68	22.01	2	-0.61	3.53e+44	4.18e+44
02232	150.48437	2.16204	1.64	24.11	0.1764	1	0.95	22.24	3	-0.34	7.69e+43	1.91e+44
05149	150.26490	1.98700	1.56	24.68	0.1947	1	0.87	25.38	0	-1.00	2.20e+43	-9.65e+43
00337	149.79177	2.08594	1.72	24.04	0.5928	1	0.53	25.79	12	-0.29	6.09e+43	1.69e+44
00436	150.30259	2.16113	1.82	21.61	0.0864	1	0.83	20.48	3	-0.54	1.37e+44	2.05e+44
00356	150.06991	2.10162	2.25	24.42	0.3135	1	0.56	23.26	2	-0.16	4.64e+43	-2.19e+44
05532	149.80403	2.28813	2.11	25.43	0.0786	4	0.13	23.92	2	-1.00	3.81e+43	-1.36e+44
05532	149.80403	2.28813	2.11	26.03	0.1548	4	0.19	23.81	3	-1.00	3.81e+43	-1.36e+44
00447	150.32879	2.38052	3.52	24.83	0.1788	1	0.80	24.78	3	0.01	1.47e+44	7.37e+44
00271	150.30890	1.91228	1.52	24.60	0.2517	1	0.82	23.66	2	-0.33	6.51e+43	1.62e+44
53584	150.32424	2.08907	1.80	24.64	0.1887	1	0.79	22.90	3	-1.00	5.81e+43	-2.28e+44
05166	149.79178	1.87289	1.57	23.84	0.2364	1	0.66	22.68	2	-0.54	7.70e+43	1.18e+44
01559	150.13272	2.16742	2.05	24.11	0.0960	1	0.84	25.45	2	-1.00	3.64e+43	-1.84e+44
05378	149.80799	2.64572	2.08	24.39	0.2358	1	0.79	21.94	2	-0.48	1.84e+44	3.18e+44
00199	149.74449	2.02756	2.45	19.84	0.0594	1	0.76	19.53	0	-0.51	9.90e+44	1.58e+45
00123	150.38279	2.55982	2.07	23.90	0.0867	4	0.78	22.64	2	-0.39	1.10e+44	2.44e+44
00391	150.32826	2.12496	1.78	23.40	0.1263	4	0.48	22.68	3	-0.34	1.08e+44	2.66e+44
00301	149.75634	2.11733	2.55	25.68	0.1917	1	0.46	24.45	2	-0.41	1.23e+44	2.59e+44
00038	150.24517	1.90009	1.56	21.54	0.0348	4	0.72	20.42	2	-0.52	2.61e+44	4.16e+44
00153	150.28482	2.39505	1.93	23.80	0.4638	1	0.77	21.66	12	-0.48	1.57e+44	2.80e+44
05428	150.55555	2.36919	1.61	24.83	0.2502	1	0.53	24.00	12	-1.00	3.16e+43	-9.05e+43
00160	150.15839	2.13960	1.82	22.83	0.3709	4	0.72	20.51	12	-0.09	5.32e+43	2.27e+44
00169	149.73590	2.02763	2.45	24.63	0.3093	1	0.60	22.97	2	-0.45	3.86e+44	7.23e+44
00320	150.34103	2.36718	1.59	24.62	0.1299	1	0.86	24.00	3	-0.48	4.22e+43	7.51e+43
02096	150.55905	2.10576	2.32	25.00	0.1416	1	0.53	22.38	2	-0.45	6.33e+44	1.23e+45
00165	150.18001	2.23129	2.15	25.96	0.1662	1	0.39	23.64	2	-0.47	1.46e+44	2.64e+44
00079	150.35358	2.34219	1.71	22.04	0.0984	1	0.87	20.94	3	-0.49	2.39e+44	4.01e+44
00189	150.12685	2.62659	1.84	24.61	0.0822	4	0.67	22.95	3	-0.40	2.75e+44	5.84e+44
02394	150.36472	2.14383	3.33	24.69	0.3639	1	0.65	24.24	12	-0.36	3.25e+44	7.47e+44
00498	150.31647	2.24678	2.49	23.12	0.1017	1	0.88	21.66	3	-0.58	1.48e+44	-5.13e+44
01548	150.02546	1.87768	1.80	24.61	0.3681	1	0.73	22.88	2	-1.00	2.71e+43	-7.46e+42
00169	149.73590	2.02763	2.45	24.48	0.2475	1	0.79	22.95	2	-0.45	3.86e+44	7.23e+44
00167	149.88715	2.11718	2.05	24.80	0.1677	1	0.61	22.65	3	-0.42	1.12e+44	2.35e+44
00318	150.24778	2.44215	3.03	25.33	0.1191	1	0.55	24.88	3	0.02	1.15e+44	6.13e+44
00197	150.16378	2.59768	1.59	23.84	0.3843	1	0.42	22.02	1	-0.57	6.15e+43	8.61e+43
00263	150.51906	2.32102	1.72	23.66	0.2739	1	0.67	23.14	12	-0.43	1.14e+44	2.25e+44
00076	150.23551	2.36176	2.50	24.82	0.0294	4	0.37	23.05	3	-0.55	4.26e+44	6.29e+44
00078	150.11959	2.29591	2.02	24.72	0.1341	1	0.88	25.49	1	-0.38	2.10e+44	4.75e+44
00285	149.82269	2.08966	2.05	23.79	0.1968	4	0.77	22.58	2	-0.59	1.29e+44	-1.40e+44
00084	150.29972	2.50692	1.50	24.12	0.4476	4	0.41	21.04	12	-0.51	8.32e+43	1.35e+44
00448	149.87442	2.03104	1.85	23.76	0.3048	1	0.73	24.37	3	-0.18	5.96e+43	2.13e+44
00294	149.90615	1.91723	1.63	23.76	0.1884	4	0.86	22.84	1	-0.42	1.28e+44	2.58e+44
00132	149.99176	2.13197	2.14	24.60	0.1110	1	0.68	22.86	2	-0.48	1.26e+44	2.25e+44
00127	150.00419	2.38917	1.85	24.91	0.2808	1	0.56	23.01	3	-0.30	8.23e+43	2.25e+44
00088	150.51814	2.52168	2.78	23.95	0.1434	1	0.70	22.94	23	-0.43	9.01e+44	1.84e+45
00139	150.04184	2.62948	1.55	23.72	0.3489	1	0.76	21.69	23	-0.37	1.43e+44	3.33e+44
00375	150.43815	2.41581	2.03	21.96	0.0777	1	0.60	20.85	23	-0.43	1.25e+44	2.45e+44

XMM ID	RA (J2000)	Dec (J2000)	z	m_{host}, AB	R_e, arc-sec	n	b/a	m_{PSF}	Flag[1]	HR	$L_{0.5-2keV}$	$L_{2-10keV}$
02148	150.37827	2.19641	1.52	23.82	0.2613	1	0.44	22.54	3	-0.30	8.89e+43	2.29e+44
70222	150.01888	2.16562	1.94	26.98	0.0741	1	0.35	24.88	3	1	-1.45e+43	1.29e+44
00525	149.92254	1.97925	1.39	23.33	0.0327	4	0.63	23.09	23	-0.49	3.30e+43	-4.08e+43
00357	150.37458	2.31023	2.15	23.45	0.2952	4	0.83	24.23	3	0.25	7.37e+43	6.63e+44
00108	150.12215	1.91318	1.57	25.40	0.2112	1	0.16	24.58	3	-0.32	8.21e+43	2.14e+44
30012	150.44012	2.24419	2.31	25.19	0.0900	1	0.82	23.58	2	-1.00	5.27e+43	-4.67e+44
01461	149.73878	2.36636	1.65	24.50	0.2373	1	0.73	23.49	2	-0.38	9.33e+43	2.14e+44
00245	150.25424	2.33063	2.46	24.63	0.1296	1	0.26	21.76	3	-0.27	8.19e+43	2.41e+44
00033	150.30956	2.39915	1.80	21.41	0.0957	4	0.76	19.33	3	-0.49	3.38e+44	5.83e+44
05531	150.38673	1.96663	1.53	23.16	0.2769	1	0.77	21.41	1	-0.40	5.03e+43	1.08e+44
00237	150.18296	2.24793	2.48	24.73	0.2064	1	0.53	23.29	3	-0.17	7.26e+43	2.61e+44
00175	149.72072	2.34902	1.54	23.43	0.4953	1	0.82	23.39	3	-0.52	2.61e+44	4.16e+44
00308	150.42195	2.38554	1.51	24.84	0.0756	4	0.89	24.20	2	-0.41	4.53e+43	9.52e+43
00322	150.40666	2.36550	2.04	23.39	0.0408	4	0.81	22.45	23	-1.00	6.25e+43	-3.48e+44
00068	150.04263	2.06329	2.02	24.65	0.2934	1	0.61	23.06	5	-0.45	2.36e+44	4.57e+44
02557	150.42257	2.01420	2.27	24.57	0.4317	1	0.47	21.99	2	-0.52	1.86e+44	2.93e+44
60102	150.19427	2.62361	1.51	24.92	0.2511	1	0.54	24.57	2	-1.00	1.11e+43	-1.72e+44
60367	149.64802	2.71675	1.720	24.83	0.4482	1	0.63	25.19	1	-1.00	1.48e+43	-1.03e+44
53723	150.44260	2.16629	2.13	24.81	0.1212	1	0.61	24.42	2	-1.00	4.42e+43	-1.90e+44
53610	150.22779	1.78681	1.68	23.53	0.4404	1	0.69	23.08	12	-1.00	2.84e+43	-1.20e+44
00282	149.87182	2.34290	1.74	23.15	0.066	4	0.82	22.17	3	-0.35	5.95e+43	1.45e+44
05138	149.83260	2.71076	2.12	24.54	0.0795	1	0.38	21.02	2	-0.47	3.10e+44	5.37e+44
00360	150.22495	2.28417	1.85	24.24	0.0765	4	0.59	24.05	2	-0.28	3.32e+43	-3.64e+44
60015	150.37801	2.31825	1.84	22.82	0.1146	1	0.76	25.52	0	0.11	-8.65e+43	1.11e+44
53749	150.26994	1.81628	2.39	24.42	0.0996	4	0.51	23.67	0	-1.00	4.94e+43	-1.58e+44
02483	150.52701	2.03845	1.65	24.25	0.2823	1	0.65	25.02	3	-0.28	5.74e+43	1.63e+44
00395	150.29918	2.63348	2.98	25.44	0.1797	1	0.37	23.49	12	-0.35	1.49e+44	3.63e+44
53584	150.32424	2.08907	1.80	24.41	0.1527	1	0.52	23.01	3	-1.00	5.81e+43	-2.28e+44
00094	150.19975	2.19087	1.51	23.27	0.2469	1	0.91	21.23	1	-0.43	1.06e+44	2.14e+44
00271	150.30890	1.91228	1.52	24.54	0.3015	1	0.61	23.63	3	-0.33	6.51e+43	1.62e+44
05243	150.58428	2.32161	1.52	24.18	0.3378	1	0.80	23.39	0	-0.38	3.96e+43	8.84e+43
05165	149.80064	1.87043	2.95	25.16	0.0897	4	0.45	23.65	12	-0.23	2.53e+44	8.01e+44
00563	150.23168	1.87490	1.76	24.45	0.1158	1	0.67	25.70	3	-1.00	2.22e+43	-1.41e+44
60026	150.38140	2.50605	1.81	24.93	0.1794	1	0.58	24.77	2	1	-8.04e+43	1.27e+44
53755	150.26240	1.79434	1.79	24.93	0.2124	1	0.57	24.50	2	-1.00	2.29e+43	-3.32e+44
53781	149.92535	2.68424	1.79	20.91	0.0279	4	0.72	21.68	1	-0.01	3.53e+43	1.73e+44
00415	149.96981	2.18349	1.56	24.52	0.1722	1	0.78	23.30	3	-1.00	1.59e+43	-1.27e+44
05428	150.55555	2.36919	1.61	24.72	0.2538	1	0.50	24.04	2	-1.00	3.16e+43	-9.05e+43
00567	150.03062	2.67879	1.94	24.05	0.2292	1	0.86	21.68	0	-0.42	9.17e+43	1.88e+44
00458	150.29125	2.29099	1.71	24.73	0.1908	1	0.68	23.56	3	-0.39	3.23e+43	-2.20e+44
00044	150.21466	2.20429	1.85	22.96	0.0849	1	0.79	21.18	2	-0.48	2.78e+44	4.92e+44
00290	149.90186	2.27069	2.09	25.13	0.1509	1	0.49	24.91	3	-1.00	4.41e+43	-2.52e+44
00128	150.19897	2.13254	2.16	22.62	0.2325	1	0.99	19.53	2	-0.30	1.23e+44	3.36e+44
00373	150.04680	1.86674	2.41	23.57	0.2319	1	0.84	24.12	1	-0.49	1.02e+44	-1.97e+44
00373	150.04680	1.86674	2.41	23.43	0.2247	1	0.93	24.16	1	-0.49	1.02e+44	-1.97e+44
00349	150.21076	2.39147	3.10	24.44	0.5415	1	0.91	23.14	23	-0.23	1.43e+44	4.55e+44
53249	150.31479	2.62875	2.33	25.12	0.2292	1	0.47	25.04	2	-1.00	4.40e+43	-6.68e+44
10360	149.73573	2.75776	2.82	25.06	0.0498	4	0.53	23.71	3	-1.00	7.12e+43	-2.28e+44
02391	150.30765	1.66688	1.56	22.93	0.3276	1	0.66	20.98	1	-0.32	8.74e+43	2.18e+44
00176	150.45413	2.80609	1.61	22.17	0.0333	4	0.65	21.50	3	-0.36	9.82e+43	2.34e+44
10889	149.59245	1.75676	1.96	18.60	0.0465	4	0.89	19.93	12	-0.61	5.00e+44	5.93e+44
54044	149.85702	2.77964	1.52	26.46	0.1734	1	0.68	24.10	23	-0.46	3.86e+43	-1.59e+44
31313	149.88897	2.84376	1.61	23.86	0.2595	1	0.49	24.63	3	-0.39	2.59e+44	5.80e+44
05544	149.62189	2.73831	1.89	22.45	0.1971	1	0.79	20.12	12	-0.55	1.37e+44	1.94e+44
60414	149.49565	1.72262	1.75	23.12	0.5610	1	0.54	22.99	1	-1.00	2.81e+43	-8.58e+43
01467	150.36184	2.77913	2.52	24.93	0.2124	1	0.56	22.40	12	-1.00	1e+44	-8.38e+44
05377	149.61070	2.70841	2.45	26.20	0.1086	4	0.17	23.15	23	-0.05	4.19e+43	1.85e+44
05218	150.74310	2.73212	2.71	24.24	0.1977	4	0.84	22.80	2	-0.30	1.55e+44	4.15e+44
02258	149.61152	2.35376	1.72	24.71	0.3204	1	0.46	24.28	2	-0.41	1.49e+44	3.08e+44
02609	149.72131	2.53998	1.74	23.21	0.3873	1	0.70	21.97	1	-0.58	1.06e+44	1.44e+44
05424	149.77770	1.69544	1.69	24.17	0.3717	1	0.76	22.82	12	-0.46	6.89e+43	-1.36e+44
53329	150.67403	2.53232	2.58	24.41	0.2730	1	0.53	22.97	23	-0.14	7.78e+43	-4.20e+44
05349	149.63927	2.00322	2.84	24.06	0.1248	1	0.84	24.41	2	0.03	1.50e+44	7.77e+44
05189	150.12506	2.86176	1.58	22.42	0.1404	4	0.56	19.74	12	-1.00	6.26e+43	-4.50e+44
02334	150.74258	1.98589	1.60	24.38	0.1143	1	0.75	22.10	3	-0.49	7.95e+43	-9.14e+43

XMM ID	RA (J2000)	Dec (J2000)	z	m_{host}, AB	R_e, arc-sec	n	b/a	m_{PSF}	Flag[1]	HR	$L_{0.5-2keV}$	$L_{2-10keV}$
02105	150.58128	1.92696	1.51	22.54	0.3933	1	0.86	20.62	1	-0.60	1.87e+44	2.36e+44
05323	149.56460	1.82311	1.51	22.47	0.1995	1	0.65	20.50	1	-0.41	8.98e+44	1.84e+45
05382	150.43992	2.70350	3.46	25.19	0.1749	1	0.76	23.75	2	-1.00	1.21e+44	-2.23e+45
02178	150.62149	2.15887	1.52	22.04	0.0666	4	0.66	20.57	3	-0.58	2.07e+44	2.72e+44
54308	149.52915	1.70081	1.71	24.47	0.2112	1	0.61	21.63	23	-0.32	7.23e+43	-2.87e+44
02169	150.63553	1.66915	1.78	23.56	0.2898	1	0.66	20.75	23	-0.57	1.60e+44	2.07e+44
02282	150.57492	1.97681	1.54	23.16	0.3315	1	0.90	21.68	3	-0.25	6.82e+43	1.96e+44
00109	150.45663	2.64818	2.05	23.01	0.189	1	0.87	20.61	12	-0.53	1.90e+44	2.94e+44
10257	150.69603	2.23638	1.60	24.16	0.7299	1	0.51	24.79	23	-1.00	1.95e+43	-1.21e+44
02265	150.05183	1.68280	2.27	22.52	0.0897	4	0.29	20.21	12	-0.36	2.84e+44	6.50e+44
02409	150.63146	2.00268	1.52	23.80	0.2724	1	0.92	23.71	3	-1.00	1.87e+43	-1.16e+44
05278	150.53860	2.85549	1.62	24.91	0.2742	1	0.42	24.74	3	0.04	4.77e+43	2.71e+44
60013	150.47938	2.79843	1.98	25.02	0.2202	4	0.35	23.36	23	-1.00	3.10e+43	-4.00e+44
00350	150.30865	2.79669	1.67	23.79	0.2784	1	0.69	24.87	2	-1.00	4.72e+43	-4.22e+44
10168	150.09462	2.83234	1.56	23.17	0.0288	4	0.90	23.01	12	-1.00	3.51e+43	-2.13e+44
05541	149.53005	2.22210	2.36	24.30	0.2769	4	0.79	22.84	3	-1.00	1.55e+44	-7.02e+44
02637	150.61225	1.99446	1.61	24.45	0.2586	1	0.44	21.68	23	-0.25	4.40e+43	1.25e+44
60168	150.70284	2.06962	2.14	23.54	0.3783	1	0.68	24.90	1	-1.00	4.57e+43	-2.15e+44
02081	150.66095	1.97540	1.57	22.50	0.4347	1	0.48	21.57	3	-0.44	3.38e+44	6.20e+44
02421	149.52908	2.38016	3.10	23.23	0.1677	1	0.66	21.22	23	-0.41	4.24e+44	8.94e+44
02792	149.67859	2.34891	2.00	25.21	0.2727	1	0.60	23.42	2	-0.36	6.68e+43	1.58e+44
05016	149.75868	2.47702	1.75	24.39	0.1785	1	0.63	25.77	3	-1.00	2.95e+43	-1.29e+44
60016	150.58772	2.77316	2.05	25.02	0.1089	1	0.69	23.41	2	0.27	-5.67e+43	1.97e+44
02650	149.59815	2.68506	2.42	24.02	0.2112	4	0.89	24.02	3	-0.38	4.40e+44	9.79e+44
05042	149.80459	1.75815	2.61	24.42	0.2268	4	0.45	25.68	23	1	-8.38e+43	8.91e+44
60116	150.29774	2.67353	1.51	23.72	0.2664	1	0.66	22.35	2	1	-3.05e+43	5.82e+43
54333	149.47107	1.68743	2.14	24.44	0.1173	1	0.31	21.57	2	-0.44	1.40e+44	-2.89e+44
05517	150.70554	2.62963	2.13	22.28	0.0912	1	0.63	20.37	3	-0.42	1.98e+44	3.96e+44
00107	150.23647	2.81367	1.53	25.02	0.213	4	0.73	22.45	3	-0.44	1.67e+44	3.34e+44
00295	150.44691	2.79149	1.53	23.39	0.2217	1	0.75	22.81	12	-0.41	4.05e+43	-3.53e+44
60368	149.97389	2.81035	1.55	23.50	0.5439	1	0.71	23.65	12	-1.00	1.55e+43	-5.01e+43
05162	149.75539	2.73856	3.52	25.05	0.318	1	0.54	23.50	2	-0.11	2.12e+44	8.62e+44
02072	150.62575	1.76547	1.59	23.62	0.1122	1	0.82	21.23	12	-0.40	1.60e+44	3.36e+44
02072	150.62575	1.76547	1.59	23.15	0.1023	4	0.66	21.29	23	-0.40	1.60e+44	3.36e+44
05006	149.74178	2.58519	2.42	25.06	0.1980	4	0.46	23.95	2	0.32	3.28e+43	3.26e+44
05022	150.62971	2.47643	2.28	25.54	0.0408	4	0.17	24.76	2	-0.42	1.76e+44	3.58e+44
05543	150.06493	2.77998	1.59	23.69	0.186	4	0.73	24.47	3	-1.00	3.28e+43	-2.13e+44
05057	150.29501	2.76463	1.69	24.14	0.3231	1	0.71	24.88	2	0.19	2.34e+43	1.70e+44
02554	149.63529	2.59884	2.55	24.39	0.2742	1	0.66	23.06	23	-0.41	2.18e+44	4.56e+44
02421	149.52908	2.38016	3.10	24.04	0.2355	1	0.14	21.12	2	-0.41	4.24e+44	8.94e+44
02276	150.53563	1.76490	2.20	23.80	0.2256	1	0.80	21.30	12	-0.51	1.34e+44	2.20e+44
54137	149.93392	2.84381	1.99	23.79	0.1701	1	0.89	21.24	2	-1.00	5.44e+43	-4.08e+44
05514	150.53844	2.78356	2.87	25.38	0.1146	4	0.53	23.74	2	-1.00	9.25e+43	-1.43e+45
02028	150.54697	1.61851	1.59	23.96	0.2706	4	0.76	23.00	2	-0.37	5.35e+44	1.21e+45
00471	150.19195	2.82916	1.57	25.09	0.2946	1	0.68	24.17	3	-0.35	6.49e+43	1.55e+44
54137	149.93392	2.84381	1.99	23.03	0.1542	1	0.81	21.38	2	-1.00	5.44e+43	-4.08e+44
00395	150.29918	2.63348	2.98	25.75	0.1815	1	0.32	23.51	1	-0.35	1.49e+44	3.63e+44
05315	149.57190	2.55450	1.57	23.52	0.2472	1	0.91	23.96	3	-1.00	1.69e+43	-9.09e+43
02081	150.66095	1.97540	1.57	22.55	0.4389	1	0.52	21.53	3	-0.44	3.38e+44	6.20e+44
60187	150.68168	1.71636	1.85	23.78	0.1497	1	0.80	24.90	3	-1.00	2.24e+43	-1.25e+42
05205	150.49132	1.79373	2.028	25.23	0.2835	1	0.66	24.12	4	0.02	4.91e+43	2.61e+44
02216	149.69876	2.44124	1.52	23.16	0.4218	1	0.68	21.53	12	-0.55	1.32e+44	1.93e+44
00220	150.33852	2.77680	2.14	24.39	0.4812	1	0.35	22.73	12	-0.31	2.06e+44	5.20e+44
00448	149.87442	2.03104	1.85	23.73	0.3054	1	0.70	24.52	3	-0.18	5.96e+43	2.13e+44
05598	149.53706	2.28017	1.70	23.94	0.3507	1	0.61	24.27	3	-1.00	4.18e+43	-1.45e+44
10332	150.49210	1.69903	2.22	24.68	0.1248	1	0.62	24.35	23	-1.00	6.16e+43	-2.76e+42
05434	149.50602	1.80919	2.40	23.89	0.3138	1	0.82	22.20	12	-0.49	3.30e+44	5.69e+44
05387	150.66891	2.51676	1.57	24.96	0.1644	1	0.81	22.32	3	-0.36	3.72e+43	8.86e+43
05464	149.57445	2.08508	1.62	24.13	0.5844	4	0.82	22.71	3	-1.00	4.63e+43	-2.18e+44
00453	150.15880	2.80851	2.83	25.11	0.1368	1	0.96	23.32	12	-0.55	3.47e+44	4.90e+44
00050	149.89196	2.28515	2.94	25.11	0.1650	1	0.41	27.08	23	-0.35	6.68e+44	1.65e+45
00133	150.27214	2.23009	2.62	25.29	0.0483	1	0.99	26.24	3	-0.37	4.13e+44	9.22e+44
00144	149.98158	2.31501	2.93	24.84	0.1974	1	0.52	26.78	3	-0.34	2.77e+44	6.97e+44
00145	150.01985	2.34914	1.80	24.91	0.2217	1	0.83	25.69	3	0.02	7.05e+43	3.80e+44
00181	150.26748	2.05561	1.72	25.06	0.2583	4	0.30	26.96	12	-0.43	1.14e+44	2.29e+44

XMM ID	RA (J2000)	Dec (J2000)	z	m_{host}, AB	R_e, arc-sec	n	b/a	m_{PSF}	Flag[1]	HR	$L_{0.5-2keV}$	$L_{2-10keV}$
00182	150.32085	2.33296	2.04	26.03	0.1458	1	0.41	25.59	3	-0.17	8.52e+43	3.01e+44
00233	150.17986	2.62140	1.93	24.94	0.216	1	0.46	25.75	2	-0.22	7.80e+43	2.53e+44
00246	150.05100	2.49385	2.30	24.14	0.2334	1	0.49	26.23	2	0.42	3.48e+43	4.30e+44
00248	150.19607	2.11927	1.88	25.29	0.3063	4	0.62	25.56	1	0.02	3.96e+43	2.11e+44
00279	150.15270	2.06784	1.65	23.87	0.339	1	0.85	26.61	1	-0.27	3.20e+43	9.47e+43
00327	150.19025	2.00033	1.91	24.19	0.3795	1	0.79	26.07	12	0.12	2.63e+43	1.67e+44
00345	150.17022	2.12144	1.81	26.12	0.1452	1	0.35	24.88	23	-0.26	5.56e+43	1.67e+44
00406	150.38741	2.59158	2.92	25.08	0.1347	4	0.72	26.75	3	-1.00	9.16e+43	-1.87e+45
00431	149.91850	2.15369	1.77	25.29	0.387	1	0.47	25.29	0	-1.00	2.87e+43	-1.73e+44
00449	150.26402	1.94734	2.13	26.37	0.1278	1	0.47	25.08	23	-1.00	8.39e+43	-2.27e+44
01666	150.19258	1.81554	1.92	25.81	0.1713	1	0.61	26.03	3	-1.00	2.59e+43	-1.44e+44
02382	150.46054	1.97675	1.67	24.86	0.2901	1	0.40	26.01	1	-0.39	8.01e+43	1.76e+44
02543	150.34372	2.14067	1.76	25.01	0.1716	1	0.93	26.64	2	-0.40	6.81e+43	1.45e+44
05178	149.82381	2.38543	1.72	23.99	0.5154	1	0.30	26.31	3	0.20	1.67e+43	1.27e+44
05187	149.98928	1.89717	2.34	24.88	0.0792	1	0.58	28.16	3	-1.00	2.05e+43	-7.95e+43
05259	150.46710	2.53200	4.45	26.00	0.033	1	0.41	27.32	3	-1.00	2.08e+44	-3.40e+45
05285	150.58005	2.47977	2.59	25.62	0.1269	1	0.42	26.08	2	-0.26	1.25e+44	3.61e+44
05347	149.66964	2.16770	3.12	26.34	0.1686	1	0.57	26.31	3	-1.00	1.69e+44	-9.08e+44
05549	149.68386	1.93303	1.66	25.62	0.1782	1	0.35	25.95	3	-0.32	1.23e+44	3.11e+44
05622	150.30872	1.89147	1.81	26.04	0.1104	1	0.50	26.43	3	0.13	3.85e+43	2.53e+44
10590	150.41410	2.10921	1.76	25.25	0.1806	1	0.81	26.08	2	-1.00	2.83e+43	-1.78e+44
30920	149.80634	2.14546	3.00	24.05	0.1044	1	0.71	26.08	2	-1.00	7.18e+43	-4.14e+44
53359	150.41193	2.31759	2.89	24.57	0.5265	1	0.61	26.24	2	-1.00	7.21e+43	-9.27e+44
54039	150.03642	2.21403	3.26	24.78	0.3867	1	0.48	26.33	12	-0.10	7.96e+43	-6.55e+44
60146	150.46826	1.74272	2.08	25.13	0.3282	1	0.61	25.91	3	0.11	2.68e+43	-2.78e+44
60235	150.29116	2.08930	1.78	23.51	0.4224	1	0.65	24.86	1	1	-1.04e+43	1.40e+44
60421	149.98010	1.92490	1.60	25.73	0.1302	1	0.42	25.77	3	-1.00	1.64e+43	-1.28e+44
60464	150.02771	2.12674	2.35	25.41	0.2358	1	0.23	26.25	2	1	-1.47e+43	2.20e+44
70135	150.00254	2.25863	2.00	23.78	0.1767	1	0.44	26.68	3	1	-2.54e+44	4.35e+44
70210	150.09497	1.97847	1.93	25.40	0.1986	1	0.88	25.60	2	1	-4.32e+44	3.09e+44
00187	150.24078	2.65906	3.36	25.64	0.1593	1	0.36	22.55	3	-0.38	3.72e+44	8.53e+44
00192	150.25131	2.73716	2.17	23.22	0.1143	4	0.54	20.94	2	-0.57	1.81e+44	2.47e+44
00275	150.15896	2.82514	1.86	23.98	0.3144	1	0.88	20.76	2	-0.52	2.55e+44	3.97e+44
00276	150.51449	2.81023	2.62	25.33	0.0921	1	0.13	21.16	23	-0.31	3.71e+44	9.92e+44
02020	150.64304	1.83646	1.51	21.92	0.1347	1	0.95	18.87	2	-0.62	3.48e+44	4.14e+44
02046	150.76222	1.86906	1.80	22.73	0.0897	1	0.39	19.97	2	-0.47	5.12e+44	9.37e+44
70225	149.61098	1.74480	1.910	23.92	0.3039	1	0.47	26.56	2	1	-6.32e+43	1.52e+44
02096	150.55905	2.10576	2.32	24.56	0.1479	1	0.45	22.32	3	-0.45	6.35e+44	1.23e+45
02200	149.60204	2.39269	1.86	22.82	0.1827	1	0.93	19.92	3	-0.53	2.18e+44	3.35e+44
02202	150.65295	1.99689	1.52	23.47	0.1116	4	0.18	19.87	2	-0.50	6.55e+43	1.07e+44
02209	150.31994	1.68160	2.06	20.54	0.03	4	0.51	20.00	12	-0.47	3.30e+44	5.79e+44
02299	150.37656	1.71788	1.57	24.08	0.1653	1	0.51	20.16	2	-0.30	5.94e+43	1.60e+44
02333	149.99158	1.72427	1.62	22.82	0.1656	1	0.75	21.12	3	-0.46	1.15e+44	2.10e+44
02336	150.60283	1.76875	1.76	25.11	0.0816	1	0.91	24.84	3	0.19	3.82e+43	2.77e+44
02344	150.73287	1.62514	2.39	21.12	0.0171	4	0.60	21.43	12	-0.58	2.28e+44	3.04e+44
02396	149.47792	2.64247	1.60	22.61	0.2208	1	0.98	19.33	12	-0.34	4.59e+43	-1.48e+44
05031	150.53669	1.84958	1.82	24.01	0.1158	1	0.89	21.82	2	-1.00	3.90e+43	-1.22e+44
05116	150.73556	2.19958	3.50	22.10	0.1035	1	0.58	20.48	12	-0.43	4.51e+44	9.06e+44
05175	150.62007	2.67141	3.14	25.91	0.1203	1	0.31	22.30	23	-0.44	5.57e+44	1.08e+45
05273	150.68208	2.78552	2.50	22.56	0.1335	1	0.42	20.09	2	-0.49	4.13e+44	6.87e+44
05290	150.10444	2.69128	1.88	24.12	0.1611	1	0.86	21.13	2	-0.38	1.43e+44	3.10e+44
05293	149.57569	2.76317	2.14	24.94	0.1884	1	0.32	22.81	2	-0.36	1.04e+44	2.33e+44
05305	149.70126	2.55577	2.50	26.98	0.0588	1	0.99	24.42	4	-1.00	4.98e+43	-3.43e+44
05331	150.60880	2.76966	3.04	23.59	0.1140	1	0.16	20.19	2	-0.59	3.55e+44	4.56e+44
05338	149.53407	1.90657	2.51	25.01	0.3669	1	0.45	22.74	23	-0.33	2.71e+44	6.66e+44
05431	150.58113	2.22106	2.03	24.97	0.1797	1	0.59	22.24	2	-0.37	8.59e+43	1.90e+44
05478	150.39676	1.73570	2.77	25.61	0.1959	1	0.28	22.92	2	-1.00	7.97e+43	-3.47e+44
05517	150.70554	2.62963	2.13	22.95	0.1392	1	0.21	20.18	3	-0.42	1.98e+44	3.97e+44
05575	149.81103	1.60974	2.41	25.59	0.1575	1	0.32	22.32	2	-1.00	2.68e+44	-4.13e+44
05625	149.54674	2.76221	2.75	25.14	0.2559	1	0.66	22.32	3	-0.56	2.69e+44	3.79e+44
30366	150.40506	2.27014	1.66	23.78	0.2409	1	0.71	20.59	2	-1.00	2.98e+43	-1.92e+44
53357	150.72068	2.69366	2.64	24.13	0.1257	1	0.13	20.19	2	-1.00	9.38e+43	-5.74e+44
60466	149.57348	2.02369	2.81	24.07	0.1077	1	0.12	21.30	2	-1.00	9.43e+43	-5.48e+44
60466	149.57348	2.02369	2.81	23.03	0.1065	1	0.56	21.61	2	-1.00	9.43e+43	-5.48e+44
00094	150.19975	2.19087	1.510	23.37	0.4677	1	0.90	21.21	1	-0.43	1.06e+44	2.14e+44

XMM ID	RA (J2000)	Dec (J2000)	z	m_{host}, AB	R_e, arc-sec	n	b/a	m_{PSF}	Flag[1]	HR	$L_{0.5-2keV}$	$L_{2-10keV}$
00197	150.16378	2.59768	1.589	24.54	0.2724	1	0.37	22.23	12	-0.57	6.14e+43	8.60e+43
00122	149.93893	2.17489	2.418	24.13	0.2148	1	0.67	25.66	3	0.37	8.30e+43	9.10e+44
00280	150.47223	2.32464	1.520	23.18	0.5121	1	0.52	25.94	3	1	-2.89e+43	1.67e+44
00304	149.88157	2.31820	1.607	22.82	0.3777	1	0.44	25.42	1	-0.29	2.59e+43	7.73e+43
00319	149.98451	2.23575	1.639	24.17	0.2931	1	0.72	25.19	2	1	-8.33e+42	1.12e+44
00435	150.32760	2.60464	2.080	23.93	0.1851	1	0.94	24.92	1	-1.00	4.08e+43	-9.70e+44
00462	149.73380	2.07635	2.047	24.82	0.1797	1	0.66	22.61	3	-1.00	7.31e+43	-1.55e+44
00504	149.87920	2.22584	3.651	24.59	0.1764	1	0.31	21.72	3	-1.00	1.78e+44	-1.55e+44
05013	149.77640	1.94016	1.820	24.19	0.1824	1	0.57	22.12	3	-0.28	7.78e+43	2.17e+44
05518	149.71558	2.01655	2.686	25.01	0.1236	1	0.23	22.43	3	-1.00	1.72e+44	-4.62e+58
53700	150.43993	1.85595	1.600	23.93	0.4362	1	0.66	24.03	12	-1.00	2.32e+43	-8.60e+43
60131	150.29725	2.14885	3.328	22.23	0.0975	1	0.78	20.79	3	0.10	1.05e+44	6.26e+44
70007	150.34768	2.39103	1.848	23.33	0.2304	1	0.81	22.52	2	1	-1.03e+44	2.03e+44
31357	150.39571	2.63535	1.700	23.62	0.4602	1	0.26	25.44	3	0.61	1.90e+43	3.92e+44
02154	149.64606	2.38804	1.845	22.85	0.1269	4	0.70	24.10	2	0.28	1.45e+44	1.28e+45
02530	149.60492	2.58660	2.640	22.22	0.459	1	0.52	22.55	3	-0.46	2.99e+44	5.51e+44
02725	149.93170	1.83014	2.676	25.67	0.0471	4	0.8	25.10	2	-0.28	1.40e+44	-4.19e+44
05127	150.67052	2.58011	1.679	22.86	0.2346	1	0.42	23.45	12	0.03	5.96e+43	3.20e+44
05533	149.59417	2.61408	2.050	24.09	0.294	1	0.66	25.87	12	-0.21	1.04e+44	3.43e+44
05573	149.52978	1.78571	1.660	23.64	0.6651	4	0.77	24.53	1	-0.19	1.27e+44	4.34e+44
31048	149.61400	2.83479	1.500	23.94	0.2259	1	0.26	24.81	3	-1.00	2.20e+43	-1.06e+44
53364	150.63166	2.61493	2.210	24.22	0.2355	1	0.56	26.69	12	-1.00	5.34e+43	-2.40e+44
70017	150.23654	2.64496	2.060	23.93	0.3318	1	0.65	26.13	2	1	-7.55e+43	1.72e+44

Table 5: Catalog of the resolved galaxies. [1] This flag specifies the disturbance class of a galaxy: 1 marks objects classified as MM, 12 - minor accretion events, 2 - lopsidedness, 23 - slightly misaligned components, 3 - undisturbed hosts

References

Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539

Alonso-Herrero, A., Pérez-González, P. G., Rieke, G. H., et al. 2008, ApJ, 677, 127

Anderson, J., Bedin, L. R., Piotto, G., Yadav, R. S., & Bellini, A. 2006, A&A, 454, 1029

Antonucci, R. 1993, ARA&A, 31, 473

Antonucci, R. R. J. & Miller, J. S. 1985, ApJ, 297, 621

Aretxaga, I., Cid Fernandes, R., & Terlevich, R. J. 1997, MNRAS, 286, 271

Arnouts, S. & Ilbert, O. 2011, Astrophysics Source Code Library, 8009

Arnouts, S., Vandame, B., Benoist, C., et al. 2001, A&A, 379, 740

Bahcall, J. N., Kirhakos, S., Saxe, D. H., & Schneider, D. P. 1997, ApJ, 479, 642

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Ballantyne, D. R., Everett, J. E., & Murray, N. 2006, ApJ, 639, 740

Barger, A. J., Cowie, L. L., Capak, P., et al. 2003, AJ, 126, 632

Barger, A. J., Cowie, L. L., Mushotzky, R. F., et al. 2005, AJ, 129, 578

Barnes, J. E. 2004, MNRAS, 350, 798

Barnes, J. E. & Hernquist, L. 1992, ARA&A, 30, 705

Barnes, J. E. & Hernquist, L. 1996, ApJ, 471, 115

Barnes, J. E. & Hernquist, L. E. 1991, ApJ, 370, L65

Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752

Bell, E. F., Zheng, X. Z., Papovich, C., et al. 2007, ApJ, 663, 834

Bertin, E. 2008, SWarp v2.17 User's guide

Bertin, E. 2010, SCAMP v1.9 User's guide

Bertin, E. & Arnouts, S. 1996, Astronomy and Astrophysics Supplement, 117, 393

Bolton, J. G., Gardner, F. F., & Mackey, M. B. 1963, Nature, 199, 682

Boutsia, K., Leibundgut, B., Trevese, D., & Vagnetti, F. 2009, A&A, 497, 81

Bouwens, R. & Silk, J. 2002, ApJ, 568, 522

Bramich, D. M., Vidrih, S., Wyrzykowski, L., et al. 2008, MNRAS, 386, 887

Brown, M. L., Ade, P., Bock, J., et al. 2009, ApJ, 705, 978

Brusa, M., Civano, F., Comastri, A., et al. 2010, ApJ, 716, 348

Brusa, M., Zamorani, G., Comastri, A., et al. 2007, ApJS, 172, 353

Bundy, K., Scarlata, C., Carollo, C. M., et al. 2010, ApJ, 719, 1969

Burles, S. & Tytler, D. 1998, ApJ, 499, 699

Caldwell, J. A. R., McIntosh, D. H., Rix, H.-W., et al. 2008, ApJS, 174, 136

Canalizo, G. & Stockton, A. 2001, ApJ, 555, 719

Cano-Díaz, M., Maiolino, R., Marconi, A., et al. 2012, A&A, 537, L8

Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99

Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, A&A, 497, 635

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Cassata, P., Cimatti, A., Franceschini, A., et al. 2005, MNRAS, 357, 903

Cattaneo, A., Blaizot, J., Devriendt, J., & Guiderdoni, B. 2005, MNRAS, 364, 407

Cisternas, M., Jahnke, K., Inskip, K. J., et al. 2011, ApJ, 726, 57

Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168

Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393

Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183

Cox, T. J., Jonsson, P., Primack, J. R., & Somerville, R. S. 2006, MNRAS, 373, 1013

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, MNRAS, 384, 386

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 399, 1755

Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, MNRAS, 349, 1397

Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 600, L127

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

Doménech-Moral, M., Martínez-Serrano, F. J., Domínguez-Tenreiro, R., & Serna, A. 2012, MNRAS, 421, 2510

Dunlop, J. S., McLure, R. J., Kukula, M. J., et al. 2003, MNRAS, 340, 1095

Ebrero, J., Carrera, F. J., Page, M. J., et al. 2009, A&A, 493, 55

Elvis, M., Civano, F., Vignali, C., et al. 2009, ApJS, 184, 158

Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1

Erben, T., Schirmer, M., Dietrich, J. P., et al. 2005, Astronomische Nachrichten, 326, 432

Fabian, A. C. 2012, ARA&A, 50, 455

Fath, E. A. 1909, Lick Observatory Bulletin, 5, 71

Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, ApJ, 600, L107

Ferrarese, L. & Ford, H. 2005, Space Sci. Rev., 116, 523

Ferrarese, L. & Merritt, D. 2000, ApJ, 539, L9

Finoguenov, A., Guzzo, L., Hasinger, G., et al. 2007, ApJS, 172, 182

Fontanot, F., Monaco, P., Cristiani, S., & Tozzi, P. 2006, MNRAS, 373, 1173

Fruchter, A. & Hook, R. N. 1997, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 3164, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. A. G. Tescher, 120–125

Fruchter, A. S. & Hook, R. N. 2002, PASP, 114, 144

Gabor, J. M., Impey, C. D., Jahnke, K., et al. 2009, ApJ, 691, 705

Gavignaud, I., Bongiorno, A., Paltani, S., et al. 2006, A&A, 457, 79

Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13

Gehren, T., Fried, J., Wehinger, P. A., & Wyckoff, S. 1984, ApJ, 278, 11

Giacconi, R., Zirm, A., Wang, J., et al. 2002, ApJS, 139, 369

Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93

Giavalisco, M., Steidel, C. C., & Macchetto, F. D. 1996, ApJ, 470, 189

Giveon, U., Maoz, D., Kaspi, S., Netzer, H., & Smith, P. S. 1999, MNRAS, 306, 637

Glikman, E., Lacy, M., & Urrutia, T. 2012, in American Astronomical Society Meeting Abstracts, Vol. 220, American Astronomical Society Meeting Abstracts 220, 420.03

Graham, A. W., Driver, S. P., Petrosian, V., et al. 2005, AJ, 130, 1535

Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580

Griffith, R. L. & Stern, D. 2010, AJ, 140, 533

Grogin, N. A., Conselice, C. J., Chatzichristou, E., et al. 2005, ApJ, 627, L97

Grupe, D., Thomas, H.-C., & Beuermann, K. 2001, A&A, 367, 470

Gunn, J. E. 1979, Feeding the monster - Gas discs in elliptical galaxies, ed. C. Hazard & S. Mitton, 213–225

Haiman, Z., Ciotti, L., & Ostriker, J. P. 2004, ApJ, 606, 763

Häring, N. & Rix, H.-W. 2004, ApJ, 604, L89

Hasinger, G. 2008, A&A, 490, 905

Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, ApJS, 172, 29

Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417

Hatziminaoglou, E., Omont, A., Stevens, J. A., et al. 2010, A&A, 518, L33

Häussler, B., McIntosh, D. H., Barden, M., et al. 2007, ApJS, 172, 615

Hawkins, M. R. S. 1996, MNRAS, 278, 787

Hazard, C., Mackey, M. B., & Shimmins, A. J. 1963, Nature, 197, 1037

Heckman, T. M. 1980, A&A, 87, 152

Heckman, T. M., Bothun, G. D., Balick, B., & Smith, E. P. 1984, AJ, 89, 958

Hernquist, L. 1989, Nature, 340, 687

Hershey, J. L. & Taff, L. G. 1998, AJ, 116, 1440

Heymans, C., Brown, M. L., Barden, M., et al. 2005, MNRAS, 361, 160

Hicken, M., Challis, P., Jha, S., et al. 2009, ApJ, 700, 331

Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, ApJS, 180, 225

Ho, L. C. 2008, ARA&A, 46, 475

Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1996, ApJ, 462, 183

Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJS, 112, 315

Hopkins, P. F., Bundy, K., Croton, D., et al. 2010, ApJ, 715, 202

Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009, ApJ, 691, 1168

Hopkins, P. F. & Hernquist, L. 2006, ApJS, 166, 1

Hopkins, P. F. & Hernquist, L. 2009, ApJ, 694, 599

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356

Hopkins, P. F., Hernquist, L., Cox, T. J., Robertson, B., & Krause, E. 2007, ApJ, 669, 45

Hopkins, P. F., Hernquist, L., Martini, P., et al. 2005, ApJ, 625, L71

Hovatta, T., Nieppola, E., Tornikoski, M., et al. 2008, A&A, 485, 51

Hutchings, J. B. & Bianchi, L. 2010, AJ, 140, 1987

Hutchings, J. B., Crampton, D., & Campbell, B. 1984, ApJ, 280, 41

Hutchings, J. B., Johnson, I., & Pyke, R. 1988, ApJS, 66, 361

Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236

Jahnke, K., Sánchez, S. F., Wisotzki, L., et al. 2004, ApJ, 614, 568

Jogee, S. 2006, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 693, Physics of Active Galactic Nuclei at all Scales, ed. D. Alloin, 143

Kauffmann, G. & Haehnelt, M. 2000, MNRAS, 311, 576

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kawaguchi, T., Mineshige, S., Umemura, M., & Turner, E. L. 1998, ApJ, 504, 671

Kawata, D. & Gibson, B. K. 2003, MNRAS, 346, 135

Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189

Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32

Kobayashi, C. 2004, MNRAS, 347, 740

Kocevski, D. D., Faber, S. M., Mozena, M., et al. 2012, ApJ, 744, 148

Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kormendy, J. & Kennicutt, Jr., R. C. 2004, ARA&A, 42, 603

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

Kotilainen, J. K., Falomo, R., Decarli, R., et al. 2009, ApJ, 703, 1663

Kron, R. G. 1980, ApJS, 43, 305

La Franca, F., Fiore, F., Comastri, A., et al. 2005, ApJ, 635, 864

Lallo, M. D., Makidon, R. B., Casertano, S., & Krist, J. E. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6270, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222

Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, ApJS, 172, 219

Lehmer, B. D., Brandt, W. N., Alexander, D. M., et al. 2005, ApJS, 161, 21

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010, MNRAS, 404, 575

Lowenthal, J. D., Koo, D. C., Guzman, R., et al. 1997, ApJ, 481, 673

Lynden-Bell, D. 1969, Nature, 223, 690

Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106

Maddox, N., Hewett, P. C., Péroux, C., Nestor, D. B., & Wisotzki, L. 2012, MNRAS, 424, 2876

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Makidon, R. B., Lallo, M. D., Casertano, S., et al. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6270, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Mao, S., Mo, H. J., & White, S. D. M. 1998, MNRAS, 297, L71

Maoz, D. 2007, MNRAS, 377, 1696

Marconi, A. & Hunt, L. K. 2003, ApJ, 589, L21

Martini, P. 2004, in IAU Symposium, Vol. 222, The Interplay Among Black Holes, Stars and ISM in Galactic Nuclei, ed. T. Storchi-Bergmann, L. C. Ho, & H. R. Schmitt, 235–241

Massey, R., Stoughton, C., Leauthaud, A., et al. 2010, MNRAS, 401, 371

Matthews, T. A. & Sandage, A. R. 1963, ApJ, 138, 30

McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202

Merloni, A., Rudnick, G., & Di Matteo, T. 2004, MNRAS, 354, L37

Meurer, G. R., Lindler, D. J., Blakeslee, J., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4854, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J. C. Blades & O. H. W. Siegmund, 507–514

Meza, A., Navarro, J. F., Steinmetz, M., & Eke, V. R. 2003, ApJ, 590, 619

Mihos, J. C. & Hernquist, L. 1996, ApJ, 464, 641

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Murray, N., Quataert, E., & Thompson, T. A. 2005, ApJ, 618, 569

Mushotzky, R. 2004, in Astrophysics and Space Science Library, Vol. 308, Supermassive Black Holes in the Distant Universe, ed. A. J. Barger, 53

Naab, T., Khochfar, S., & Burkert, A. 2006, ApJ, 636, L81

Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, A&A, 530, A122

Papovich, C., Dickinson, M., Giavalisco, M., Conselice, C. J., & Ferguson, H. C. 2005, ApJ, 631, 101

Papovich, C., Giavalisco, M., Dickinson, M., Conselice, C. J., & Ferguson, H. C. 2003, ApJ, 598, 827

Peacock, J. A. 2003, Royal Society of London Philosophical Transactions Series A, 361, 2479

Peng, C. Y. 2007, ApJ, 671, 1098

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097

Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, MNRAS, 401, 2148

Pereyra, N. A., Vanden Berk, D. E., Turnshek, D. A., et al. 2006, ApJ, 642, 87

Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei

Ptak, A., Yaqoob, T., Mushotzky, R., Serlemitsos, P., & Griffiths, R. 1998, ApJ, 501, L37

Rees, M. J. 1984, ARA&A, 22, 471

Reichardt, C. L., Ade, P. A. R., Bock, J. J., et al. 2009, ApJ, 694, 1200

Reid, B. A., Percival, W. J., Eisenstein, D. J., et al. 2010, MNRAS, 404, 60

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

Rhodes, J. D., Massey, R., Albert, J., et al. 2006, in The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode, ed. A. M. Koekemoer, P. Goudfrooij, & L. L. Dressel, 21

Rhodes, J. D., Massey, R. J., Albert, J., et al. 2007, ApJS, 172, 203

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

Richstone, D., Ajhar, E. A., Bender, R., et al. 1998, Nature, 395, A14

Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, ApJS, 152, 163

Robaina, A. R., Bell, E. F., van der Wel, A., et al. 2010, ApJ, 719, 844

Salpeter, E. E. 1955, ApJ, 121, 161

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

Sánchez, S. F., Jahnke, K., Wisotzki, L., et al. 2004, ApJ, 614, 586

Sandage, A. 1973, ApJ, 183, 711

Sanders, D. B. & Mirabel, I. F. 1996, ARA&A, 34, 749

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988a, ApJ, 325, 74

Sanders, D. B., Soifer, B. T., Elias, J. H., Neugebauer, G., & Matthews, K. 1988b, ApJ, 328, L35

Santini, P., Rosario, D. J., Shao, L., et al. 2012, A&A, 540, A109

Sarajedini, V. L., Koo, D. C., Klesman, A. J., et al. 2011, ApJ, 731, 97

Schawinski, K., Simmons, B. D., Urry, C. M., Treister, E., & Glikman, E. 2012, MNRAS, 425, L61

Schawinski, K., Treister, E., Urry, C. M., et al. 2011, ApJ, 727, L31

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

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

Schmidt, M. 1963, Nature, 197, 1040

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

Schrabback, T., Erben, T., Simon, P., et al. 2007, A&A, 468, 823

Schrabback, T., Hartlap, J., Joachimi, B., et al. 2010, A&A, 516, A63

Schramm, M., Wisotzki, L., & Jahnke, K. 2008, A&A, 478, 311

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

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

Seyfert, C. K. 1943, ApJ, 97, 28

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

Shields, G. A., Menezes, K. L., Massart, C. A., & Vanden Bout, P. 2006, ApJ, 641, 683

Silk, J. & Rees, M. J. 1998, A&A, 331, L1

Silverman, J. D., Green, P. J., Barkhouse, W. A., et al. 2008, ApJ, 679, 118

Silverman, J. D., Kampczyk, P., Jahnke, K., et al. 2011, ApJ, 743, 2

Silverman, J. D., Lamareille, F., Maier, C., et al. 2009, ApJ, 696, 396

Soltan, A. 1982, MNRAS, 200, 115

Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, MNRAS, 320, 504

Soria, R., Fabbiano, G., Graham, A. W., et al. 2006a, ApJ, 640, 126

Soria, R., Graham, A. W., Fabbiano, G., et al. 2006b, ApJ, 640, 143

Springel, V. 2000, MNRAS, 312, 859

Springel, V. & Hernquist, L. 2005, ApJ, 622, L9

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stalin, C. S., Gopal-Krishna, Sagar, R., & Wiita, P. J. 2004, MNRAS, 350, 175

Stauffer, J. R. 1982, ApJ, 262, 66

Stockton, A. 1982, ApJ, 257, 33

Surace, J. A. & Sanders, D. B. 1999, ApJ, 512, 162

Surace, J. A., Sanders, D. B., Vacca, W. D., Veilleux, S., & Mazzarella, J. M. 1998, ApJ, 492, 116

Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201

Terlevich, R., Tenorio-Tagle, G., Franco, J., & Melnick, J. 1992, MNRAS, 255, 713

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. G. Larson, D. Campbell, 401

Toomre, A. & Toomre, J. 1972, ApJ, 178, 623

Treister, E. & Urry, C. M. 2006, ApJ, 652, L79

Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, ApJ, 574, 740

Trevese, D., Boutsia, K., Vagnetti, F., Cappellaro, E., & Puccetti, S. 2008, A&A, 488, 73

Trevese, D., Kron, R. G., Majewski, S. R., Bershady, M. A., & Koo, D. C. 1994, ApJ, 433, 494

Trujillo, I. & Aguerri, J. A. L. 2004, MNRAS, 355, 82

Trujillo, I., Förster Schreiber, N. M., Rudnick, G., et al. 2006, ApJ, 650, 18

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886

Urrutia, T., Lacy, M., & Becker, R. H. 2008, ApJ, 674, 80

Urry, C. M. & Padovani, P. 1995, PASP, 107, 803

van den Bergh, S. 2001, AJ, 122, 621

van den Bergh, S., Abraham, R. G., Whyte, L. F., et al. 2002, AJ, 123, 2913

Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, ApJ, 601, 692

Wang, J. X., Malhotra, S., Rhoads, J. E., & Norman, C. A. 2004, ApJ, 612, L109

White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341

Williams, R. E., Baum, S., Bergeron, L. E., et al. 2000, AJ, 120, 2735

Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335

Windhorst, R. A., Taylor, V. A., Jansen, R. A., et al. 2002, ApJS, 143, 113

Wold, M., Brotherton, M. S., & Shang, Z. 2007, MNRAS, 375, 989

Wolf, C., Meisenheimer, K., Rix, H.-W., et al. 2003, A&A, 401, 73

Woltjer, L. 1959, ApJ, 130, 38

Wright, E. L. 2006, PASP, 118, 1711

Wyithe, J. S. B. & Loeb, A. 2003, ApJ, 595, 614

Yencho, B., Barger, A. J., Trouille, L., & Winter, L. M. 2009, ApJ, 698, 380

Yu, Q. & Tremaine, S. 2002, MNRAS, 335, 965

Yuan, F. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 373, The Central Engine of Active Galactic Nuclei, ed. L. C. Ho & J.-W. Wang, 95

Zackrisson, E., Bergvall, N., Marquart, T., & Helbig, P. 2003, A&A, 408, 17

Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2006, AJ, 132, 1496

Zel'dovich, Y. B. & Novikov, I. 1964, Sov. Phys. Dokl., 158, 811

Zheng, X. Z., Bell, E. F., Somerville, R. S., et al. 2009, ApJ, 707, 1566