
ALMA MATER STUDIORUM
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

DIPARTIMENTO DI FISICA E ASTRONOMIA
Dottorato di Ricerca in Astronomia
CICLO XXVIII

BLACK HOLE AND GALAXY GROWTH OVER
COSMIC TIME: THE CHANDRA COSMOS LEGACY
SURVEY

Dottorando
STEFANO MARCHESI

Relatore
Prof. **CRISTIAN VIGNALI**

Co-relatori
Dr. **FRANCESCA CIVANO**
Dr. **ANDREA COMASTRI**

Coordinatore
PROF. LAURO MOSCARDINI

Esame finale anno 2015

Settore Concorsuale: 02/C1 - Astronomia, Astrofisica, Fisica della Terra e dei Pianeti
Settore Scientifico-Disciplinare: FIS/05 - Astronomia e Astrofisica

E prego Madre Terra con lo sguardo
al cielo: illumina la strada in questo
buco nero.

Elio e le Storie Tese, La Visione

I can see for miles.

The Who

Contents

Executive summary	1
1 X-ray surveys and their role in AGN selection and investigation	3
1.1 Active Galactic Nuclei taxonomy	3
1.2 Black hole and galaxy growth history	4
1.3 AGN emission mechanisms from radio to UV	8
1.4 AGN emission mechanisms in the X-ray	10
1.5 Advantages of the X-ray analysis to study AGN	12
1.6 X-ray surveys and the “wedding cake” strategy	14
1.6.1 X-ray surveys in the COSMOS field	15
1.7 AGN X-ray space density as a SMBH evolution estimator	18
1.7.1 Phenomenological models of AGN luminosity function	18
1.7.2 $z > 3$ AGN space density as a SMBH early growth estimator	21
2 The <i>Chandra</i> COSMOS Legacy survey: overview and point source catalog	25
2.1 Observations	26
2.2 Data processing	26
2.2.1 Astrometric corrections	27
2.2.2 Exposure maps and data mosaic creation	29
2.2.3 Background maps creation	30
2.3 Data Analysis: source detection and photometry	33
2.3.1 Point source catalog	35
2.3.2 Matching with XMM-COSMOS catalog	42
2.4 Sky coverage and survey sensitivity	44
2.5 Number counts	46
2.6 Summary and Conclusions	49
2.7 CID-42: a candidate recoiling SMBH in the <i>Chandra COSMOS-Legacy</i> survey	54
2.7.1 Data analysis	55
2.7.2 P-Cygni profile analysis	57
2.7.3 Potential causes of the absorption line disappearance	59
3 The <i>Chandra COSMOS-Legacy</i> optical/infrared counterparts	61
3.1 Identification datasets	61
3.2 X-ray source identification method	63
3.2.1 Method	63

3.2.2	Choice of the Likelihood Ratio threshold	66
3.3	X-ray source identification results	67
3.3.1	Identification rates	70
3.3.2	Counterparts with $2 < \text{SNR} < 5$	70
3.3.3	Solving the cases of ambiguous sources	71
3.3.4	Final results for optical counterparts	72
3.3.5	Sources in C-COSMOS with updated optical counterpart	73
3.3.6	Sources in XMM-COSMOS with updated optical counterpart	73
3.3.7	Radio-detected counterparts	74
3.4	X-ray to optical positional offset and optical properties versus identification class	74
3.5	Spectroscopic and photometric redshift distribution	75
3.5.1	Spectroscopic redshifts	75
3.5.2	Photometric redshifts	77
3.5.3	Redshift summary	77
3.5.4	High-redshift sample	80
3.5.5	Spectroscopic and photometric types	81
3.5.6	X-ray luminosity	82
3.6	X-ray, optical and infrared properties of <i>Chandra COSMOS-Legacy</i> sources	85
3.6.1	Redshift Evolution of hardness ratio	85
3.6.2	X-ray to Optical Flux Ratio	89
3.6.3	X/O-hard band luminosity relation	92
3.6.4	Luminosity dependence of the AGN obscured fraction	94
3.7	Conclusions	97
4	The $z \geq 3$ sample of the <i>Chandra</i> COSMOS Legacy survey	101
4.1	Introduction	101
4.2	Properties of the high-redshift AGN sample	101
4.2.1	Spectroscopic redshifts	102
4.2.2	Photometric redshifts	102
4.2.3	Summary	104
4.2.4	Optically unidentified sources	105
4.2.5	Optical properties	105
4.2.6	X-ray properties	109
4.3	0.5-2 keV AGN number counts	110
4.4	2-10 keV comoving space density	113
4.4.1	Obscured versus unobscured AGN	117
4.5	Comparison with merger models	121
4.5.1	Alternatives to mergers	122
4.6	Black hole mass estimate at $z > 3$	124
4.7	Discussion and conclusions	127
5	Discussion and conclusions	131
A	List of new <i>Chandra COSMOS-Legacy</i> observations	135
B	The <i>Chandra</i> COSMOS X-ray catalog	137

C The multiwavelength catalog of *Chandra* COSMOS Legacy sources 139

Executive summary

One of the most active but least known epochs in astrophysics is the period between re-ionization ($z \sim 8$), when the growth of structures becomes highly non-linear and the first stars form, and $z \sim 2$, where major virialization occurs and star formation (SF) and supermassive black hole (SMBH) accretion peaks. Most SMBHs – and definitely the most massive ones – had to grow during an active accretion phase, when they would be visible as an active galactic nucleus (AGN), implying that most galaxies had an AGN phase in their past (Soltan 1982).

X-ray data play an important role in the selection of AGN, because at X-ray energies the contamination from non-active galaxies, many of which are star-forming, is far less significant than in optical and infrared surveys (Donley et al. 2008, 2012; Stern et al. 2012, Lehmer et al. 2012). Moreover, X-ray surveys with *Chandra* and *XMM-Newton* are very effective in selecting both unobscured and obscured AGN, including also a fraction of heavily obscured AGN at $z \simeq 1-2$ (Comastri et al. 2011; Georgantopoulos et al. 2013; Lanzuisi et al. 2015). Therefore, X-ray surveys can be used in combination with surveys in other bands to study the co-evolution of SMBHs and galaxies once large samples of AGN, both obscured and unobscured, are available. The moderate luminosity AGN that dominate the X-ray background need deep, moderate-area surveys ($\geq 1 \text{ deg}^2$), at sufficient depth to detect AGN at $z \leq 6$. Moreover, spectroscopic information, deep enough to detect faint sources (with L^* luminosities) even at $z \simeq 3$, is also required.

In this thesis, I will present the results of the *Chandra COSMOS-Legacy* survey (Civano et al. 2015; Marchesi et al, submitted), a 4.6 Ms X-ray survey covering the equatorial COSMOS area (Scoville et al. 2007a). *COSMOS-Legacy* is unusual in terms of size (2 deg^2) and depth (flux limit $f=2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the 0.5–2 keV band) in comparison with other X-ray surveys, which are either deep on pencil-beam area (e.g., 4 Ms CDF-S, Xue et al. 2011) or shallow on large-area surveys (e.g., Stripe82, LaMassa et al. 2013). In this regard, *COSMOS-Legacy* represents the path for surveys with future facilities, like *Athena* (Nandra et al. 2013) and *X-ray Surveyor* (Vikhlinin et al. 2012).

The final *Chandra COSMOS-Legacy* catalog contains 4016 point-like sources, more than 20% than other other contiguous surveys with similar or larger area, except XXL (Pierre et al., submitted). The plethora of multiwavelength data available in the COSMOS field allows us to compute a photometric redshift for 97% of the *Chandra COSMOS-Legacy* sources, all optically/infrared identified. Optically obscured AGN can be related to the main phase of BH accretion in a gas-rich environment. In the X-rays, where the obscuration is significantly less effective than in the optical, we can therefore fully characterize the AGN peak phase. In *Chandra*

COSMOS Legacy, 65% of the sources are optically obscured, thus allowing us to analyze this strategic BH growth phase on a wide range of redshifts and luminosities. Moreover, over 90% of the *Chandra COSMOS-Legacy* sources are below the knee of the luminosity function at any redshift, thus probing the faint end of the AGN population.

With *Chandra COSMOS-Legacy* it is possible to explore with excellent statistics the high-redshift Universe, using the largest X-ray selected sample of high-redshift AGN on a contiguous field (174 sources at $z \geq 3$). The rest frame 2-10 keV comoving space density at $z \geq 3$ can be used to place constraints on the BH formation scenario. In fact, the shape of the space density is linked to the time-scale of accretion of SMBHs and is therefore a tool to investigate the SMBH formation and growth scenario, eventually distinguishing between major-merger driven accretion and secular accretion. In the high-luminosity range of our survey ($L_X \geq L^*$), our data show a decline in the AGN space density, consistent with the one observed in optical surveys at higher luminosities (e.g. Masters et al. 2012; McGreer et al. 2013) and predicted by different X-ray background synthesis models (e.g. Gilli et al. 2007; Ueda et al. 2014; Aird et al. 2015). With our dataset, which samples lower AGN luminosities and SMBH masses than optical surveys, it will also be possible to test the predictions of different classes of black hole seed models. Moreover, our data significantly disagree with the prediction of models of BH activation through major merger (e.g., Shen 2009). These models have been calibrated on optical quasar surveys at luminosities a factor of 10 higher than those of *Chandra COSMOS-Legacy* ($L_{bol} > 10^{46}$ erg s $^{-1}$), and strongly overpredict our number counts at faint luminosities by a factor of 3-10. This exercise on the AGN space density suggests that in our luminosity range the BH triggering through secular accretion is likely preferred to a major-merger triggering scenario. This result is in agreement with other results obtained with different approaches, like those of Allevato et al. (2014 and in preparation), based on the AGN clustering, or the one of Cisternas et al. (2013), based on galaxy morphology.

The *Chandra COSMOS-Legacy* dataset, combined with the other multiwavelength COSMOS catalogs, can be used to answer questions related to a large number of astrophysical topics, such as those related to the interplay between accretion onto SMBHs and host galaxy growth; the accretion processes in low-mass and low-luminosity AGN; the growth and properties (e.g., bolometric luminosity, accretion rate) of X-ray selected SMBHs at early times.

Chapter 1

X-ray surveys and their role in AGN selection and investigation

1.1 Active Galactic Nuclei taxonomy

About 1-10% of all galaxies show luminosities (up to $L_{bol} \sim 10^{48}$ erg s⁻¹, e.g., Wu et al. 2010) much larger than those observed in ordinary galaxies, and that cannot be explained with stellar processes. Given that these luminosities were found to be produced from the center of the galaxy (within 1 pc³), these objects were called Active Galactic Nuclei (AGN). A standard AGN spectrum can be approximated with a power-law over an extended range of wavelengths, therefore suggesting a non-thermal emission mechanism; however, over in certain ranges of the emission spectrum thermal emission may be dominant as well.

AGN are mainly classified on the basis of their optical and radio properties. Optically, AGN are divided in Type 1, if a broad (with velocities of the emitting gas of ~ 2000 – 15000 km s⁻¹) emission line is observed in the optical spectra, or in Type 2 otherwise. The radio emission power is instead adopted to distinguish between Radio Loud and Radio Quiet AGN.

While different classes of AGN are observed, the general agreement is that all these objects can be described by a unified model (e.g., Antonucci 1993; Urry & Padovani 1995; Chiaberge et al. 2000). In this scenario, different type of AGN are mainly caused by different orientation effects, although also intrinsic physical properties can contribute to different observational effects. The main elements of a typical AGN are the following (see, e.g., Peterson et al. 2003):

- A **supermassive black hole (SMBH)** with $M_{BH} = 10^6$ – $10^9 M_{\odot}$.
- A **rotating accretion disk** falling towards the black hole. The temperature of the gas forming the disk can reach $T \sim 10^6$ K and is related to the distance from the black hole (BH), regions of the disk closer to the BH having higher temperatures. The accretion disk is usually linked to the AGN emission in optical/UV (see Section 1.3).
- The **Broad Line Region (BLR)**, made by dense ($n_e \sim 10^{10-11}$ cm⁻³) and warm ($T \sim 10^4$ K) gas clouds. These clouds are very close to central SMBH (with distances up to $\sim 10^{-3}$ pc). The broadening observed in the lines is

mainly caused by Doppler motion of single clouds, with typical velocities $v \sim 10^3 \text{ km s}^{-1}$.

- A **dusty torus** surrounding the SMBH, at distances $d \sim 0.3\text{--}3 \text{ pc}$. The torus temperature is lower than the dust sublimation temperature, which varies with the different dust composition: average sublimation temperatures are $T \sim 10^3 \text{ K}$ (e.g., Maiolino et al. 2007). While the presence of a torus is necessary to justify the existence of obscured Type 2 AGN, its structure is still debated. First models (e.g., Urry & Padovani 1995) proposed a homogeneous or clumpy (Nenkova et al. 2008; Elitzur 2008) dust distribution, with different regions having different temperature and densities. A different class of models (e.g., Elvis 2000) suggested that the obscuration observed in Type 2 AGN can be explained by outflowing winds driven by magnetic fields.
- The **Narrow Line Region (NLR)**. As for the BLR, the NLR is made by gaseous clouds. However, these clouds are at distances of 100–1000 pc from the SMBH, and have lower densities ($n_e \sim 10^{3-4} \text{ cm}^{-3}$) and temperatures ($T \sim 10^3 \text{ K}$) than the BLR. In this physical scenario, narrow forbidden lines (like [OIII] at 4959 and 5007 Å) can be emitted, because in this low density range radiative de-excitation processes have shorter time-scales than collisional ones. The narrow lines have width of $\sim 10^2 \text{ km s}^{-1}$
- The accretion disk is surrounded by a gaseous **hot corona**, with $T \sim 10^9 \text{ K}$: the causes of this high temperature have not been yet understood. Similarly, the structure of the corona is debated, with different models suggesting a clumpy (e.g., Schnittman & Krolik 2010) or a diffuse (e.g. Haardt & Maraschi 1991) structure.
- Radio-loud AGN show prominent **relativistic jets**. These jets are perpendicular to the torus plane and emit through synchrotron mechanism, probably caused by the black hole magnetic field (e.g., Pudritz et al. 2012). The jets can reach scales of 1 Mpc, and the interaction between jets and the surrounding intergalactic medium can produce features like hot spots and lobes.

Different classes of objects can be described assuming that the components reported above are observed from different lines of sight (l.o.s). For example, Type 2 AGN are objects where the dusty torus hides the BLR to the observer, only the NLR is visible and no broad line are detected in the optical spectra. In Type 1 AGN, instead, the l.o.s. does not intercept the torus, and both broad and narrow lines can be detected.

1.2 Black hole and galaxy growth history

It is widely believed that galaxies and their central supermassive black holes undergo closely coupled evolution. SMBH masses in the nuclei of nearby galaxies correlate with bulge luminosity (Kormendy & Richstone 1995) and stellar velocity dispersion, with a limited scatter (Magorrian et al. 1998; Gebhardt et al. 2000; Ferrarese & Merrit 2000; Merloni et al. 2010; McConnell & Ma 2013; see also Figure 1.1).

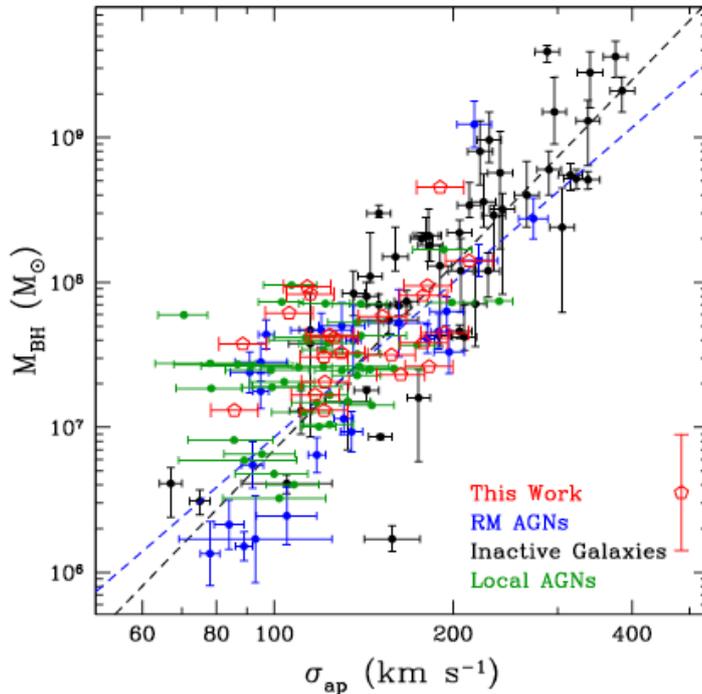


Figure 1.1: Compilation of BH masses against stellar velocity dispersion of their host galaxy bulges. Taken from Bennert et al. (2011).

Most SMBH - and definitely the most massive ones - had to grow during an active accretion phase, when they would be visible as an AGN, which implies that most bulges had an active phase in their past (Soltan 1982).

In the local Universe, $\sim 5\text{--}10\%$ of the galaxies show evidence of nuclear activity on the basis of optical spectral emission (e.g. Veilleux & Osterbrock 1987; Maiolino & Rieke 1995; Kewley et al 2001, 2006). In principle, using mid-infrared (mid-IR) spectroscopy it would be possible to detect AGN (through high-excitation emission lines like [Ne V] at $14.3\ \mu\text{m}$ or [O IV] at $25.9\ \mu\text{m}$) even in objects where the AGN narrow-line (NL) region is obscured by dust in the host galaxy (e.g. Satyapal et al. 2008, Goulding & Alexander 2009). However, the statistics from reliable mid-IR spectroscopy (hundreds of objects) is so far significantly poorer than the one available from optical spectroscopy (thousands of objects) and effective only at low redshifts.

While several works, on different classes of AGN (e.g., Best et al. 2005 with optically luminous AGN, i.e., objects with $L_{[OIII]} > 1.2 \times 10^{41}\ \text{erg s}^{-1}$, and radio-luminous AGN; Goulding et al. 2010 with low-luminosity AGN), show that in the nearby Universe mass accretion onto the central SMBH is a relatively common process, the BH fraction itself does not provide an estimation of how fast the BH increases its mass. To do so, it is strategic to estimate reliable BH masses and therefore estimate the AGN accretion rates and Eddington ratios, i.e., the ratio between the AGN Luminosity and the Eddington luminosity. The Eddington luminosity is defined as the maximum luminosity that an object of mass M can produce by

spherical accretion of fully ionized hydrogen, and is $L_E \sim 1.26 \times 10^{38} \frac{M}{M_\odot} \text{ erg s}^{-1}$. Moreover, different techniques have been developed to estimate BH masses, mainly based on the optical spectra of Broad Line AGN. We describe these techniques in section 1.3.

The different BH mass estimate techniques provide a significant statistics of BH masses (i.e., ~ 50 – 100 objects) and allowed to estimate that AGN in the local Universe have Eddington ratios in the range $\lambda = [10^{-6} - 10^{-3}]$ (Ho 2008; Goulding et al. 2010). Similar Eddington ratios are too small to justify the BH masses that we observe: therefore, the vast majority of massive BHs in the nearby Universe should have grown with considerably larger Eddington ratios in the past. Therefore, to understand what triggered the AGN activity at high-redshift and which kind of BH and host-galaxy growth processes were dominant in the past, is necessary to sample a sizable population of objects on a large range of both redshift and luminosity.

The tightness of the BH-bulge mass relationship (see Figure 1.1) suggests that in bulge-galaxies the BH and its host-galaxy have a related growth (on a scale of $\sim 1000 M_\odot$ of star formation for each ~ 1 - $2 M_\odot$ of black hole growth) over cosmic time. However, the star formation and black hole growth processes happen on significantly different scales ($\sim \text{kpc}$ versus $< 1 \text{ pc}$) and in principle no causal link between the two processes is expected. Actually, the observed AGN/star-formation ratio shows a large scatter when the whole SF activity is taken into account (e.g., Netzer et al. 2007; Wild et al. 2007; Baum et al. 2010), but the scatter decreases while the star-formation is measured only in the central kpc of the host galaxy (e.g., Wild et al. 2007). Moreover, the fraction of AGN increases with the galaxy IR luminosity (8 – $1000 \mu\text{m}$), which is a proxy of the star-formation rate: ~ 50 – 80% of objects with $L_{IR} > 10^{11} L_\odot$ also show evidence of AGN (e.g., Alexander et al. 2008; Lehmer et al. 2010; Nardini & Risaliti 2011). A similar scenario has also been confirmed by the existence of a proportional trend between the AGN obscured fraction, in both optical/IR and in the X-ray, and the host-galaxy SFR, which suggests that SMBHs preferentially grow in star-forming, dust-rich, environments (Chen et al. 2015). However, this relation has not been found in several previous works (e.g., Sturm et al. 2006; Zakamska et al. 2006, 2008; Mainieri et al. 2011; Merloni et al. 2014), which instead show no evidence of different morphological and SF properties between obscured and unobscured AGN; this discrepancy can be explained with different methods of obscuration estimation.

Many different works tried to determine the main AGN activity triggering process. The main candidate are major merger processes, which can accrete clumps of material with low angular momentum onto the galaxy center (Bellovary et al. 2013) and are the dominant accretion process for the more massive black holes ($> 50\%$ of the mass in black holes with $M > 10^9 M_\odot$ come from mergers, Dubois et al. 2012). Evidence of major merger in AGN was searched using both morphological evidence of merger (e.g., Darg et al. 2010; Schawinski et al. 2010; Koss et al. 2010; Cisternas et al. 2013) and detection of AGN pairs within ~ 5 – 100 kpc , final product of a merger, using both optical imaging (Liu et al. 2011) and spectroscopy (e.g. Komossa et al. 2008; Colpi & Dotti 2009; Rosario et al. 2011). These works have not produced a commonly accepted scenario so far; nonetheless, is largely accepted that bulge-dominated objects may have undergone at least one major-merger event during their cosmic life (e.g. Bournaud et al. 2005; Hopkins et al. 2010).

While so far we focused only on BH hosted in bulge-dominated galaxies, a significant fraction of AGN have been found also in late-type galaxies with only evidence of pseudo-bulges. No correlation has so far been found between BH mass and pseudo-bulge luminosity (e.g., Greene et al. 2008; Jiang et al. 2011; Kormendy et al. 2011). A similar evidence suggest that BHs in late-type galaxies could have had a different evolutionary path than BHs in bulges: for example, NL Seyfert 1 AGN (i.e., objects with no intrinsic obscuration), reside mainly in late-type systems, and may have never been effected by galaxy major merger, growing only through secular processes (Mathur et 2011; Orban de Xivry et al. 2011). Given that secular processes are significantly less effective than major mergers in accreting the BH mass, in these objects the mass of the BH would mainly come from the original black hole seed (e.g., Kormendy & Kennicutt 2004; Volonteri & Natarajan 2009; see also Section 1.7.2). In Figure 1.2 we show a simple representation of these two different SMBH growth scenarios.

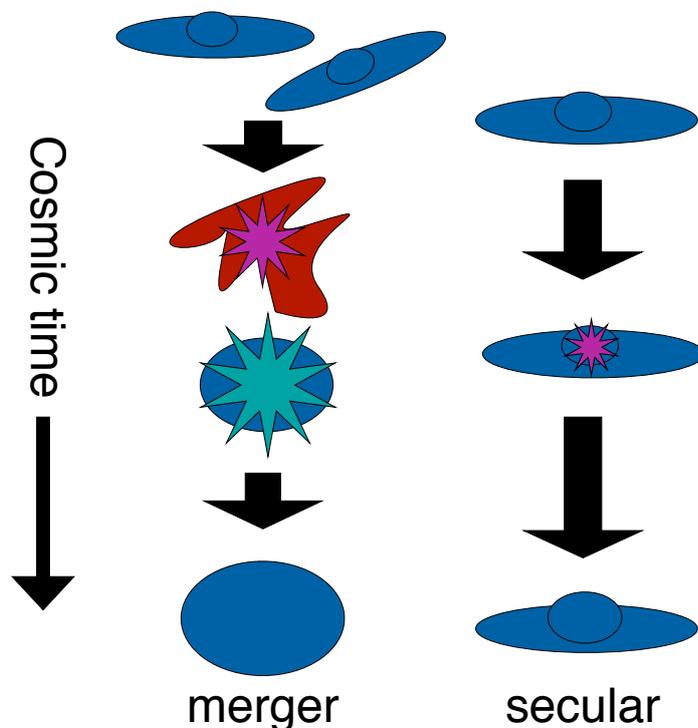


Figure 1.2: A schematic representation of an different SMBH growth scenarios.
Taken from Alexander & Hickox (2012).

Galaxies and AGN also show coeval cosmic “downsizing”: more luminous AGN and more massive galaxies formed earlier (and therefore their number density peaks at higher redshift) than less luminous AGN and less massive galaxies (Cowie et al. 1996). Massive galaxies exhibit a peak in star formation at $z \simeq 2$ (Cimatti et al. 2006; Madau and Dickinson 2014), and SMBH growth peaks in the same redshift range ($z=2-3$), as the quasar luminosity function (Hasinger et al. 2005; Silverman et al. 2008; Hasinger 2008; Ueda et al. 2014; Aird et al. 2015; Miyaji et al. 2015; see also Figure 1.3). Moreover, star formation rate in lower mass galaxies peaks at

$z=1-1.5$, and the same trend is followed by lower mass and lower luminosity AGN (La Franca et al. 2005; Bongiorno et al. 2007). These evidences can again be explained with a “BH -galaxy co-evolutionary scenario”, where the AGN feedback during its accretion phase can cause massive gas outflows and therefore a burst of star formation (e.g., Feruglio et al. 2010; Maiolino et al. 2012; Cicone et al. 2014).

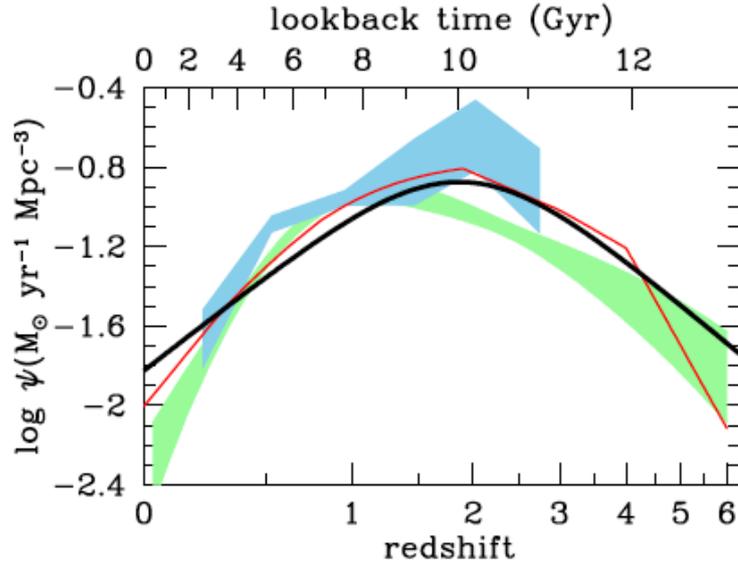


Figure 1.3: Evolution with redshift of the SFR density, compared with black-hole accretion estimates (scaled up by a factor 3300) from Shankar et al. (2009, red curve), Aird et al. (2010, green area) and Delvecchio et al. (2014, blue area). Taken from Madau & Dickinson (2014).

1.3 AGN emission mechanisms from radio to UV

The bolometric output of an AGN is due to the combination of different mechanisms generated in different regions of the source. In Figure 1 we show the schematic representation of the Spectral Energy Distribution (SED) of an AGN based on observations of radio quiet quasars (e.g., Elvis et al. 94; Richards et al. 2006). In the following paragraphs we briefly describe the different mechanisms responsible for emission at different frequencies, following the review by Risaliti & Elvis (2004). An extended description of the X-ray emission mechanisms is shown in Section 1.4.

- **Radio** (yellow line), between 0.1 and 10 GHz. AGN radio emission is mainly produced and observed in jets and lobes. This emission is caused by a synchrotron process from relativistic electrons accelerated by a magnetic field. The spectral shape of the emission process is a power-law,

$$F_\nu \propto \nu^{-\alpha}. \quad (1.1)$$

where F_ν is the source flux at frequency ν , while α is the energy spectral index, with typical observed values in the range $\alpha \sim [0-2]$. The radio emission

contribution to the AGN bolometric luminosity is almost negligible for radio-quiet AGN (yellow solid line in Figure 1.4). Similarly, the contribution of radio-loud AGN (yellow dotted line) to the source bolometric luminosity is marginal, although some orders of magnitude more luminous in the radio than radio-quiet objects.

- **millimeter (mm)–Far Infrared (FIR)**, from 3 cm to 200 μm . In this energy range, the contribution of the nuclear emission to the total one is uncertain because the emission is dominated by cold dust associated to SF processes. Nonetheless, it is commonly accepted that the AGN emission at these wavelengths, when present, is caused by synchrotron emission from relativistic electrons, the same process observed in radio-loud AGN. The AGN contribution is negligible, compared to the star formation one, for radio-quiet AGN, while this could not be true only for the most powerful radio-loud AGN (dotted yellow line).
- **Infrared (IR, 1-200 μm)** (red dashed line). The IR emission is usually linked to the dusty torus surrounding the SMBH. UV photons emitted by the accretion disk (see below) are absorbed by the dust grains, and then re-emitted in the IR. The spectral shape is formed by three different components: (i) a minimum at $\sim 1\text{--}2 \mu\text{m}$, corresponding to 1000–2000 K. At these temperatures, the most refractory dust grains sublimate. In this wavelength range, emission from stars and cold galactic dust may be significant and therefore hide the AGN emission. (ii) The so-called “IR-bump”, usually between 10 and 30 μm , caused by dust thermal emission (with temperatures between 50 and 1000 K). This is the most important AGN feature in the IR, because at these wavelengths the host galaxy contribution is smaller than in the near-IR. The flattening of the SED in the mid-IR can consequently be used to select AGN (Donley et al. 2012). (iii) A steep decline at $\lambda > 50\text{--}100 \mu\text{m}$. In this wavelength range the SF contribution is the dominant one in the whole galaxy SED.

The total IR contribution to the bolometric luminosity is on average $\sim 30\%$.

- **Optical/UV** (blue dashed dotted line). In this wavelength regime the emission is dominated by “Big Blue Bump” at $\lambda \sim 1000\text{--}4000 \text{ \AA}$, caused by the combination of black-body emission at different temperatures from the accretion disk:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1}, \quad (1.2)$$

resulting in a power-law spectrum $F \propto \nu^{1/3}$. The peak of AGN emission is usually found in the UV, associated with the Big Blue Bump.

AGN optical/UV spectra also show evidence of numerous emission lines (e.g., Vanden Berk et al. 2001; Telfer et al. 2002). These lines can be permitted and “broad” (with corresponding velocities of the emitting gas of $\sim 2000\text{--}15000 \text{ km s}^{-1}$). The Broad Line Region (see also 1.1) is probably made by dense ($n_e \sim 10^{10\text{--}11} \text{ cm}^{-3}$) gas clouds extremely close to the SMBH ($d \sim \text{few } 10^{-3} \text{ pc}$).

The width of broad lines (BL, commonly H β , Mg II and C IV) can be linked to the BH mass using the equation

$$M_{BH}(H\beta) = 1.05 \times 10^8 \left(\frac{L_{5100}}{10^{46} \text{ ergs}^{-1}} \right)^{0.65} \left[\frac{FWHM(H\beta)}{10^3 \text{ km s}^{-1}} \right]^2 M_{\odot}. \quad (1.3)$$

This expression is based on the assumption that the BLR is completely virialized (see Onken et al. 2004; Grier et al. 2013) and on R_{BLR} - L_{5100} relation calibrated in the low-redshift Universe using a more time-consuming technique, the so-called ‘‘reverberation mapping’’ (e.g., Peterson 1993; Peterson et al. 2004; Kaspi et al. 2005). The reverberation mapping BH mass estimation formula is

$$M_{BH} = \frac{f R_{BLR} \Delta v^2}{G}, \quad (1.4)$$

where Δv is the velocity dispersion of the gas close to the BH, estimated through Doppler broadening; f is a parameter connected to the shape of the BLR, which is so far not well constrained; and R_{BLR} is the radius of the BLR. To estimate R_{BLR} , repeated observations over several years are required. In fact, it is well known (e.g., Peterson & Horne 2004) that variations in the emission line flux are strongly related to variations in the continuum emission, produced by the accretion disk. Moreover, the emission line variation shows a delay with respect to the continuum one, and this delay can be linked to the distance between the BLR and the accretion disk.

A second type of lines is also observed in AGN spectra: these lines are forbidden and ‘‘narrow’’ ($v \sim$ few hundreds km s $^{-1}$). As we explained in section 1.1, the Narrow Line Region is believed to be more distant from the BH ($d=100$ - 1000 pc) than the BLR, and significantly less dense ($n_e \sim 10^{3-4}$ cm $^{-3}$), therefore allowing forbidden-line emission.

Finally, while AGN emission is dominant in the optical/UV band for very bright quasars, this is not true for less luminous AGN, where the galaxy emission from star-forming processes and/or old stars can be the main contribution to the total SED. Consequently, to properly select AGN is necessary to move to higher energies, where the total emission is less biased by the galaxy contribution.

1.4 AGN emission mechanisms in the X-ray

A visual summary of the different X-ray emission mechanisms is shown in Figure 1.5. We can distinguish five main components:

1. The AGN X-ray emission is mainly originated by the **Inverse Compton (IC) up-scattering** of UV photons from the BH thermal accretion disk. The IC process is caused by the hot, energetic corona surrounding the SMBH (e.g., Haardt & Maraschi 1993; Mushotzky et al. 1993; Reynolds and Nowak 2003; Fabian 2006). The spectral outcome of this process is a power law (magenta line in Figure 1.5) and is due to the superposition of different scattering orders of the initial black body profile:

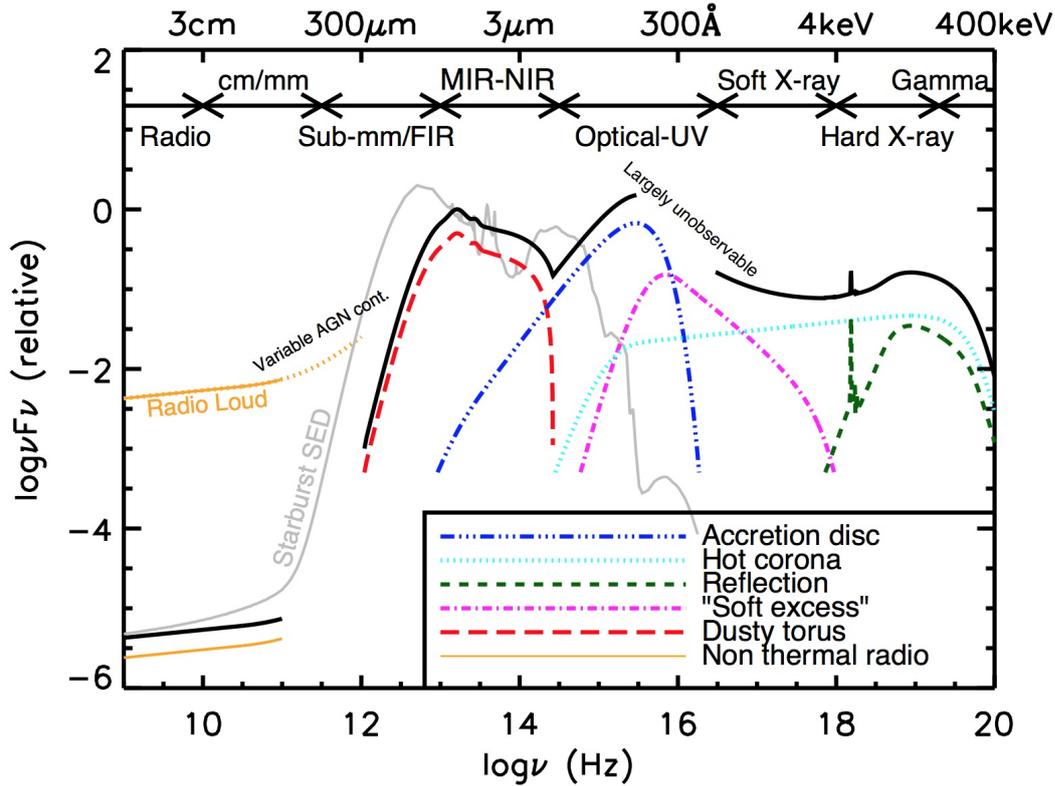


Figure 1.4: A schematic representation of an AGN SED. The total SED is plotted as a black solid curve, while the individual components are shown using different colors (and with an arbitrary normalization offset). Also shown is an example radio–UV SED of a starburst galaxy (grey curve; SED of M82, from the GRASIL library; Silva et al. 98). The AGN contribution to the total SED, for a radio-quiet AGN, becomes dominant in the UV region, in the so-called “blue bump”, where the emission is due to the accretion disk. Figure taken from the personal website of Chris Harrison, <http://astro.dur.ac.uk/~cpnc25/research.html>.

$$F_\nu \propto \nu^{-\Gamma} [\text{photons/s/keV}]. \quad (1.5)$$

Here, $\Gamma = \alpha + 1 \approx 1.8 - 2.0$, with $\alpha = \frac{-\ln \tau}{\ln A}$ being the spectral index. τ is the optical depth of the hot plasma, while $A = e^y$ is the average energy gain in each process of scattering, y being the Comptonization parameter $y = \frac{4kT}{m_e c^2} \tau$. kT is the temperature of the hot corona. Observationally, the power-law energy cut-off is at $\sim 100-300$ keV, the thermalization temperature of electron and photons.

2. The so-called “**soft excess**” below 1 keV (cyan line), i.e., an excess of X-ray emission compared to the extrapolation to lower energies of the best-fitting 2-10 keV continuum. This excess was originally thought to be linked to the high-energy tail of the Big Blue Bump (see Section 1.3). However, more recent works (e.g., Done et al. 2012) suggest that other processes, like absorption and reflection, may contribute to the soft excess. Unfortunately, Galactic cold

gas absorption between $\sim 100 \text{ \AA}$ and 912 \AA , the so-called Lyman continuum edge, does not allow to make observations below 0.1–0.3 keV and therefore verify or reject the hypothesis that the soft excess and the Big Blue Bump have actually the same origin.

3. **Warm absorber** (absorption features in the magenta line power-law profile). Absorption features in the soft X-ray spectra are probably due to an outflowing gas with an equivalent hydrogen column density of about $\sim 10^{22} \text{ cm}^{-2}$. This outflowing gas is not related to the BLR, which has a higher degree of ionization; it could however be the same outflowing absorber observed in many UV spectra, which show blueshifted absorption lines (see e.g., Giustini et al. 2011). These extreme outflows could also be responsible to the quenching of the star formation, and offer a physical justification to the tight SMBH-galaxy relation we described in Section 1.2 (Tombesi et al. 2015).
4. **Compton hump**. The power-law emission from the accretion disk is reflected by the Compton-thick ($\tau > 1$ and $N_H > 10^{24} \text{ cm}^{-2}$) disk through Thomson scattering. This reflection produces a hardening of the spectrum, with a peak in the SED at $\sim 30 \text{ keV}$ (green line). The intensity of the reflection component can be significantly different in different objects, and is related to different parameters, like the geometry of the reflecting thick disk and its ionization state.
5. **Iron $K\alpha$ line at 6.4 keV** (red line). This line is the byproduct of a Fe-K $n=2-1$ transition of iron not highly ionized (i.e., $\leq \text{FeXVII}$). The line has typical equivalent width (EW) of 100–200 eV, the EW being defined as

$$EW = \frac{1}{I_c(\nu_l)} \int I_l(\nu) d\nu, \quad (1.6)$$

where $I_c(\nu_l)$ is the continuum intensity at the line energy, while $I_l(\nu)$ is the intensity of the line itself.

The iron $K\alpha$ line is generally connected to fluorescence emission (e.g., Fabian et al. 2000; Fabian & Miniutti 2005), in the accretion disk and/or in the dusty torus. If the line is emitted from the accretion disk, it presents black-hole induced relativistic effects, i.e., a redshifted double-peaked broadened line profile, with an enhanced blue peak and a large red wing.

1.5 Advantages of the X-ray analysis to study AGN

The X-ray emission contribution to the AGN bolometric luminosity is of $\sim 10\%$, but X-rays play an important role in the selection of AGN, and are strategic to study SMBH accretion properties, for many reasons. First of all, at these energies the contamination from non-nuclear emission, generally linked to star-formation processes, is considerably less significant than in optical and infrared (Donley et al. 2008, 2012; Stern et al. 2012; see also Figure 1.4). Objects with $L_X > 10^{42} \text{ erg s}^{-1}$ in the 2-10 keV band can safely be assumed as AGN, given that non-AGN X-ray emission in galaxies, mainly from X-ray binaries and diffuse hot gas (with a more

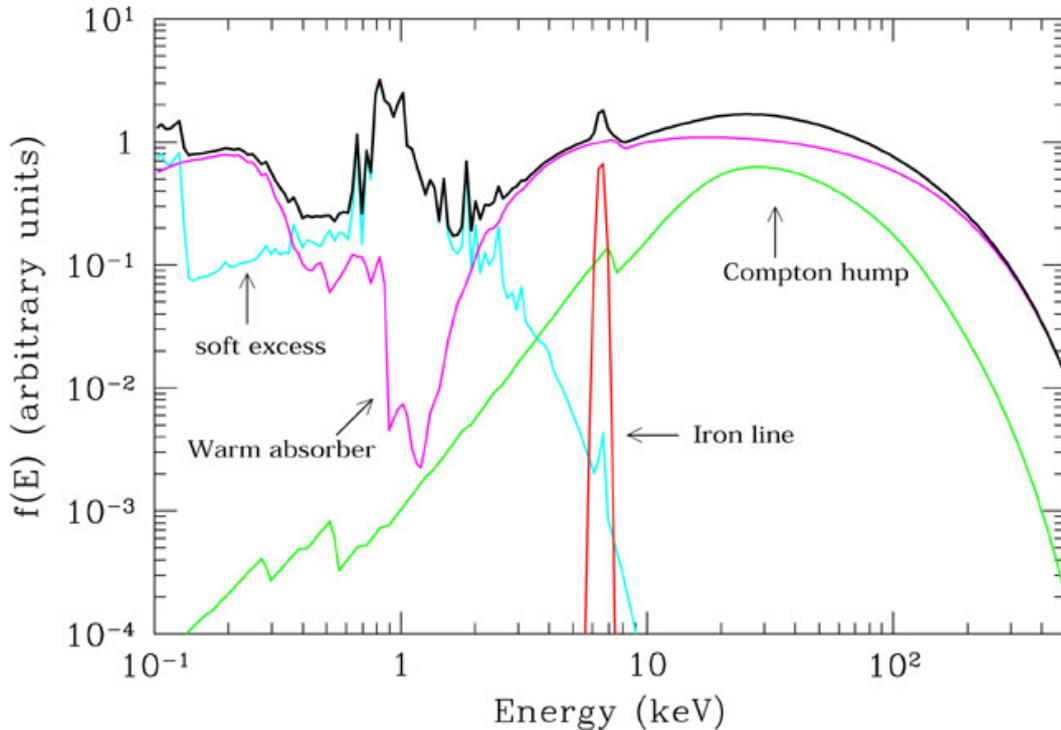


Figure 1.5: Typical X-ray spectrum of a type I AGN (black line). The main primary continuum component is a power law with an high energy cut-off at $E \sim 100\text{--}300$ keV, absorbed at soft energies by warm gas with $N_H \simeq 10^{21}\text{--}10^{23}$ cm^{-2} (magenta line). A cold reflection component is also shown (green line). The most relevant narrow feature is the iron $K\alpha$ emission line at 6.4 keV (red line). Finally, the so-called "soft excess" (cyan line) is shown. Taken from Risaliti & Elvis (2004).

significant contribution in the soft band, see, e.g., Kim & Fabbiano 2012; Civano et al. 2014; Paggi et al. 2015), has found to be at most $\sim 10^{42}$ erg s^{-1} (e.g., Ranalli et al. 2003; Mineo et al. 2012; Fragos et al. 2013).

Moreover, even in deep X-ray surveys the integrated emission is dominated by AGN. We show in Figure 1.6 the number counts for the 4 Ms CDF-S from Lehmer et al. (2012). As can be seen, the AGN contribution (blue) is the dominant one at any flux in the 2-8 keV band; in the soft 0.5-2 keV band, the AGN contribution is instead dominant down to a flux $f \simeq 10^{-17}$ $\text{erg s}^{-1} \text{cm}^{-2}$, i.e. at the flux limit of the 4 Ms CDF-S survey. Only at fainter fluxes, which are going to be reached with the new 7 Ms CDF-S survey, the galaxy contribution will be dominant.

Unlike optical surveys, which are biased towards luminous, unobscured AGN, X-ray surveys with *Chandra* and *XMM-Newton* are very effective at selecting both unobscured and obscured AGN. These survey can include also a fraction of AGN in the Compton-thick regime, with optical depth $\sigma_T \sim 1$, i.e., with column densities, N_H , up to 10^{24} cm^{-2} , reaching redshifts $z \simeq 1\text{--}2$ (Comastri et al. 2011; Georgantopoulos et al. 2013; Lanzuisi et al. 2015; Buchner et al. 2015). Recently, several works with the hard X-ray telescope *NuSTAR* (Harrison et al. 2013) were also able to detect a number of sources above the 10^{24} cm^{-2} threshold. Due to *NuSTAR* lower sensitivity compared to *Chandra* and *XMM-Newton*, these candidate

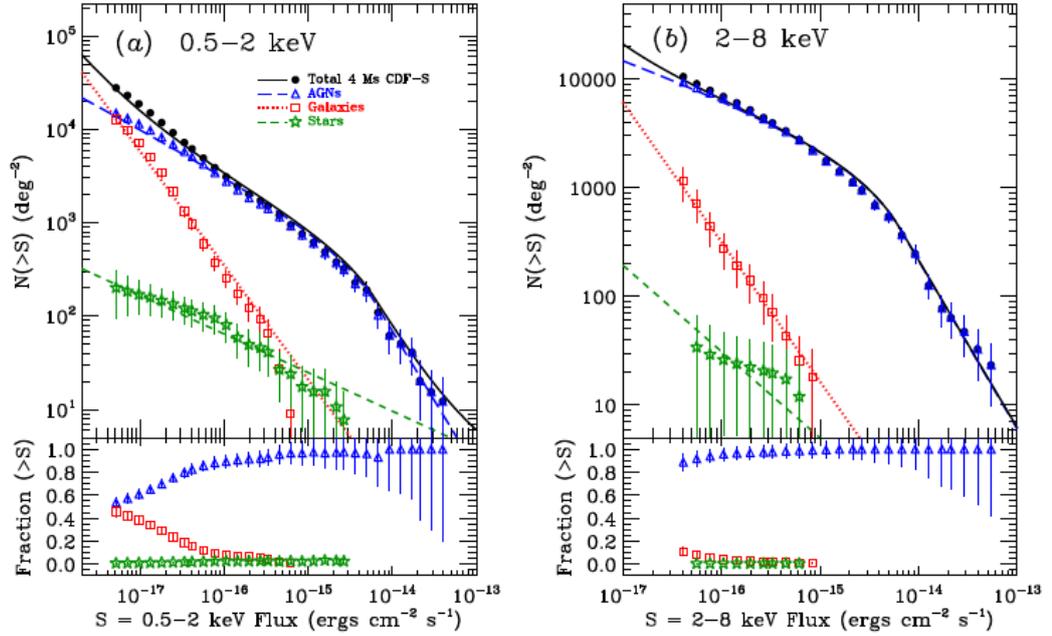


Figure 1.6: Cumulative number counts for the 4 Ms CDF-S in 0.5–2 keV (left) and 2–8 keV (right) bands. The total number counts are plotted in black. AGN (blue), galaxy (red) and stars (green) contribution is also plotted. In the bottom panel the fractional contribution of each source class to the total number counts is shown. The AGN contribution is dominant even at faint fluxes: only at fluxes fainter than 5×10^{-18} erg s⁻¹ cm⁻² in the 0.5–2 keV band the galaxy contribution becomes the most significant one. Figure from Lehmer et al. (2012).

Compton-thick sources have been detected at lower redshift ($z < 1$; Lansbury et al. 2014, 2015; Civano et al. 2015).

Finally, with X-ray we can sample lower bolometric luminosities than with optical surveys. In fact, at low optical luminosities, the standard color-color quasar identification procedure becomes less reliable, because stars cannot be easily disentangled from quasars: consequently, low-luminosity optical surveys have so far produced AGN luminosity functions with results in significant disagreement (see Ikeda et al. 2011; Glikman et al. 2011). Moreover, optical surveys are biased against obscured sources, whose contribution becomes also more significant at low luminosities. Given that with X-ray emission we can track objects at lower luminosities than with optical surveys, X-ray selected AGN could be used to track different BH accretion processes and, at high redshift, different populations of BH seeds (see Section 1.7.2).

1.6 X-ray surveys and the “wedding cake” strategy

The co-evolution of SMBH and galaxies can be studied with sizable samples of AGN, both obscured and unobscured, with sufficient multiwavelength data to disentangle selection effects. To access the moderate luminosity AGN that dominate the X-ray

background requires a deep moderate-area survey ($\geq 1 \text{ deg}^2$ in area, at sufficient depth to detect AGN up to $z \sim 6$), on areas wide enough to measure large-scale structures and find rare objects. Moreover, spectroscopic information deep enough to detect faint sources (with L^* luminosities) even at $z \simeq 3$ is also required.

In Figure 1.7 we report a summary of X-ray extragalactic surveys, taken from Brandt & Alexander (2015). Thanks to the development and launch of both *Chandra* and XMM-*Newton* satellites, in the last 15 years many X-ray surveys were performed on both deep and wide-area fields. These surveys produced catalogs of X-ray emitting AGN and galaxies, which have then been combined with extended multiwavelength spectroscopic and photometric information.

As shown in Figure 1.8, all the surveys lie in a “survey locus” (yellow area), computed taking into account the region of the diagram covered by 80% to 20% of the maximum area of each survey. These contiguous surveys follow a “wedding cake” strategy, being layered in decreasing area and increasing depth (see Figure 1.8). Following this strategy, all surveys obtain a comparable number of detected objects, covering a broad range of both redshifts and luminosities.

Wide/shallow surveys occupy one extreme of the “wedding cake”. Surveys like XBootes (9 deg^2 ; Murray et al. 2005), Stripe 82X (31.2 deg^2 , LaMassa et al. 2013a and submitted), XXL (50 deg^2 , Pierre et al. submitted) and 3XMM ($\simeq 880 \text{ deg}^2$, Rosen et al. submitted) are designed to cover large volumes of Universe and their main aim is to find rare objects, i.e., high-luminosity and/or high-redshift AGN. However, their statistics becomes poor in the luminosity range where low and intermediate-luminosity AGN are located.

At the opposite extreme are narrow/ultra-deep surveys like the 4 Ms *Chandra* Deep Field South (CDF-S, 0.1 deg^2 ; Xue et al. 2011; other 3 Ms of *Chandra* time have were granted in *Chandra* Cycle 15). Similar surveys g, even at medium to high redshifts (Luo et al. 2011; Lehmer et al. 2012; see also Figure 1.6). However, even if deep pencil-beam surveys can potentially detect faint AGN up to $z \geq 5$, they lack of statistically thorough samples at any redshift (e.g., Weigel et al. 2015 showed that the CDF-S does not appear to contain any AGN at $z > 5$).

To complete the census of the AGN population is therefore necessary to bridge the gap between deep and wide-area surveys. A similar task has mainly been performed with X-ray surveys on the 2.2 deg^2 of the COSMOS field. This whole thesis is focused on the *Chandra* COSMOS-*Legacy* survey, while we extensively describe previous works with both XMM-*Newton* and *Chandra* in Section 1.6.1.

1.6.1 X-ray surveys in the COSMOS field

The Cosmic Evolutionary Survey (COSMOS; Scoville et al. 2007) field is a $1.4^\circ \times 1.4^\circ$ field located close to the celestial equator, thus being visible by all astronomical facilities, especially by all large optical/IR telescopes. The COSMOS field alignment is east–west, north–south, and the field is centred at R.A.= $10^h 00^m 28.6^s$, declination= $+02^\circ 12' 21.0''$ (J2000).

Thanks to its strategic location, the COSMOS field has already been targeted with multiple instruments, and a large multiwavelength photometric and spectroscopic database is already available (Schinnerer et al. 2007; Sanders et al. 2007; Taniguchi et al. 2007; Capak et al. 2007; Koekemoer et al. 2007; Zamojski et al.

Survey Name	Rep. Eff. Exp. (ks)	Solid Angle (arcmin ²)	Representative Reference
<i>Chandra</i> (0.3–8 keV)			
<i>Chandra</i> Deep Field-South (CDF-S)	3870	465	Xue et al (2011)
<i>Chandra</i> Deep Field-North (CDF-N)	1950	448	Alexander et al (2003)
AEGIS-X Deep	800	860	Goulding et al (2012)
SSA 22 protocluster	392	330	Lehmer et al (2009a)
HRC Lockman Hole	300	900	PI: S.S. Murray
Extended CDF-S (E-CDF-S)	250	1,128	Lehmer et al (2005)
AEGIS-X	200	2,412	Laird et al (2009)
Lynx	185	286	Stern et al (2002)
LALA Cetus	174	297	Wang et al (2007)
LALA Boötes	172	346	Wang et al (2004)
C-COSMOS and COSMOS-Legacy	160	6,120	Elvis et al (2009)
SSA 13	101	345	Barger et al (2001b)
Abell 370	94	345	Barger et al (2001a)
3C 295	92	274	D’Elia et al (2004)
ELAIS N1+N2	75	590	Manners et al (2003)
WHDF	72	286	Bielby et al (2012)
CLANS (Lockman Hole)	70	2,160	Trouille et al (2008)
SEXTI ^a	45	7,920	Harrison et al (2003)
CLASXS (Lockman Hole)	40	1,620	Trouille et al (2008)
13 hr Field	40	710	McHardy et al (2003)
ChaMP ^a	25	34,560	Kim et al (2007)
XDEEP2 Shallow	15	9,432	Goulding et al (2012)
<i>Chandra</i> Source Catalog (CSC) ^a	13	1,150,000	Evans et al (2010)
Stripe 82X— <i>Chandra</i> ^a	9	22,320	LaMassa et al (2013b)
NDWFS XBoötes	5	33,480	Murray et al (2005)
<i>XMM-Newton</i> (0.2–12 keV)			
<i>Chandra</i> Deep Field-South (CDF-S)	2820	830	Ranalli et al (2013)
Lockman Hole	640	710	Brunner et al (2008)
<i>Chandra</i> Deep Field-North (CDF-N)	180	752	Miyaji et al (2003)
13 hr Field	120	650	Loaring et al (2005)
ELAIS-S1	90	2,160	Puccetti et al (2006)
Groth-Westphal	81	730	Miyaji et al (2004)
COSMOS	68	7,670	Cappelluti et al (2009)
Subaru <i>XMM-Newton</i> Deep Survey (SXDS)	40	4,100	Ueda et al (2008)
Marano Field	30	2,120	Lamer et al (2003)
HELLAS2XMM ^a	25	10,440	Baldi et al (2002)
XMM-LSS XMDS	23	3,600	Chiappetti et al (2005)
3XMM ^a	15	2,300,000	Watson (2012)
Stripe 82X— <i>XMM-Newton</i> ^a	15	37,800	LaMassa et al (2013a)
XMM-LSS	10	39,960	Chiappetti et al (2013)
XMM-XXL	10	180,000	Pierre (2012)
Stripe 82X— <i>XMM-Newton</i> Targeted	8	129,600	PI: C.M. Urry
<i>XMM-Newton</i> Slew Survey (XMMSL1) ^a	0.006	8×10^7	Warwick et al (2012)

Figure 1.7: Selection of extragalactic X-ray surveys with *Chandra* and *XMM-Newton*. Taken from Brandt & Alexander (2015).

2007; Lilly et al. 2007; Trump et al. 2007; Ilbert et al. 2009; McCracken et al. 2010; Laigle et al. submitted).

The whole COSMOS field was first covered in the X-rays with XMM-COSMOS (Hasinger et al. 2007; Cappelluti et al. 2009; Brusa et al. 2010, hereafter B10). The XMM-COSMOS survey consists of 55 *XMM-Newton* pointings, for a total exposure of ~ 1.5 Ms over 2.13 deg^2 of the COSMOS field. The average, vignetting-corrected, depth of the observation is of $\simeq 40$ ks. The catalog contains 1887 sources, down

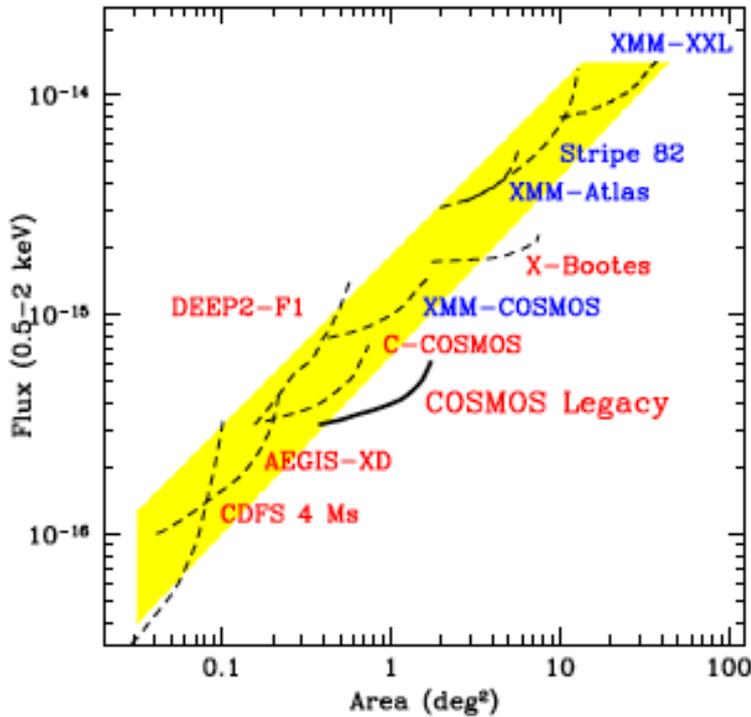


Figure 1.8: Area–flux curves for *Chandra* (red) and XMM-*Newton* (blue) contiguous X-ray surveys. Each survey has been plotted using each sensitivity curve starting from the flux corresponding to the area that is 80% of the maximum area for that survey to the flux corresponding to the 20% of the total area. The plotted surveys are: CDFS 4Ms (Xue et al. 2011), XDEEP2-F1 (Goulding et al. 2011), AEGIS-XD (Nandra et al. 2015), C-COSMOS (Elvis et al. 2009), XMM-COSMOS (Cappelluti et al. 2009), X-Bootes (Murray et al. 2005), XMM-Atlas (Ranalli et al. 2015), Stripe 82 (LaMassa et al. 2013a), XMM-XXL (PI: Pierre; see also Pierre et al. 2004). The survey locus described in the last section is drawn in yellow.

to flux limits of $\sim 5 \times 10^{-16}$ erg cm $^{-2}$ s $^{-1}$, $\sim 3 \times 10^{-15}$ erg cm $^{-2}$ s $^{-1}$ and $\sim 7 \times 10^{-15}$ erg cm $^{-2}$ s $^{-1}$ in the 0.5–2 keV, 2–10 keV and 5–10 keV, respectively. $\sim 98\%$ of the sources have an optical counterpart, $\sim 95\%$ have IRAC near-IR counterparts and $\sim 78\%$ of the sources have MIPS 24 μ m counterparts. About 99% of the sources also had either a photometric or a spectroscopic redshift, while $\sim 50\%$ of the sources have a spectroscopic redshift. Thanks to XMM-COSMOS, the AGN sample at high-X-ray luminosities ($L_X > 10^{44}$ erg s $^{-1}$) has been analyzed in many different works on different topics (e.g. Brusa et al. 2009; Allevato et al. 2011; Mainieri et al. 2011; Bongiorno et al. 2012; Lusso et al. 2012, 2013; Merloni et al. 2014; Miyaji et al. 2015 among others).

The central 0.9 deg 2 of the COSMOS field have also been covered with 1.8 Ms of *Chandra* observations (Elvis et al. 2009, E09; Puccetti et al. 2009, hereafter

P09; Civano et al. 2012a, hereafter C12). The C-COSMOS catalog contains 1761 sources down to flux limits ~ 3 times fainter than XMM-COSMOS, i.e., $\sim 2 \times 10^{-16}$ erg cm $^{-2}$ s $^{-1}$ and $\sim 9 \times 10^{-16}$ erg cm $^{-2}$ s $^{-1}$ in the 0.5-2 keV and 2-10 keV bands, respectively. C-COSMOS also detected sources in the 0.5-10 keV band, down to a flux limit $\sim 7 \times 10^{-16}$ erg cm $^{-2}$ s $^{-1}$. This improved sensitivity is due to *Chandra* low background and to its subarcsecond accuracy angular resolution, significantly better than the XMM-*Newton* one ($\simeq 0.5''$ and $\simeq 6''$ for *Chandra* and XMM-*Newton* full width half maximum, FWHM, respectively). The combination of area and sensitivities in C-COSMOS allowed to study faint and/or rare systems (e.g. Fiore et al. 2009; Civano et al. 2010, 2012b; Capak et al. 2011; Lackner et al. 2014) and to measure large-scale clustering in the Universe (Allevato et al. 2014).

1.7 AGN X-ray space density as a SMBH evolution estimator

Several works established that the SMBH activity changes over cosmic time: at any luminosity, the number of AGN is larger at $z > 1$ than in the local Universe. A similar behaviour has been observed also in SFR evolution, star-forming galaxies being significantly more common at $z > 1$.

These independent results suggest that SMBH and host galaxy should experiment co-evolution during their accretion history. However, such a relation is less constrained at high redshift ($z > 3$), before the peak in SF and AGN activity, where the statistics is significantly poorer. Moreover, recent works (e.g. Trakhtenbrot et al. 2015) revealed that the co-evolutionary scenario could be less reliable at $z > 3$ (see Chapter 1 on different accretion scenarios).

In this section we describe the main results obtained so far in studying the AGN evolution, first focusing on different phenomenological model calibrated at low redshifts, then presenting a summary of the works who investigated the high redshift ($z > 3$).

1.7.1 Phenomenological models of AGN luminosity function

We discussed in section 1.2 the existence of a co-evolutionary trend between AGN and their host galaxies. Massive galaxies exhibit a peak in star formation at $z \simeq 2$ the same redshift range ($z=2-3$) where the supermassive black hole (SMBH) activity peaks, as seen with the quasar luminosity function.

During the last fifteen years, phenomenological models of the AGN luminosity function have been developed using hard X-ray surveys. Many works are in general agreement towards fitting their X-ray luminosity function using analytical formulae. The models are usually based on a smoothed double power-law: at the low-luminosity end the data are fitted with a shallow slope, while at high luminosity with a steep one. Many works have used the so-called luminosity-dependent density evolution (LDDE) model (e.g. Ueda et al. 2003; Hasinger et al. 2005; Ueda et al. 2014; Miyaji et al. 2015; Buchner et al. 2015). The most recent versions of this model show evidence of the so-called ‘‘AGN downsizing’’: the peak of the AGN space density is at $z \simeq 2-3$ for more luminous AGN ($L_X > 10^{45}$ erg s $^{-1}$), for which the peak is followed by an exponential decline down to $z \simeq 6$; on the other hand, less

luminous AGN ($L_X < 10^{45}$ erg s $^{-1}$) show a peak shifted towards more recent times, $z \simeq 1-2$, and this peak is then followed by a slow decline to the highest redshifts reached so far ($z \simeq 3$).

A different model, the luminosity and density evolution model (LADE), has been instead proposed by Aird et al. (2010): in this model the shape of the XLF is the same at all redshifts, but with a shift in luminosity and a decrease in density at increasing redshifts. The agreement between LADE and LDDE models is good at $z < 2-3$, while at higher redshifts their predictions become significantly different. A further model, the flexible double power-law (FDPL), has been proposed in a more recent work of the same group (Aird et al. 2015): the agreement between the FDPL and the LDDE models show an improvement, with respect to the LADE model, at high redshifts.

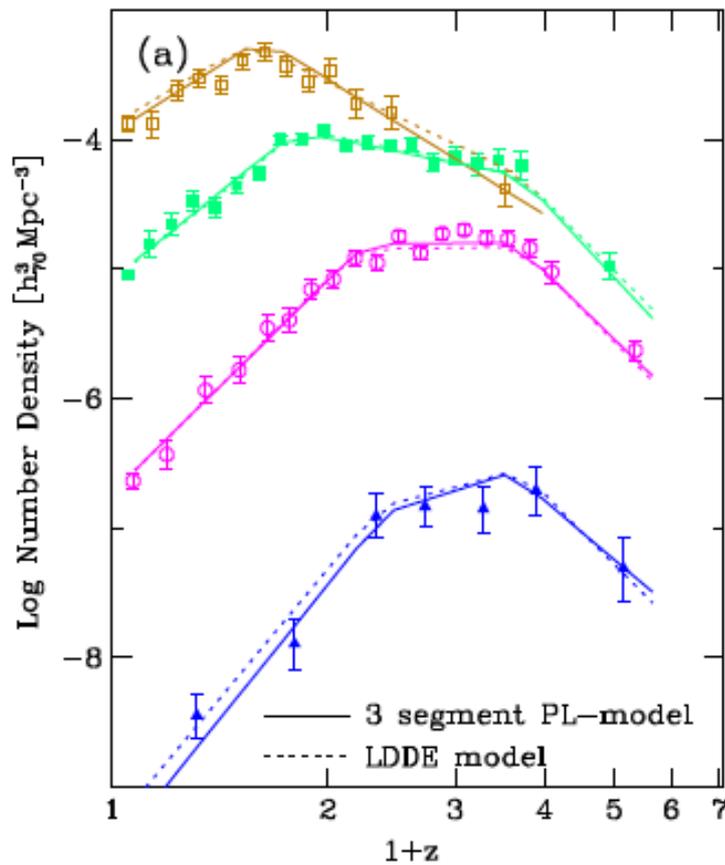


Figure 1.9: AGN Comoving number density redshift evolution in different luminosity bins: $L_X=[10^{42}-10^{43}]$ erg s $^{-1}$ (brown open squares); $L_X=[10^{43}-10^{44}]$ erg s $^{-1}$ (green squares); $L_X=[10^{44}-10^{45}]$ erg s $^{-1}$ (magenta circles); $L_X=[10^{45}-10^{46}]$ erg s $^{-1}$ (blue triangles). More luminous AGN peak at higher redshifts. Figure from Miyaji et al. (2015).

The role of obscured AGN

Obscured AGN (i.e., objects with column density $N_H > 10^{22} \text{ cm}^{-2}$) represent a significant fraction of the whole AGN population (up to 80%), as demonstrated by AGN population synthesis models for the cosmic x-ray background (CXB; e.g., Setti & Woltjer 1989; Comastri et al. 1995; Gilli et al. 2007). The X-ray obscuration can be caused by both photoelectric absorption and Compton scattering, the former being more effective at softer energies. In the energy range $E=[1-10] \text{ keV}$ the absorption contribution, usually linked to the dusty torus (e.g. Bianchi et al. 2012), becomes more important as the N_H increases, with a stronger flux depletion at lower energies. Consequently, the observed photon index is flattened ($\Gamma \sim 1.4$, as observed in the CXB, instead than $\Gamma \sim 1.8$).

As we explained in Section 1.5, X-rays provide a better tool to detect obscured AGN than optical/NIR observations, given that X-ray emission is less biased towards obscuration. Obscuration can be estimated in the X-rays using different techniques, such as spectral analysis (e.g., Lanzuisi et al. 2013a), hardness ratio analysis (i.e., the ratio between counts detected in different bands, e.g., Brusa et al. 2009) and X-ray colors (Iwasawa et al. 2012). Optical Type 2 AGN can instead be classified through optical spectroscopy (e.g., Mignoli et al. 2013), color-color analysis (e.g., Donley et al. 2012) and line diagnostic diagrams (e.g., Bongiorno et al. 2012). It is worth noticing that, optical and X-ray classifications can lead to different estimations of obscuration for the same source (e.g., Merloni et al. 2014).

Observational evidence of a decrease with luminosity of the fraction of obscured AGN at $z < 3$ has been found in many works in both the X-ray (Ueda et al. 2003; Steffen et al. 2003; Barger & Cowie 2005; La Franca et al. 2005; Akylas et al. 2006; Treister & Urry 2006; Della Ceca et al. 2008; Hasinger 2008; Ebrero et al. 2009; Treister et al. 2009; Brusa et al. 2010; Burlon et al. 2011; Sazonov et al. 2012; Ueda et al. 2014; Buchner et al. 2015) and in the optical/IR (Simpson 2005; Treister et al. 2008; Bongiorno et al. 2010; Assef et al. 2013; Lusso et al. 2013). A physical explanation to this effect is that the covering factor of the obscuring material gets smaller in more luminous AGN (e.g. Maiolino et al. 2007; Mor & Trakhtenbrot 2011). A similar trend can be related to the so-called “receding torus scenario” (Lawrence 1991; Nenkova et al. 2008), where the torus height is the same at each radius and does not change significantly with the AGN luminosity. In this model, the anti-correlation between covering factor and luminosity is caused by the increase of the dust sublimation radius at increasing luminosities. A similar physical behaviour could also explain another observational effect, the so-called “Iwasawa-Taniguchi” effect (Iwasawa & Taniguchi 1993; Page et al. 2004; Guainazzi et al. 2006), i.e., the decreasing EW of the Fe $K\alpha$ line with the AGN luminosity. A similar model can however explain the observed anti-correlation in the optical/IR, while the X-ray are not sensitive to dust obscuration. Therefore, a different model was proposed by Lamastra et al. (2006), which assumed that the Compton-thin obscuration (i.e., $10^{22} \leq N_H < 10^{24} \text{ cm}^{-2}$) is caused by the interstellar medium in the host galaxy. In a similar scenario, the X-ray anti-correlation is explained with the fact that the gravitational pull is larger in SMBHs with larger mass (and, consequently, with larger luminosity, on statistical basis). Another model links the anti-correlation to massive outflows of material caused by the AGN radiation pressure (as observed in

several works, e.g., Fischer et al. 2010; Feruglio et al. 2010; Rupke & Veilleux 2011; Sturm et al. 2011).

Finally, while many different works proposed physical explanation for the decrease with luminosity of the fraction of obscured AGN, other works explained the trend with incompleteness and/or selection effects (Dwelly & Page 2006; Lawrence & Elvis 2010; Merloni et al. 2014).

1.7.2 $z > 3$ AGN space density as a SMBH early growth estimator

For a complete analysis of the way SMBH and galaxies evolve and co-evolve before their density peak, large samples of AGN at both high redshifts and low luminosities are required. The rest-frame 2-10 keV comoving space density at $z \geq 3$ can be used to place constraints on the BH formation scenario. In fact, the shape of the space density is linked to the time-scale of accretion of SMBHs and is therefore a tool to investigate the SMBH formation and growth scenario, eventually distinguishing between major-merger driven accretion and secular accretion.

Several optical surveys have already computed the space density and the luminosity function of high- z AGN (e.g., Richards et al. 2006; Willott et al. 2010; Glikman et al. 2011; Masters et al. 2012; McGreer et al. 2013; Ross et al. 2013; see also Figure 1.10); however, all these works were limited to high luminosities ($-27.5 < M_{AB} < -25.5$) at $z > 3$ and therefore present large uncertainties in their faint end. At lower optical luminosities, the AGN selection is less reliable and low-luminosity optical surveys have so far produced disagreeing AGN luminosity functions (see Ikeda et al. 2011; Glikman et al. 2011). Moreover, optical surveys are biased against obscured sources, whose contribution becomes also more significant at low luminosities. X-ray selection provides a solution to overcome these issues.

In the last ten years several X-ray surveys (in the 2-10 keV band) were sensitive enough to investigate this redshift range. Two pioneering works were performed in the COSMOS field, using XMM-*Newton* (Brusa et al. 2009, $N_{AGN}=40$), and *Chandra*, on the central 0.9 deg² (C-COSMOS, Elvis et al. 2009, Civano et al. 2012) (Civano et al. 2011, $N_{AGN}=81$), reaching a luminosity limit of $L_{2-10keV}=10^{44.2}$ erg s⁻¹ and $L_{2-10keV}=10^{43.55}$ erg s⁻¹, respectively. Vito et al. (2013, $N_{AGN}=34$) were able to extend their analysis down to $L_{2-10keV} \simeq 10^{43}$ erg s⁻¹, using the 4 Ms *Chandra* Deep Field South (CDF-S, Xue et al. 2011) catalog; the same group (Vito et al. 2014) studied the 2-10 keV luminosity function in the redshift range $z=[3-5]$, combining deep and shallow surveys ($N_{AGN}=141$). Kalfountzou et al. (2014) combined the C-COSMOS sample with the one from the wide and shallow ChaMP survey (Kim et al. 2007; Green et al. 2009; Trichas et al. 2012) to have a sample of $N_{AGN}=211$ at $z > 3$ and $N_{AGN}=27$ at $z > 4$, down to a luminosity $L_{2-10keV}=10^{43.55}$ erg s⁻¹. Finally, Georgakakis et al. (2015) combined data from different surveys to obtain a sample of 340 sources at $z > 3$ over about three orders of magnitude, $L_{2-10keV} \simeq [10^{43}-10^{46}]$ erg s⁻¹. All these works show a decline of the AGN space density at $z > 3$, but they are not able to put better constraints at $z > 4$, due to the lack of good statistics. Moreover, when combining different surveys one has to assume a completeness correction, therefore introducing uncertainties in the final result.

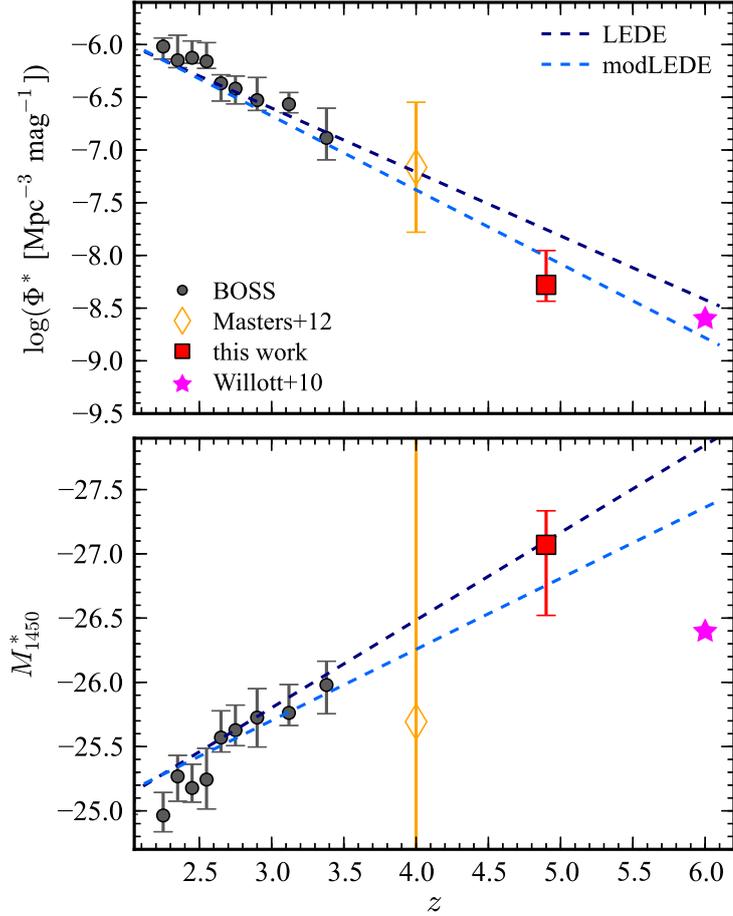


Figure 1.10: Evolution with redshift of the quasar luminosity function (QLF) normalization (Φ^*) and break luminosity (M_{1450}^*). Data from the BOSS DR9 QLF (Ross et al. 2013, black diamonds), McGreer et al. (2013, red square), Masters et al. (2012, yellow diamond) and Willott et al. (2010, magenta star). Two different fits to the BOSS data are shown as dark and light blue lines. Taken from McGreer et al. (2013).

In Chapter 4 we show the $z > 3$ space density from the *Chandra COSMOS-Legacy* Survey, and we compare it with the predictions of both phenomenological models of AGN evolution and physical models of quasar activation through major merger.

Black hole seeds

The analysis of statistically significant samples of high redshift AGN can also provide better constraints on different models of SMBH seed formation.

Several recent works show evidence of optically selected SMBH with mass $> 10^9 M_\odot$ at $z > 6$ (Willott et al. 2003; Fan et al. 2006a,b; Willott et al. 2009; Mortlock et al. 2011), i.e., less than 1 Gyr after the Big Bang. The existence of these objects require both high accretion rates (even in super-Eddington regime) and high-mass BH seeds, i.e., objects with $M \sim 10^{3-5} M_\odot$. Similar seeds are usually linked to the

direct collapse of massive gas clouds of dense gas (Haehnelt & Rees 1993; Loeb & Rasio 1994; Eisenstein & Loeb 1995; Bromm & Loeb 2003; Koushiappas et al. 2004; Begelman et al. 2006; Lodato & Natarajan 2006). Dense gas clouds in the inner part of the first proto-galaxies are ideal massive BH seeds candidates, because of their high density and very low metallicity: in low-metallicity environments cooling processes are less effective, the gas cloud fragmentation is less likely and all the gas can contribute to the BH seed formation. Another model adopted to explain massive BH seeds with $M \sim 10^{2-4} M_{\odot}$ is based on stellar-dynamical processes instead of gas-dynamical processes (Devecchi & Volonteri 2009). In this scenario, stellar collisions in compact star clusters (Schneider et al. 2006; Clark et al. 2008) can produce “very massive stars” (VMS) with $M > 100 M_{\odot}$, eventually leaving a massive BH remnant.

A second class of seeds, low-mass seeds with $M \sim 10^2 M_{\odot}$, are associated to remnants of population III stars (e.g., Madau & Rees 2001; Volonteri & Begelman 2010). A similar class of objects can have, as a byproduct, SMBH with intermediate mass (i.e., $M_{BH} = 10^{6-8} M_{\odot}$) at $z \leq 3$ and bolometric luminosity $L_{bol} \sim 10^{46} \text{ erg s}^{-1}$. Similar AGN cannot be easily selected using optical information, but can instead be observed in the X-rays (see Section 1.5).

We show a visual summary of different BH seeds formation processes in Figure 1.11 (from Volonteri 2012).

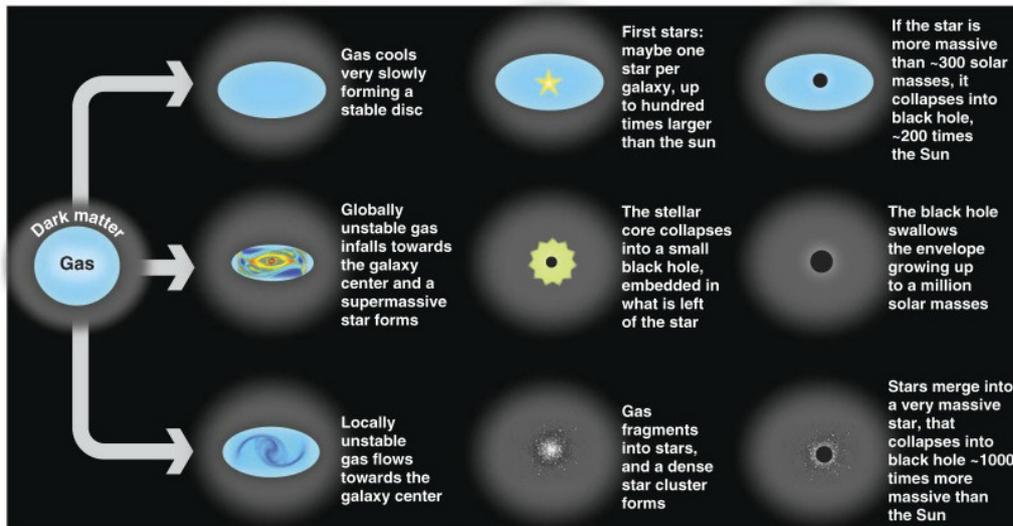


Figure 1.11: Summary of different models of BH seed formation. Taken from Volonteri (2012)

Chapter 2

The *Chandra* COSMOS Legacy survey: overview and point source catalog

In this chapter we present the *Chandra COSMOS-Legacy* survey¹: the results of these chapter are also presented in Civano et al. (submitted). The *Chandra COSMOS-Legacy* survey is the combination of the old C-COSMOS survey (1.8 Ms) with 2.8 Ms of new *Chandra* ACIS-I (Garmire et al. 2003) observations (56×50 ks pointings) approved during *Chandra* Cycle 14 as an X-ray Visionary Project (PI: F. Civano; program ID 901037). *COSMOS-Legacy* uniformly covers the ~ 1.7 deg² COSMOS/*HST* field at ~ 160 ks depth, expanding on the deep C-COSMOS area (dashed green square in Figure 2.1) by a factor of ~ 3 at $\sim 3 \times 10^{-16}$ erg cm⁻² s⁻¹ (1.45 vs 0.44 deg²), for a total area covered of ~ 2.2 deg².

An extended description of the multiwavelength identification of the X-ray sources will be presented in chapter 3.

In Section 2.1, we present the observations and tiling strategy. In Section 2.2 we detail all the steps of the data processing, including astrometric corrections, exposure and background map production. The data analysis procedure is instead described in Section 2.3, with some references and comparison with the one adopted for C-COSMOS as explained in P09. The point source catalog and the source properties are presented in Section 2.3.1, while in Sections 2.4 and 2.5, the survey sensitivity and the number counts in both soft and hard band, and also dividing the sources in obscured and unobscured, are presented. Finally, in Section 2.7 we show the results reported in Lanzuisi et al. (2013b) about the XMM-*Newton* spectral analysis of CID-42, a candidate recoiling SMBH detected by *Chandra COSMOS-Legacy*.

We assume a cosmology with $H_0 = 70$ km s⁻¹ Mpc⁻¹, $\Omega_M = 0.29$ and $\Omega_\Lambda = 0.71$; magnitudes are reported in the AB system if not otherwise stated. Throughout this work, we make use of J2000.0 coordinates. The data analysis is performed in three X-ray bandpasses 0.5–2 keV (soft band, S), 2–7 keV (hard band, H), and 0.5–7 keV

¹Throughout the rest of the work we use the term C-COSMOS to refer to the original survey of the inner field, and the name *Chandra COSMOS-Legacy* survey to refer to the full, combined survey, including the new data presented here.

(full band, F), while sensitivity and fluxes have been computed in the 0.5–2, 2–10 and 0.5–10 keV bands for an easy comparison with other works in the literature.

2.1 Observations

The half-a-field shift tiling strategy was designed in order to cover uniformly, in depth and point spread function size, the COSMOS *Hubble* area (cyan outline in Fig. 2.1; Scoville et al. 2007b), by combining the old C-COSMOS observations (green outline in Fig. 2.1) with the new *Chandra* ones (red outline in Fig. 2.1). To achieve this, 56 ACIS-I pointings (numbered black points in Fig. 2.1) were used, 11 of which were scheduled as two or more separate observations because of satellite constraints, for a total of 68 pointings. Moreover, the observing roll angle was constrained to be within 70 ± 20 or 250 ± 20 degrees. The main properties of the new *Chandra* COSMOS Legacy observations are shown in Appendix A.

The observations took place in four blocks: November, 2012 to January, 2013; March to July, 2013; October, 2013 to January, 2014; and March, 2014. The mean net effective exposure time per field was 48.8 ks, after all the cleaning and reduction operations (see Section 2.2). The maximum exposure was 53 ks (observation 15227) while the minimum exposure was 45.2 ks (combined observations 15208 and 15998).

The sequence of the observations was designed to start from the N-E top corner tile of C-COSMOS moving towards W and proceeding clockwise around the central C-COSMOS area, in such a way that the outer frame of the C-COSMOS survey overlaps with the inner frame of the new *Chandra* observations. The tiling number and the total area covered is shown in Figure 2.1.

Using this tiling strategy we achieve an approximately uniform combined point spread function (PSF) across the survey. The mean combined PSF width (size at 50% of the encircled energy fraction, EEf, in the 0.5-7 keV band; see Section 2.4 for details on the PSF maps), weighted on the exposure, peaks at around $3''$ (see Figure 2.2). As shown in Figure 2.2, 80% of the field has a PSF in the range $2''$ - $4''$. As a comparison, in a single-pointed survey (regardless of exposure time), the PSF size distribution has a larger spread, and although $\sim 30\%$ of the field has a PSF $< 2''$ the PSF can reach a substantially larger size ($> 4''$) in 40% of the field.

2.2 Data processing

The data reduction was performed following the procedures described in E09 for C-COSMOS, using standard *Chandra* CIAO 4.5 tools (Fruscione et al. 2006) and CALDB 4.5.9. We also reprocessed the 49 C-COSMOS observations in order to use them in concert with the new observations for source detection in the area where the new observations overlap with the old ones and to compute the sensitivity of the whole survey (see the comparison between fluxes in Section 2.3.1).

We used the `chandra_repro` reprocessing script, which automates the CIAO recommended data processing steps and creates new level 2 event files, applying the *VFAINT* mode for ACIS background cleaning to all the observations. We then performed the following steps before starting data analysis: astrometric correction and reprocessing of all the observations to a standard frame of reference using the

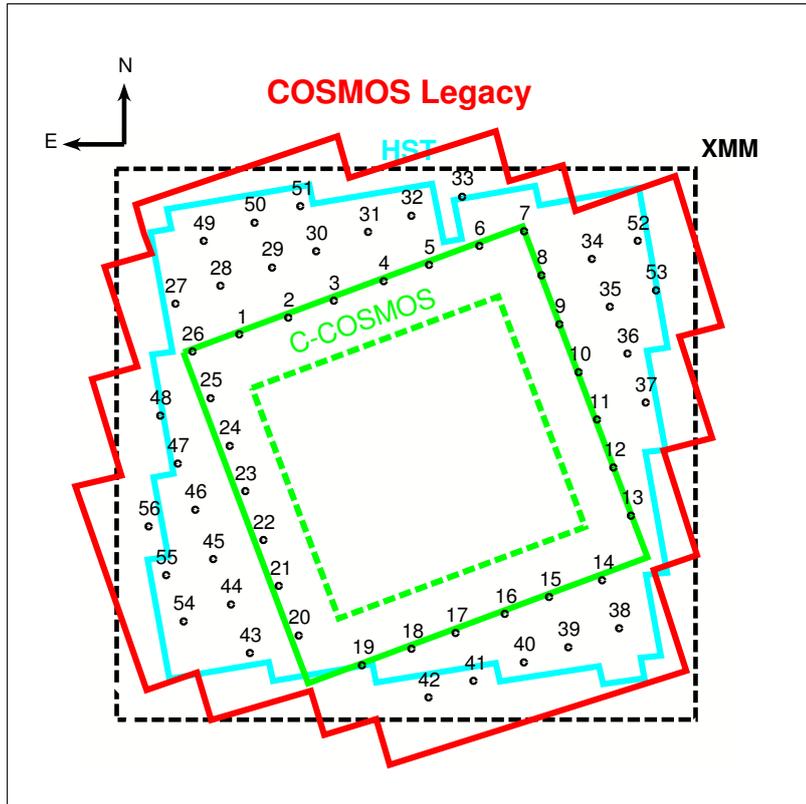


Figure 2.1: *COSMOS-Legacy* tiling (red) compared to the area covered by HST (cyan), C-COSMOS (green solid: total area; green dashed: deeper area) and XMM-COSMOS (black). The ordering numbers of new observations are marked (see the Appendix A for a list of all the observations.)

new aspect solution (Section 2.2.1); mosaic and exposure map creation in three standard *Chandra* bands (Section 2.2.2): 0.5-7 keV, 0.5-2 keV and 2-7 keV; background map creation, using a two-components model to take into account both the cosmic background contribution and the instrumental one (Section 2.2.3).

2.2.1 Astrometric corrections

Even though *Chandra* data astrometry is accurate to $0.6''$ (at 90% confidence, see Proposer User Guide² Chapter 5), in order to produce a sharp X-ray mosaic and to match the positions of X-ray sources with the optical catalog for which the positional accuracy is $\sim 0.2''$ (Capak et al. 2007, Ilbert et al. 2009, Laigle et al. submitted), we performed source detection on each individual observation to register them to a common optical astrometric frame. This work has been done on the new observations and also on the C-COSMOS outer frame fields overlapping with the new data.

We generated a list of detected sources using the CIAO wavelet source detection tool *WAVDETECT* on each single observation binned at $1''$ and adopted a false-

²http://cxc.cfa.harvard.edu/proposer/POG/html/chap5.html#tth_fig5.5

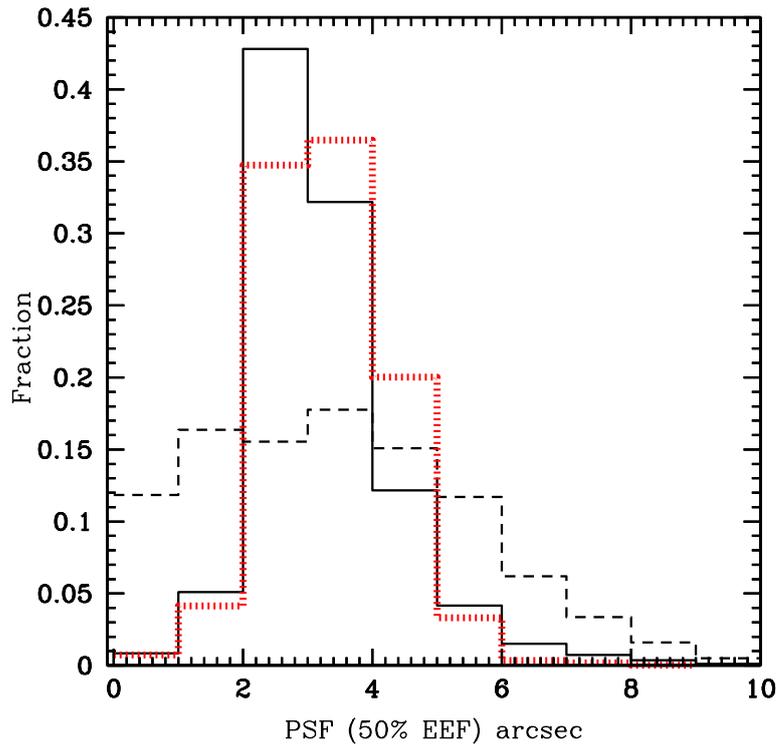


Figure 2.2: Normalized distribution of the combined point spread function (50% of the EEf in the 0.5-7 keV band) size in arcseconds measured in *COSMOS-Legacy* (solid histogram) and in a single pointing survey (dashed line). In red, the distribution of the combined PSF (the mean value) for all the detected sources.

positive detection probability threshold corresponding to ~ 10 spurious sources per field. Of the detected sources (on average 150 sources per field), we considered in each field those with significance $> 3.5\sigma$ and within $360''$ from the aim point. In *Chandra* data, the positional accuracy of significant sources is $< 1''$ even at $10'$ off-axis and it is energy independent (K. Glotfelty, private communication). Therefore, choosing sources within $6'$ of the aim-point provides a sample of sources with very good centroid estimate ($< 0.3''$) for astrometric purposes. Using the CIAO tool `reproject_aspect`, these sources were then compared to the CFHT *MegaCam* catalog of i-band selected sources (McCracken et al. 2012) with optical AB magnitudes in the range 18-23. At least 4 sources in each field, not on the same side of the aim-point, are needed to compute meaningful rotational and translation transformations. In our analysis, we used on average 12 sources (up to 22 sources) per field, with 75% of the fields having more than 10 sources used to perform the reprojection.

With the corrected aspect solution, we reprocessed the level 1 data using `chandra_repro` and performed the `WAVDETECT` detection again to compute the new separation between X-ray and optical positions. The resulting standard deviation on the shift computed from the detected sources within $6'$ is $0.36''$ and $0.51''$ on the RA and Dec, respectively. After matching all the X-ray fields to the same astrometric op-

tical frame, 95% of the X-ray sources used for the astrometry correction have a distance to their optical counterpart smaller than $1.4''$, 10% lower than the value before the correction ($1.53''$). The improvement in the position increases to 20% when considering 90% of the sources ($1.26''$ to $1.02''$) and 30% when considering a smaller sample of 68% of the sources (from $0.72''$ to $0.51''$; see Figure 2.3). This is consistent with, and slightly better than, what was found for C-COSMOS (see E09, Figure 6).

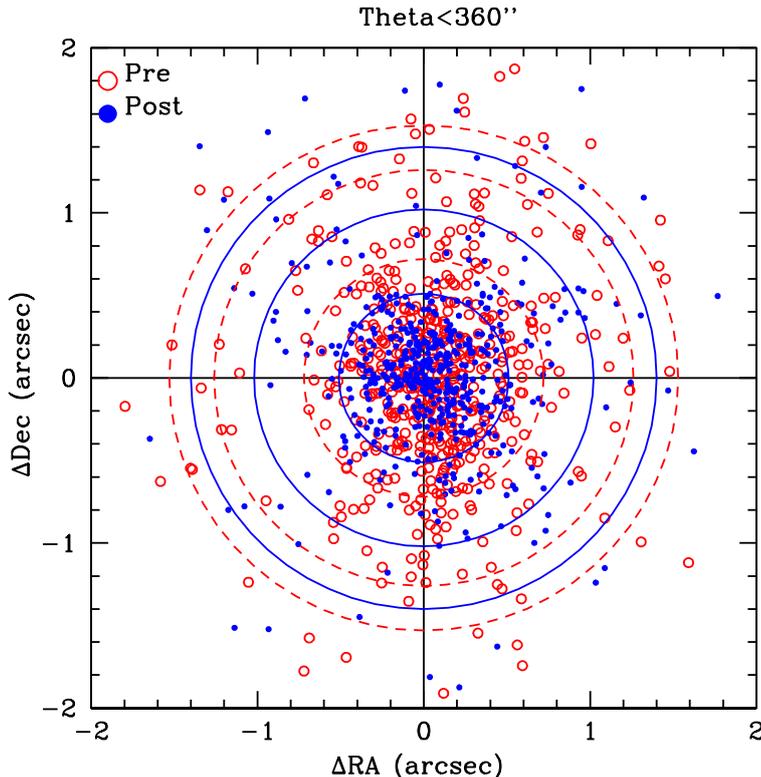


Figure 2.3: The X-ray to I-band separation (ΔRA , ΔDec) in arcsecond for the X-ray sources within $6'$ from the aim point detected in each single observations before (red open circles) and after (blue solid circles) the aspect correction. The circles encompass 68%, 90% and 95% of the sources before (red dashed) and after (blue solid) the correction.

2.2.2 Exposure maps and data mosaic creation

We created exposure maps in three bands using the standard CIAO procedure. The spectral model used for the map creation is a single power-law with slope $\Gamma=1.4$ and Galactic absorption ($N_H=2.6 \times 10^{20} \text{ cm}^{-2}$; Kalberla et al. 2005). The choice of a spectral slope $\Gamma=1.4$ is not only because of consistency with E09 and P09, but it is also because this slope is the same of the cosmic X-ray background (e.g., Hickox & Markevitch 2006) and therefore well represents a mixed distribution of obscured and unobscured sources at the fluxes covered by *COSMOS-Legacy*. Instrument maps, generated with MKINSTMAP for each CCD in each observation, were used as

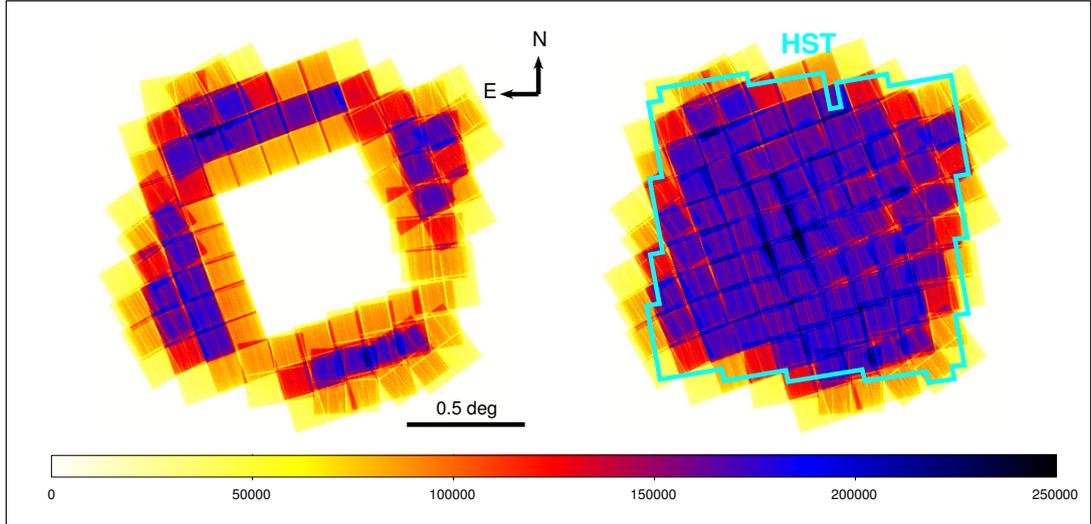


Figure 2.4: The mosaic of exposure maps for the new observations (left) and for the whole *COSMOS-Legacy* survey (right) in the full band. The color bar gives the achieved effective exposure in units of seconds. We reached a uniform coverage of ~ 160 ks over the full HST area (cyan polygon).

input files for the MKEXPMAP tool, which computes an exposure map for each CCD separately. These exposure maps were combined in a single exposure map for each observation using DMREGRID with a binning of 2 pixels.

Figure 2.4 shows a composite image of the effective exposure time (in seconds) in the full band for both the new observations (left) and the whole *COSMOS-Legacy* (right). As can be seen, the central 1.5 deg^2 , covering almost entirely the HST area, have a uniform depth of $\simeq 160$ ks.

The data mosaic image was created in three bands using the HEASoft `addimages` tool, which adds together a set of images using sky coordinates. In Figure 2.5, the three color image, created by combining the exposure corrected images in three non-overlapping bands (0.5-2.0 keV, 2.0-4.5 keV, and 4.5-7.0 keV as red green and blue, respectively) is shown. The combined image was then Gaussian smoothed with a 3 pixel radius. A filter was then applied to isolate sources from the background level, as well as to increase the contrast and color vibrancy of those sources. This process was repeated 3 times.

2.2.3 Background maps creation

The *Chandra* background consists of two different components: the cosmic X-ray background and a quiescent instrumental background due to interactions between the ACIS-I CCD detectors and high-energy particles. We followed the procedure described in Cappelluti et al. (2013) to create background maps, which we used for the selection of reliable sources in our detection procedure and for the computation of the sensitivity curves.

The background maps were computed for each observation separately in the full, soft and hard bands. We ran WAVDETECT with a threshold parameter $\text{sigthresh}=10^{-5}$,

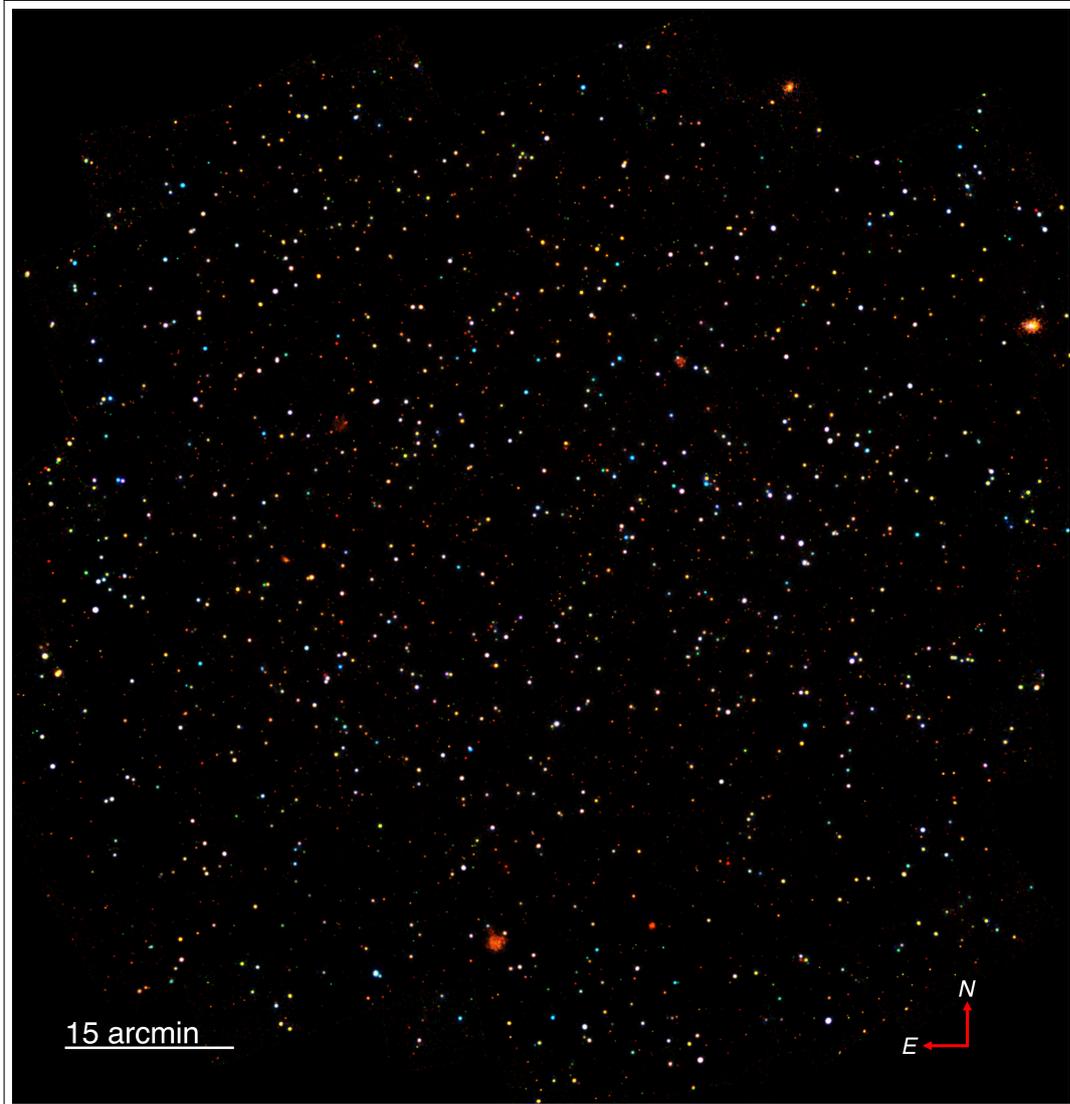


Figure 2.5: Three color image of the whole *COSMOS-Legacy* field (0.5-2.0 keV, 2.0-4.5 keV, and 4.5-7.0 keV as red green and blue, respectively).

corresponding to ~ 100 spurious sources per field (see Section 2.2.1), large enough to select also sources with significant signal only in stacked emission. We then removed these sources from the science images by excising a region corresponding to the source size (using a 3σ value) as computed by the detection tool. We then uniformly distributed the remaining counts, rescaled by the ratio between the whole area of the observation and the area without the removed sources. These files were then used as initial background.

We then downloaded “stowed background” data from the *Chandra* archive³. Stowed background files are particle-only background files and are obtained when the ACIS detector is out of the focal plane. These files were then rescaled using the procedure described in Hickox & Markevitch (2006): we measured the ratio between the number of counts in our initial background (C_{data}) and in the stowed image (C_{stow}) in the energy range 9.5-12 keV. In this band, the effective area of *Chandra* is $\simeq 0$ and consequently all the counts have a non-astrophysical origin.

The stowed background, rescaled to our data by C_{data}/C_{stow} , was then subtracted from the initial background to obtain a first version of the cosmic X-ray background. The counts of this map were then renormalized using the exposure maps to create an exposure-corrected cosmic background.

Finally, we performed a Monte Carlo simulation using the exposure-corrected cosmic X-ray background and the stowed background as input files. We simulated 1000 images for each of the two backgrounds using the IDL routine *poidev* to obtain a Poissonian realization of each map, and then we obtained our final homogeneous background map adding together the two mean simulated images. In order to use these maps for sensitivity computations and in our detection algorithm, a Gaussian smoothing (with a scale of 20 pixels) was applied to this final background map using the FTOOL `fgauss`.

The distribution of the computed background (in counts/arcsec²) in the three bands is reported in Figure 2.6. The overall background count distribution is consistent with the one found in C-COSMOS (see Figure 4 of P09). In the full band the main peak is at around 0.13 counts/arcsec² and this corresponds to the deepest part of the exposure. In C-COSMOS, the deep and shallow areas were roughly of the same size and therefore the background distribution had two clear peaks of approximately the same height, while in *COSMOS-Legacy*, the area with higher exposure is 3 times larger than the shallow area. This is represented in the background distribution as well. The number of background counts is consistent with the expectation for *Chandra* given the distribution of our exposure times.

³<http://cxc.harvard.edu/ciao/threads/acisbackground/>

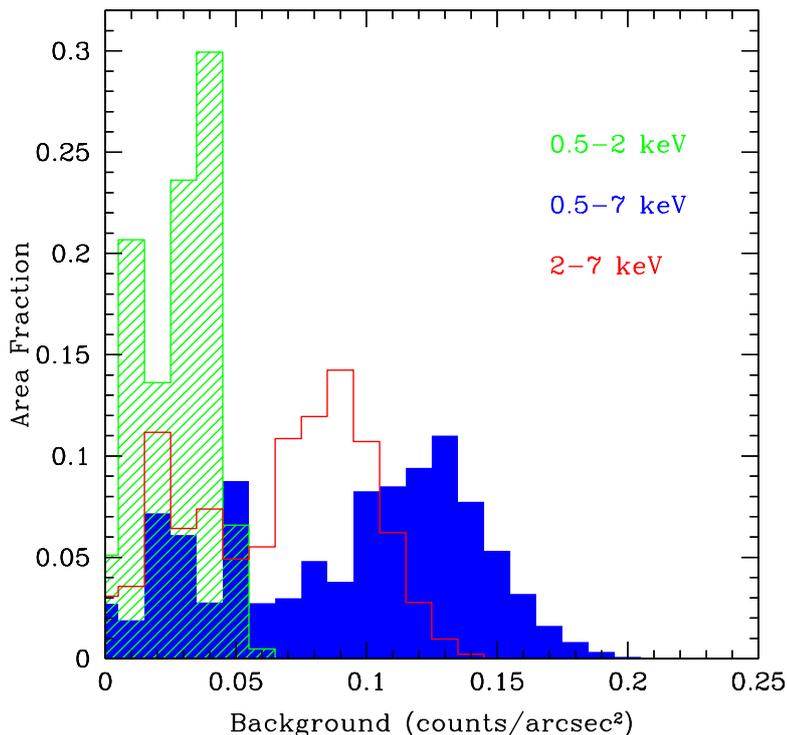


Figure 2.6: Distributions of background counts per square arcsecond in the full (solid blue histogram), soft (shaded green histogram) and hard (empty red histogram) bands.

2.3 Data Analysis: source detection and photometry

The analysis presented in the following focuses only on point sources. A parallel effort on the detection of extended sources will be presented by Finoguenov et al. (in preparation). To avoid contamination by extended sources, we used the XMM-COSMOS catalog of extended sources (Finoguenov et al. 2007, Kettula et al. 2013) and visually inspected all the brightest ($L_X > 10^{41}$ erg s $^{-1}$ in 0.5-2 keV) ones to check if a point source is detected inside them by *Chandra*.

Puccetti et al. (2009) extensively discussed and compared different source detection techniques concluding that the best procedure for C-COSMOS was a combination of `PWDetect` (Damiani et al. 1997) and the `Chandra Emldetect` (`CMLDetect`) Maximum Likelihood algorithm. As shown by P09 using extensive simulations, one of the strongest features of `PWDetect` is its ability to locate X-ray sources with extreme accuracy ($0.02'' \pm 0.15''$, P09 Table 1), while `CMLDetect` is the best tool to perform source photometry and derived source significance. The *COSMOS-Legacy* survey shares the same tiling layout, exposure time per field and roll angle range of C-COSMOS, hence, we can follow the P09 procedure and use the same significance threshold for source detection.

The original version of `CMLDetect`, called `emldetect` (Crudace et al. 1988, Hasinger et al. 1993), is part of the *XMM-Newton* SAS package and is based on a

code originally developed for *ROSAT* data. `CMLDetect` has been adapted to run on *Chandra* data by replacing the *XMM-Newton* PSF library with the *Chandra* one (see Krumpel et al. 2015 for another application of `CMLDetect`). Moreover, this new tool can also work with different PSFs simultaneously.

`PWDetect` was developed to properly treat *Chandra* data with PSF varying across the field and is based on the wavelet transform (WT) of the X-ray image. A WT is the convolution of an image with a “generating wavelet” kernel which depends on position and length scale (a free parameter). For this survey, and for *Chandra* data in general, the length scale varies from $0.5''$ to $16''$ in steps of $\sqrt{2}$. These steps cover all possible *Chandra* PSFs (the largest are those at large off axis angle θ_i). Both radial and azimuthal PSF variations are accounted for by `PWDetect`, which first assumes a Gaussian PSF and then corrects by a PSF shape factor, calibrated with respect to source positions on the CCD.

`PWDetect` works on stacked observations only if co-aligned (same aim point and roll-angle), as is the case for 11 of our fields which are observations split into multiple parts. Therefore, `PWDetect` was run on each of our new 56 fields setting the detection limit to 3.8σ corresponding to a probability of a spurious detection to $\simeq 10^{-4}$ with the aim of creating a large catalog of detections to be fed to `CMLDetect`. Also, given that the outer frame of C-COSMOS overlaps with the new survey, we run `PWDetect` on 20 old fields (fields 1-1 to 1-6, 1-6 to 6-6, 6-6 to 6-1 and last 6-1 to 1-1 as in Table 3 of E09). For overlapping regions between different pointings, we performed a positional cross correlation (using a $2''$ radius) and if a source was detected in more than one field, we chose the position of the source at the smallest θ_i , i.e. the one with the best PSF. We performed a visual inspection of all the sources having multiple matches within $5''$. About 90% of the pairs in the range $2-5''$ were actually false detections, mainly caused by PSF tail detection of bright sources.

The positions obtained with `PWDetect` were then fed as input to `CMLDetect`, to obtain photometric information and significance for each source. `PWDetect` can be used to obtain net counts, rates and fluxes, but we opted to use `CMLDetect` because it can work on a mosaic, while `PWDetect` cannot. Moreover, P09 has shown that `PWDetect` count rates are systematically less accurate than those of `CMLDetect` (the median ratio between the output detected and input simulated count rates ranges from 86 to 94% for `PWDetect` versus 97 to 105% for `CMLDetect`, independently from the photon energy). `CMLDetect` performs a simultaneous maximum likelihood PSF fitting for each input candidate source, previously obtained using `PWDetect`, to all images at each position and, working on a mosaic, can provide a refined position of the source and count rates. This procedure was run in three bands: full (0.5-7 keV), soft (0.5-2 keV) and hard (2-7 keV). With the goal of not missing close pairs, we run `CMLDetect` allowing to slightly change the input position provided by `PWDetect`. We also run `CMLDetect` allowing only the detection of point-like sources, without taking in account those extended.

The best-fit maximum likelihood parameter in `CMLDetect`, `DET_ML`, is related to the Poisson probability that a source candidate is a random fluctuation of the background (P_{random}), as follows:

$$DET_ML = -\ln(P_{random}). \quad (2.1)$$

As a consequence, sources with small values of `DET_ML` have high values of P_{random} and are then likely to be background fluctuations. We chose a threshold significance value of 2×10^{-5} , that corresponds to `DET_ML`=10.8 (i.e., a source needs `DET_ML`>10.8 in at least one of the three bands to be included in the final catalog). This value is the same used in C-COSMOS and represents the best compromise between completeness and reliability as shown by P09 in Figure 11 and 12. 75% of the sources detected by `PWDetect` in a single field with `DET_ML`>10.8 and fed to `CMLDetect`, were found to be above the threshold in output.

To improve the final completeness of the catalog, we also search for less significant sources, up to about 100 times higher P , which corresponds to a threshold `DET_ML`=6. This lower threshold catalog, similarly to what was done in C-COSMOS, is only used for sources already detected with `DET_ML`>10.8 in another band.

As determined by P09, the chosen `DET_ML` threshold implies a completeness of 87.5% and 68% for sources with at least 12 and 7 full band counts, of 98.2% and 83% for the soft band, 86% and 67% for the hard band. At this significance level and the same count limits, the reliability is $\sim 99.7\%$ for the three bands.

2.3.1 Point source catalog

Source numbers

We positionally matched the three single-band, `CMLDetect` output catalogs (including all the sources to `DET_ML`=6) to each other using a cross-correlation radius of $3''$. We first matched the full band detected source catalog to the soft band one, then the full with the hard band catalog and finally the soft and hard band one. The mean (median) separation between detections of the same source in two different bands is $0.43''$ ($0.23''$) for full to soft and full to hard, with 90% of the matches within $1''$. For soft to hard matches, the mean (median) separation is instead $0.73''$ ($0.56''$), with $\sim 80\%$ within $1''$. We also performed a visual inspection of the whole sample, and we also make use of the catalog of optical/IR identifications, that will be presented in chapter 3, to solve ambiguous cases. After the visual inspection, we found a total of 20 “fake pairs”, i.e. sources with separation $>3''$ in one combination of bands (5 sources with separation full-hard $>3''$ 6 with separation full-soft $>3''$ 18 with separation soft-hard $>3''$; some sources have separation $>3''$ in more than one combination of bands). All but four of these sources have exposure ≤ 80 ks and are located in the external part of the field of view, therefore having large PSF and significant uncertainty on the position.

In Table 2.1, we report the total number of new sources for each combination of bands, while in Table 2.3 we report the number of sources detected in each band at the two adopted thresholds (`DET_ML`>10.8 and $6 < \text{DET_ML} < 10.8$). The number of detections with `DET_ML`>10.8 in at least one of three X-ray bands is 2273. The number of expected spurious sources with `DET_ML`>10.8 is reported in each band for two count limits in Table 2.2.

In the area where the new data overlap with the outer C-COSMOS frame, the exposure time is now double with respect to the previous mean exposure time (142 ks versus 72 ks), and 385 new sources are detected in addition to the 694 sources already in E09. For the last 694 sources with doubled exposure time, 676 have been

detected in the new data as well. The eighteen C-COSMOS sources not detected in the new data had DET_ML E09 values in the three bands close to the threshold; moreover, 10 of them were detected only in 2 out of 3 bands in E09 and the remaining eight were detected only in 1 band. 13 of these sources have a detection in full band, all with DET_ML<14; 9 have a detection in soft band, all with DET_ML<16.5; 6 have a detection in hard band, all with DET_ML<14. The number of C-COSMOS sources not detected here is $\sim 1\%$ of the total (1761), in acceptable agreement with the number of expected spurious sources (see P09 and also Table 2.2).

In Table 2.1, we include the number of sources in each combination of bands for the C-COSMOS area including the new data and also (in parentheses) the number of sources as in E09. The same old and new numbers are included in Table 2.3.

Among the 676 C-COSMOS sources with new data, only $\sim 10\%$ in each of the three bands have a DET_ML value which is lower in the combined data with respect to the C-COSMOS catalog. Moreover, only $\sim 1.5\%$, $\sim 2\%$ and $\sim 3\%$ of the 676 sources were significant in C-COSMOS, i.e., they had DET_ML>10.8, and now have DET_ML<10.8 in the full, soft and hard band, respectively. These results confirm the reliability of the detection method, and the consistency between the analysis performed in E09 and P09 and the one performed here.

The total number of sources summing the two datasets is reported in the last column of Table 2.1. Adding the new observations, we more than doubled the sample with respect to C-COSMOS, obtaining a catalog of 4016 sources, the largest sample of X-ray sources homogeneously detected and with uniform multiwavelength data (see Section 2.6 and chapter 3 for a complete discussion of the optical/infrared counterparts identification process). In comparison, other contiguous surveys with similar area in the literature have about 20% fewer sources than *COSMOS-Legacy* (see 3362 sources in Stripe 82 by LaMassa et al. 2013a,b; 3293 in X-Bootes by Murray et al. 2005; 2976 in X-DEEP2 by Goulding et al. 2012).

In Figure 2.7, we show the signal-to-noise ratio (SNR = count rate/count rate error) as a function of the DET_ML for the new sources with DET_ML>10.8. In excellent agreement with the finding in C-COSMOS, the SNR increases smoothly with increasing DET_ML, with a dispersion of a factor of 2 at both low and high DET_ML values.

Source positional errors

To compute the positional errors associated with the X-ray centroids given in the catalog ($\sqrt{\sigma_{R.A.}^2 + \sigma_{dec}^2}$), we followed the prescription of P09 defining $err_pos = r_{PSF}/\sqrt{S}$, where S is the number of net (i.e. background subtracted) source counts in the full band in a circular region of radius r_{PSF} containing 50% of the encircled energy in the observation where the source has the smallest off-axis angle. The positional errors are generally in very good agreement with those resulting from *CMLDetect*. In Figure 2.8, the positional error distribution is presented for all the new sources (solid), the old C-COSMOS sources (long dashed) and the updated C-COSMOS distribution (short dashed). The sources plotted in the lowest bin are those with positional errors values smaller than $0.1''$. The peak of the new sources distribution is $\sim 0.6''$ and 85% of the sources have a positional error $< 1''$, while C-COSMOS source distributions peaks at around $0.4''$. This difference (the somewhat

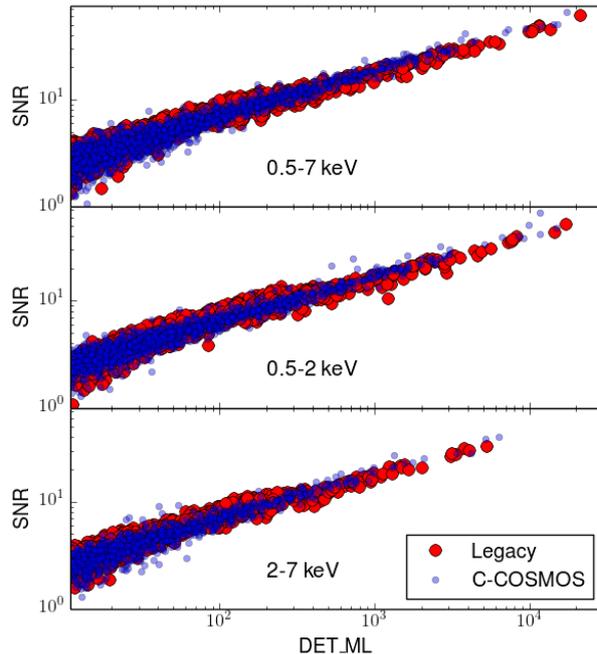


Figure 2.7: Signal-to-noise ratio as a function of DETML for sources detected in three bands. The new *Chandra* sources are plotted as red circles, the C-COSMOS sources as blue ones. We plot only sources with $\text{DETML} > 10.8$.

larger positional errors for the sources detected with the new data than for those detected in C-COSMOS) is due to the fact that, as shown in Fig. 2.9, the net counts distribution for the sources in the new data peaks at a lower value than for the C-COSMOS sources (therefore giving a smaller denominator in the formula of the positional error).

Source counts and fluxes

The count rates in three bands reported here were obtained with *CMLDetect*. Vignetting and quantum efficiency were taken into account when measuring the effective exposure time. The count rate error at 68% confidence level was computed using the equation $err_rate = \frac{\sqrt{C_{S,90\%} + (1+a) \times B_{90\%}}}{0.9 \times T}$, where C_S are the source counts estimated by aperture photometry using, for each observation where the source was detected, an extraction radius including 90% of the EEf; B are the background counts estimated in the same aperture on the background maps used in *CMLDetect* and corrected by a factor $a=0.5$, introduced to account for the uncertainties on the background estimation in a given position (see P09); T is the vignetting corrected exposure time.

In Figure 2.9, the net count distributions for the new sources in three bands are compared to those in E09 (C-COSMOS old) and also to the updated counts distribution of C-COSMOS (C-COSMOS new). The total is the sum of the new

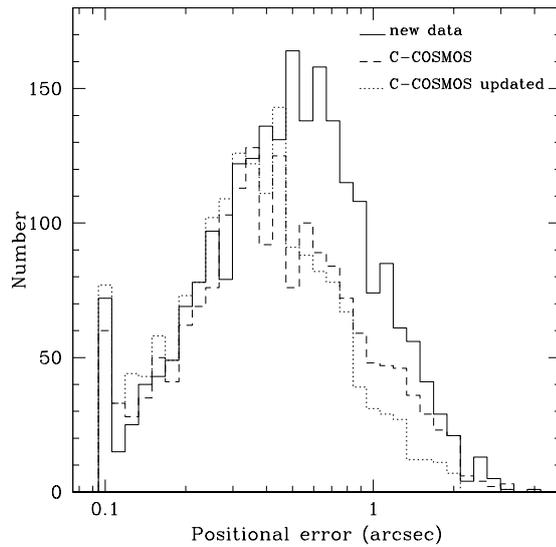


Figure 2.8: Positional error distribution for the new *COSMOS-Legacy* data (solid line), the original C-COSMOS (dashed line) and the updated C-COSMOS (dotted line).

detections plus the updated C-COSMOS. The median (mean) value of net counts in the whole dataset in full, soft and hard bands is 30, 20 and 22 (80, 60, 43), respectively, compared to C-COSMOS where we had 33, 22 and 23 (88, 65, 46). The number of counts for the 676 C-COSMOS sources in the new dataset is on average 60-80% larger than the number of counts in C-COSMOS only. As a consequence, the updated C-COSMOS count histograms in Figure 2.9 are all shifted to a higher numbers of counts. While in the full band the peak of the distribution is still around 30 counts, we more than double the number of sources with more than 70 full band counts, for which it is possible to perform basic X-ray spectral analysis, from 390 (Lanzuisi et al. 2013a) to ~ 950 sources in *COSMOS-Legacy*.

The fluxes were obtained from the count rates using the relation

$$F = R/(CF \times 10^{11}), \quad (2.2)$$

where R is the count rate in each band and CF is the energy conversion factor computed using the online tool PIMMS⁴, assuming a power-law spectrum with slope $\Gamma=1.4$ and a Galactic column density $N_H=2.6 \times 10^{20} \text{ cm}^{-2}$. Due to the fact that the observations have been taken in two different *Chandra* cycles, i.e. Cycle 8 for C-COSMOS and Cycle 14 for the new data, we used as CF a weighted mean of the factors in the different cycles, depending on the exposure time for each source accumulated in each cycle, to take into account for its variation ($\sim 10\%$ between the two cycles). The Cycle 14 (Cycle 8) CF are 0.66 (0.75), 1.63 (1.87) and 0.36 (0.39) counts $\text{erg}^{-1} \text{ cm}^2$ for full, soft and hard bands, respectively. As in C-COSMOS, we computed fluxes in 0.5-10, 0.5-2 and 2-10 keV. The conversion factors are sensitive

⁴<http://cxc.harvard.edu/toolkit/pimms.jsp>

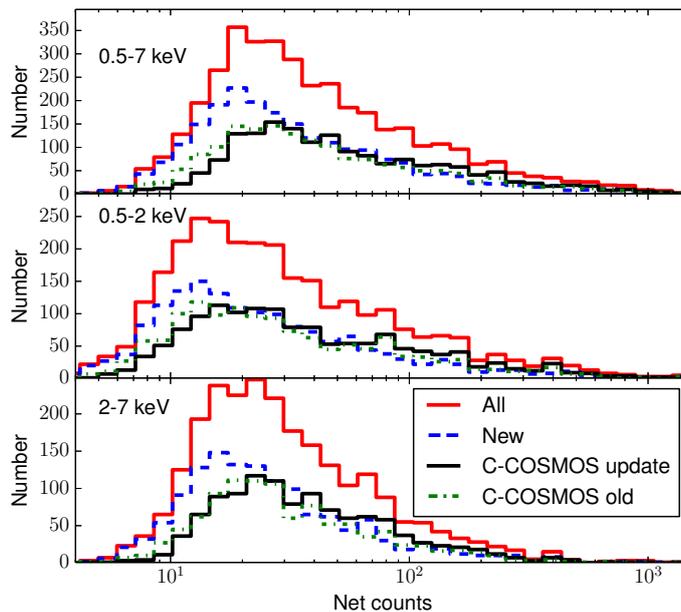


Figure 2.9: Source count distributions in three bands: full (top), soft (center) and hard (bottom) for *COSMOS-Legacy* (solid red), new data only (blue dashed) C-COSMOS old (green dot-dashed) and updated (black solid). Sources with upper limits have not been included.

to the assumed spectral shape: for $\Gamma=2$, there is a change of 40% in the full band CF, of >5% in the soft band and of >20% in the hard band.

For the 676 C-COSMOS sources detected in the new data as well, we computed new X-ray fluxes. In Figure 2.10 the normalized distribution of weighted flux differences between new and old fluxes are plotted for the three bands. From Gaussian fitting of the distributions, we find centroids at $(F_{new} - F_{old}) / \sqrt{F_{new,error}^2 + F_{old,error}^2} = -0.23, -0.24, -0.33$ and standard deviations of ~ 1 for all the bands, showing a remarkable agreement between old and new fluxes. The distributions show wings to both negative and positive values. Malmquist bias is most likely responsible for the negative wing, while variability for the positive one.

The distributions of X-ray fluxes for the whole *Chandra* COSMOS Legacy survey in the full, soft and hard bands is shown in Figure 2.11, where it is also compared with C-COSMOS (the new version with just the updated fluxes, given the excellent agreement) and XMM-COSMOS. The new survey is about ~ 2.5 times deeper than XMM-COSMOS in the 0.5-2 keV band and ~ 2 times in the 2-10 keV band, and more than doubles the number of C-COSMOS sources in the same flux range. In the same Figure we compare our data with the 4 Ms CDFS (Xue et al. 2011) and the large area Stripe82 survey (LaMassa et al. 2013a,b) source flux distributions, respectively to the left and to the right of *COSMOS-Legacy* flux distribution. The combination of the three surveys (the deepest, the intermediate and among the widest, see also Section 2.6) allows us to cover more than 4 orders of magnitude

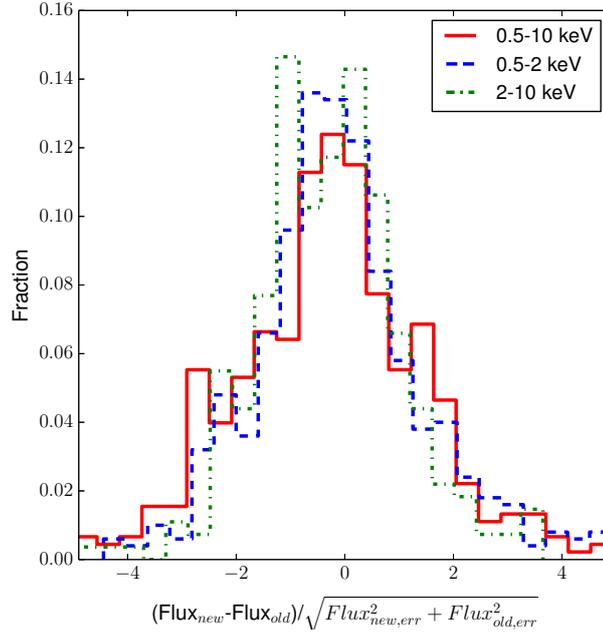


Figure 2.10: Normalized distributions of relative differences between the fluxes weighted by their errors for the 676 sources detected in C-COSMOS and also in the new data at $\text{DET_ML} > 10.8$. Sources with upper limit have not been included.

in flux.

Upper limits (90% confidence level) on net counts, count rates and fluxes are given for all sources found in one band but not detected in another band. The upper limits were computed with the same procedure adopted for C-COSMOS. First we computed T , the total number of counts measured at the source position, in a band where the source was not detected above the detection threshold. The counts were extracted through aperture photometry in the same way we described in section 2.3.1. We then computed B , the expected background counts. Assuming that X are the unknown net counts, we can define the 90% upper limit on X ($X(90\%)$) as the number of counts necessary to observe T (or less) counts with a 10% probability. Using the Poisson probability distribution function is then possible to compute $X(90\%)$, iteratively solving the equation

$$0.1 = e^{-(X+B)} \sum_{i=0}^T \frac{(X+B)^i}{i!} \quad (2.3)$$

for different values of X (see, e.g., Narsky 2000). Equation 2.3.1 does not take in account the expected background counts statistical fluctuations. For this reason, a further correction has to be applied:

$$X_{corr}(90\%) \sim X(90\%) + 1.282\sigma(B), \quad (2.4)$$

computed assuming that the 90% lower limit on B is $B(90\%) = B - 1.282\sigma(B)$. The value 1.282 is an appropriate value for the 90% probability (see, e.g., Bevington &

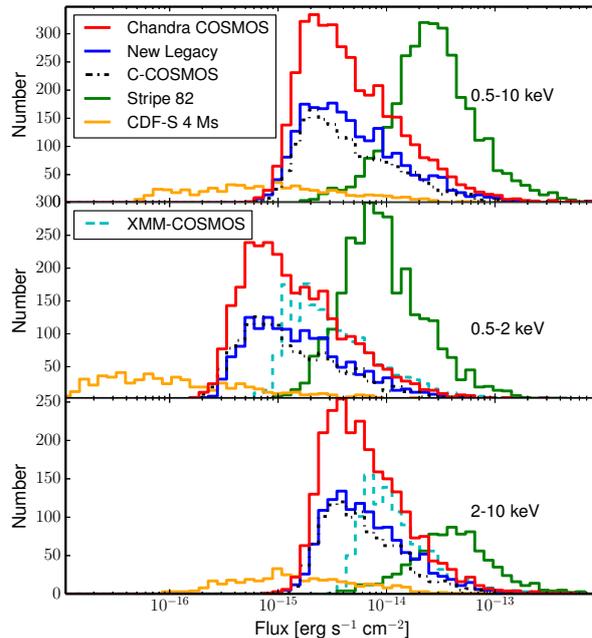


Figure 2.11: Flux distributions for sources detected in full (top), soft (center) and hard (bottom) bands for *COSMOS-Legacy* (solid red), new data only (blue solid), C-COSMOS updated (black dotted) and XMM-COSMOS (cyan dashed) sources. We also include the CDFS 4 Ms source flux distribution (Xue et al. 2011; orange) and the Stripe82 sources (LaMassa et al. 2013a,b; green). Sources with upper limits have not been included.

Robertson 1992). $X_{corr}(90\%)$ is the final net counts upper limit for a source with no detection in a given band. Upper limits on rates and fluxes are then obtained by multiplying $X_{corr}(90\%)$ by the source exposure and, for the fluxes, also by the CF described above.

Hardness ratio analysis

In order to provide a rough estimate of the X-ray spectral shape of the sources, in particular of the intrinsic obscuration (see section 3.6.1 for a more detailed discussion) for all the sources in the catalog, including the C-COSMOS sources, we computed the hardness ratio defined as $HR = \frac{H-S}{H+S}$, where H are the net counts in the hard band and S are those obtained in the soft band. Given the low number of counts for most of the sources (see Figure 2.9), we used BEHR (Bayesian Estimation of Hardness Ratios, Park et al. 2006) which is particularly effective in the low count regime, not needing a detection in both bands to work.

We extracted aperture photometry counts from each observation where the source was detected, using the PSF radius at encircled energy fraction (EEF)=0.9. We also extracted the background counts from the same observations, using an annulus with $r_{min} = r_{PSF} + 8$ pixels and $r_{max} = r_{PSF} + 40$ pixels, where r_{PSF} is the

PSF radius at encircled energy fraction (EEF)=0.95 (in pixel). In the background extraction, we excluded the contamination by other nearby detected sources using an exclusion radius equal to r_{PSF} . Total counts, background counts and the ratio between the sum of background areas and the sum of source areas, both in soft and hard bands, were then fed as input parameters to BEHR.

For most sources (>3000) BEHR finds a detection on the HR, and for 989 sources an upper or lower limit (616 and 371 sources respectively). The typical error on the HR is ~ 0.2 . In Figure 2.12, we plot the distribution of the HRs for the measured values (black solid line), for the lower limits (red) and the upper limits (blue). The mean (median) HR value is -0.09 (-0.17) for the measured values and it moves to lower values when including upper and lower limits (-0.11 and -0.19 for the mean and the median, respectively). A Gaussian fit returns a peak at -0.20 with a 1σ dispersion of 0.32, however a single Gaussian is not clearly a best representation of the HR distribution. A double Gaussian fit (black dotted lines; the single Gaussian components are plotted with a black dashed line) returns a peak at -0.31 and one at 0.12 with a 1σ dispersion of 0.18 and 0.38, respectively.

Hardness ratio is not a fully reliable measurement of obscuration, because of the complexity of the spectral shape, the large error bars due to low counts statistic and the redshift dependency (see section 3.6.1); however it is possible to roughly assume an HR value to divide the sources in obscured and unobscured. We use here $HR = -0.2$, which has been shown to be a fair value to separate sources with column densities above and below 10^{22} cm^{-2} (Lanzuisi et al. 2013a, Civano et al. 2012) at all redshifts. A total of 1993 sources, $50^{+17\%}_{-16\%}$ of the entire sample (errors have been computed using HR 1σ errors) are therefore classified as obscured. Tentatively, the double Gaussian fit of the HR distribution could also be interpreted as to be due from two populations of sources, the obscured population peaking at positive HRs and the unobscured population peaking at negative HR. The broad dispersion of the Gaussian peaking at positive HR could be due to high redshift obscured sources whose HR would be negative even if obscured. A more detailed analysis on the obscured AGN fraction is presented in section 3.6.4.

Source catalog

In Appendix B of this work we show the columns of the final catalog we developed. It contains the new 2273 sources (named as “lid” in column 1) combined with the updated C-COSMOS catalog of 1743 sources (named as “cid” in column 1): for each source we provide all the measurements discussed above.

2.3.2 Matching with XMM-COSMOS catalog

We matched the *COSMOS-Legacy* sources with those in XMM-COSMOS (Capelluti et al. 2009). There are 1714 secure XMM-COSMOS sources with at least one counterpart in *COSMOS-Legacy*, 824 of which have at least one counterpart in the new data. There are 46 XMM-COSMOS sources outside the area covered by *COSMOS-Legacy* (see Fig. 2.1) and 126 with no *Chandra* counterparts. In summary, 93% of the XMM-COSMOS sources within the *COSMOS-Legacy* area have at least one *Chandra* counterpart.

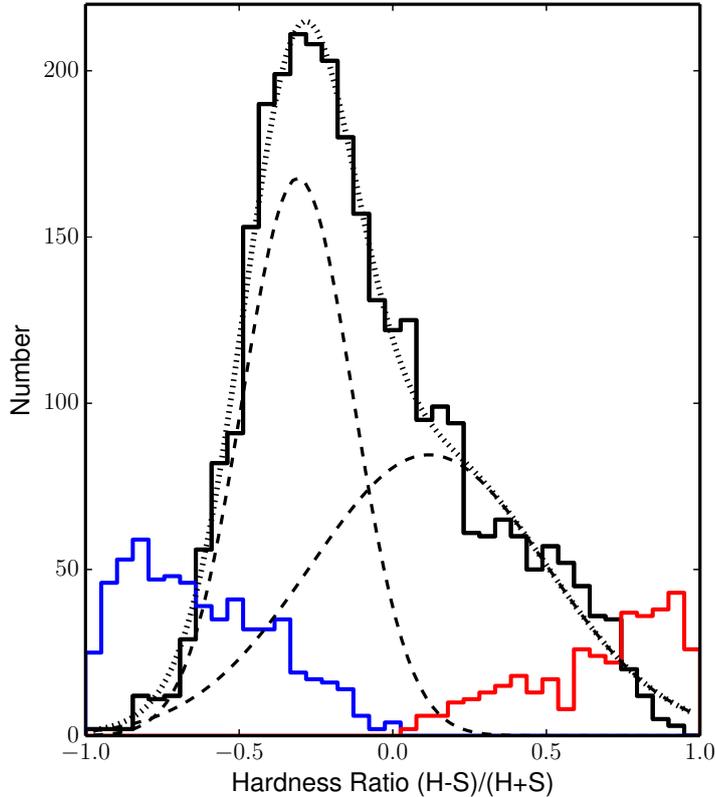


Figure 2.12: HR distributions for the whole sample (black), upper limits (blue) and lower limits (red). The result of a double Gaussian fit (black dotted line) and its two components (black dashed lines) are also shown.

The 126 sources with no *Chandra* counterparts can be divided in three groups: the 25 sources (20%) with *Chandra* exposure <40 ks; the 60 sources (48%; 13 of these sources have also *Chandra* exposure <40 ks) with XMM-COSMOS DET_ML <15 in all of the three bands (0.5-2 keV, 2-8 keV, 4.5-8 keV); last, the 54 sources with XMM-COSMOS DET_ML >15 in at least one band and *Chandra* exposure >40 ks.

For the first group, the low exposure time could be the reason of the non detection, while for the second a non-detection in *Chandra* can be explained with a flux fluctuation within the flux uncertainty. We visually inspected the sources in the last group and we found that seven of them are located inside a bright cluster, and therefore have been not resolved into point sources by our analysis. For the remaining 47 sources the *Chandra* signal is weak or negligible, and therefore these sources could be candidate variable AGN. In particular, XMM-ID 30748 has DET_ML 20 times larger than the detection threshold in XMM-COSMOS: this source was detected only in the 0.5-2 keV band, with a flux $f=2.7\times 10^{-15}$ erg s $^{-1}$ cm $^{-2}$ and a photometric redshift $z=2.71$. Despite being interesting and worth further analysis on the variability, this is beyond the scope of this work.

Bands	New	C-COSMOS	Legacy
F+S+H	1140	1047 (922)	2187
F+S	536	397 (474)	933
F+H	448	231 (257)	679
F	121	49 (73)	170
S	21	17 (32)	38
H	7	2 (3)	9
Total	2273	1743 (1761)	4016

Table 2.1: Number of sources with $\text{DET_ML} > 10.8$ in at least one band, for each combination of X-ray bands. The columns labelled as C-COSMOS include the updated numbers, using the information from the new data and in parenthesis also the old numbers as in Elvis et al. (2009).

Bands	New		C-COSMOS		Legacy	
	> 7	>12	> 7	>12	> 7	>12
F	5	5	6	6	12	11
S	4	3	4	3	9	7
H	3	3	4	3	8	7

Table 2.2: Number of expected spurious sources with $\text{DET_ML} > 10.8$ with at least 12 and 7 full band counts, corresponding to a reliability of 99.7% for the new data, the old C-COSMOS data (as in P09 Section5) and in the whole *COSMOS-Legacy*.

There are 58 XMM-COSMOS sources that have been resolved by the smaller *Chandra* PSF into two distinct sources using a maximum radius of $10''$ for the match. In Figure 2.13 we show an example of these sources. Two XMM-COSMOS sources have been resolved into three *Chandra* sources using a maximum radius of $10''$. As a comparison, 25 XMM-COSMOS sources (Brusa et al. 2010) were resolved into two separate C-COSMOS sources. More details on the optical counterparts of the XMM-COSMOS sources resolved in two *Chandra* ones are given in section 3.3.6.

Finally, it is worth noticing that there is a remarkably good agreement between XMM-COSMOS and *Chandra* fluxes. We rescaled the *Chandra COSMOS-Legacy* fluxes using the same slope used for XMM-COSMOS ($\Gamma=2$ in soft band and $\Gamma=1.7$ in hard band) and we found that the median value of the flux ratio $f_{\text{XMM}}/f_{\text{Chandra}}$ is 1.13 in the soft band and 1.22 in the hard band.

2.4 Sky coverage and survey sensitivity

The sky coverage of a survey is the area covered as a function of the flux limit. We computed it in the three standard energy bands (F, S and H) using the exposure and background maps (see Section 2.2.2 and 2.2.3) produced for the source detection, and assuming a power-law spectrum with $\Gamma=1.4$ and Galactic $N_H=2.6 \times 10^{20} \text{ cm}^{-2}$. X-ray observations have a flux limit that changes over the field of view because the *Chandra* PSF changes in both size and shape as a function of the distance from the aim point and because the effective area is vignetted. In this survey, where the

Band	DET_ML \geq 10.8			6 < DET_ML < 10.8		
	New	C-COSMOS	Legacy	New	C-COSMOS	Legacy
Full (F)	2146	1667 (1655)	3813	99	57 (71)	156
Soft (S)	1538	1382 (1340)	2920	159	79 (88)	238
Hard (H)	1325	1115 (1017)	2440	271	165 (165)	436

Table 2.3: Number of sources detected in each band at the two adopted thresholds. The columns labelled as C-COSMOS include the updated numbers using the information from the new data and, in parenthesis, also the old numbers as in Elvis et al. (2009).

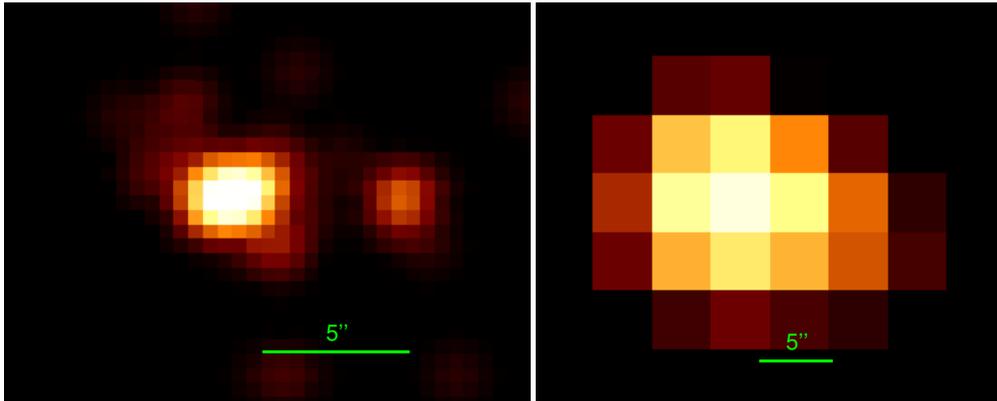


Figure 2.13: Left: two *Chandra COSMOS-Legacy* sources, lid_688 and lid_689. The separation between the two sources is $d \sim 5.3''$. Right: the two sources are not resolved in XMM-COSMOS, where only one object (xid_5307) is detected.

total coverage is obtained using multiple overlapping pointings, every source was observed in up to six different positions on the detector, resulting in a quite uniform average PSF (Figure 2.2).

The procedure we used to compute *COSMOS-Legacy* survey sky coverage is closely similar to that used by P09 for C-COSMOS, but makes use of a PSF map for each observation instead of an analytical form of the PSF as function of the off-axis angle. This is a more time consuming approach but one that returns a more detailed sensitivity map, which can be valuable in other studies (e.g., clustering analysis and correlation functions) or simply for source photometry (Section 2.3.1).

For each observation we used the CIAO tools `mkpsfmap` and `dmimgadapt` to create a background map convolved with the PSF map in such a way that at each position of the map, the count value corresponds to the number of counts in an aperture corresponding to 50% of the encircled energy fraction at that position.

For each position in the whole mosaic (788×776 pixels wide), we computed the minimum number of counts C_{min} needed to exceed the background fluctuations, assuming the same probability for spurious sources (i.e., DET_ML threshold) used in the C-COSMOS and *COSMOS-Legacy* catalogs for the Poisson statistics, i.e. 2×10^{-5} . We used the relation

$$P_{Poisson} = e^{-B} \sum_{k=C_{min}}^{\infty} \frac{B^k}{k!} = 2 \times 10^{-5}, \quad (2.5)$$

where B is the total background counts computed at each position of the grid, by summing the background counts in each observation covering that given position. Equation 2.5 is solved iteratively to find C_{min} ; then the count rate limit, R_{lim} is obtained using

$$R_{lim} = \frac{C_{min} - B}{f_{psf} \times T_{exp}}, \quad (2.6)$$

where T_{exp} is the total, vignetting corrected, exposure time at each position on the grid, while f_{psf} is the encircled count fraction of the PSF. In C-COSMOS, this value was tuned to reproduce the simulation results and then it was fixed to $f_{psf}=0.5$, however any number in the range 0.5-0.9 produced similar results with variations of the order of few percent in the resulting sensitivity.

Finally, we converted the count rate limit R_{lim} into the flux limit using the same conversion factors used for the sources in the catalog based on the position (see Section 2.3.1).

The sky coverage of the *Chandra* COSMOS Legacy survey in the three energy bands is shown in Figure 2.14. We compare our results with those of C-COSMOS (black solid lines) and XMM-COSMOS (blue dashed lines): the new survey covers a similar area to XMM-COSMOS and almost three times the area of C-COSMOS at faint fluxes (e.g., $\sim 5 \times 10^{-15}$ erg cm $^{-2}$ s $^{-1}$ in the soft band) and ~ 2 times at bright fluxes (e.g., $> 10^{-15}$ erg cm $^{-2}$ s $^{-1}$ in the soft band).

We have verified that the limits at 50% completeness for the Legacy catalog are consistent with those computed and reported in Table 2 of P09 of 1.7×10^{-15} , 4.5×10^{-16} and 2.9×10^{-15} erg cm $^{-2}$ s $^{-1}$ in the F, S and H bands, respectively. At this limit, *COSMOS-Legacy* increases by a factor of 3 the area covered with respect to C-COSMOS.

2.5 Number counts

The $\log N$ - $\log S$, i.e. the number of sources $N(> S)$ per square degree detected at fluxes brighter than a given flux S (erg s $^{-1}$ cm $^{-2}$), provides a first estimate of source space density as a function of flux and therefore information on the cosmic population to compare with different models of population synthesis. Given that multiple $\log N$ - $\log S$ curves have been published in the literature, it is also a standard check to validate the many calibration steps used to produce a catalog of X-ray point like sources.

We constructed the $\log N$ - $\log S$ curve for *COSMOS-Legacy* in both the 0.5-2 keV and 2-10 keV bands. Following P09, we included only sources with DET_ML >10.8 and we applied a cut in SNR (> 2 and > 2.5 in soft and hard) to limit the Eddington bias effect, which could have a significant (up to 30-50%) contribution at the lowest fluxes. This choice avoids sources with large statistical uncertainties on their fluxes and limits the errors due to the sky coverage uncertainties at the faint end. With the adopted thresholds in SNR, the agreement measured in P09 between simulated input and output $\log N$ - $\log S$ is better than 5%. The number of sources not included because of the SNR cut is $\sim 1\%$ in the soft and $\sim 5\%$ in the hard band.

The adopted SNRs imply the following flux limits: 2.3×10^{-16} erg s $^{-1}$ cm $^{-2}$ in the soft band and 1.6×10^{-15} erg s $^{-1}$ cm $^{-2}$ in the hard band. These are the same

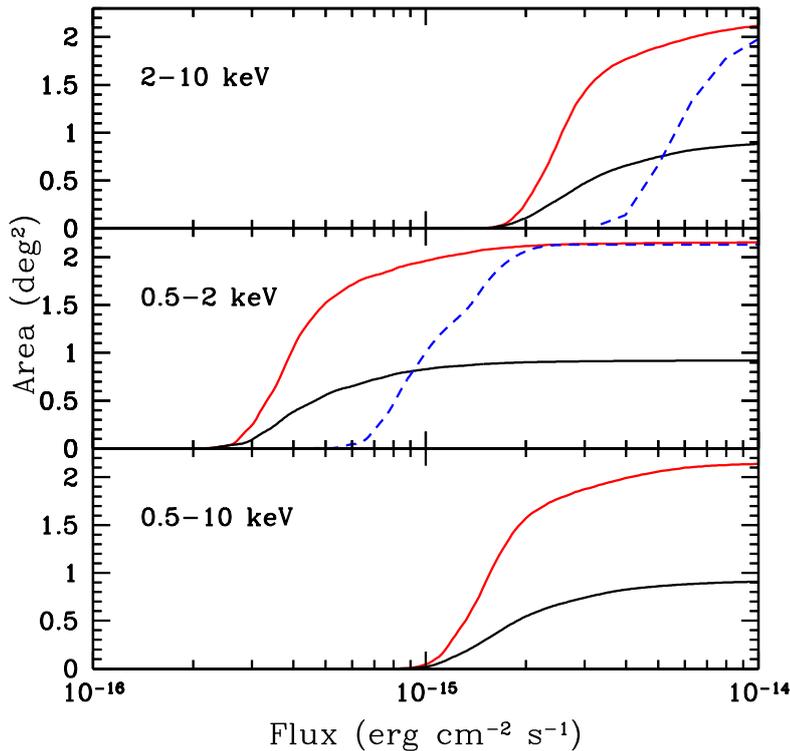


Figure 2.14: Area-flux curve for *COSMOS-Legacy* (red solid line) in hard (top), soft (center) and full (bottom) bands. The coverage of C-COSMOS (black solid line) and XMM-COSMOS in the soft and hard bands (Cappelluti et al. 2009; dashed blue line) are shown for comparison.

flux limits of C-COSMOS, which is expected given that the new observations have the same maximum exposure. The final number of sources used here for the number counts with the above constraints are 2768 in the soft band (1310 from C-COSMOS and 1458 from the new sample) and 2221 in the hard band⁵ (1045 from C-COSMOS and 1176 from the new sample).

We show the results obtained with these source selections in Figure 2.15 (top panels): the normalized Euclidean curves, i.e. with $N(> S)$ multiplied by $S^{1.5}$, are presented in order to enhance the differences between different surveys. In the same figure we include the C-COSMOS (E09) and XMM-COSMOS points (Cappelluti et al. 2009). We also compare our $\log N$ - $\log S$ relationships with those from previous X-ray surveys, spanning from wide (Stripe82 XMM: LaMassa et al. 2013a; 2XMM: Mateos et al. 2008), to moderate (XDEEP2: Goulding et al. 2012), to small areas (4 Ms CDFS: Lehmer et al. 2012). As XDEEP2 and CDFS define their hard band in a slightly different energy range, we converted their energy to 2-10 keV to perform an adequate comparison.

COSMOS-Legacy $\log N$ - $\log S$ covers 3 and 2.5 orders of magnitude in flux in the soft and hard band, respectively, with 2-8% errors at fluxes $< (1-3) \times 10^{-14}$ erg

⁵We also applied a cut in exposure time at 40 ks in the hard band to limit sources (65 in total) at the edges of the field with high background level.

$\text{cm}^{-2} \text{s}^{-1}$, respectively. The excellent statistics allows us us to considerably reduce the uncertainties (20-30%) in the number counts also at bright fluxes, which are now $\simeq 40\%$ smaller than in C-COSMOS.

In the soft band, the agreement between our survey and previous works is excellent at all fluxes, even at $S > 2 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$, where the uncertainties are larger due to the low number of detections (65 sources). In the hard band instead, *COSMOS-Legacy* number counts agree with other surveys at faint fluxes, while we observe an excess, even if within the errors, of sources with respect to C-COSMOS and other works at the bright end (i.e. $S > 2 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$).

We also compare our results with predictions of two different phenomenological models, Gilli et al. (2007) and Treister et al. (2009), assuming column densities in the interval $N_H = 10^{20-26} \text{ cm}^{-2}$ and redshift $z=0-6$. In Fig. 2.15 (bottom panels), we show the ratio of *COSMOS-Legacy* number counts to both models in the soft and hard bands (left and right). At the faint end of the soft band, our results are in agreement with the model prediction within 1-5%, with the Gilli et al. (2007) model (solid points) over-predicting and the Treister et al. (2009) model (open points) under-predicting the counts. At bright fluxes, where the sample is limited by the statistics, the differences between models and data become larger even exceeding 10%. In the hard band, both models reproduce well the observed data within 1-2% below $> 2 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$ and the difference becomes more pronounced at bright fluxes ($> 10\%$)

The Gilli et al. (2007) and Treister et al. (2009) models are based on different assumptions on the fraction of obscured sources and on the assumed luminosity and redshift dependences. Therefore, their differences are more marked when considering obscured and unobscured sources separately. We used the hardness ratio, as defined in Section 2.3.1, to divide the sample using $\text{HR} > -0.2$ for obscured sources and $\text{HR} < -0.2$ for unobscured sources. In the soft (hard) band there are 1057 (1325) obscured sources and 1711 (896) unobscured ones. In Figure 2.16, we present the number counts in the soft and hard bands (left and right) for both obscured (red) and unobscured (blue) sources. A clear difference is observed in the number counts of obscured and unobscured in the soft band, where we observe a ratio of up to ~ 10 at bright fluxes, while it almost disappears in the hard band, where the ratio is very small at all fluxes. This implies that the difference must be dictated by obscuration effects.

The models from Gilli et al. (2007, solid line) and Treister et al. (2009, dashed line), assuming column densities above and below 10^{22} cm^{-2} (red and blue, respectively) are plotted in the same Figure. In the soft band, both model predictions of the number of unobscured sources are in agreement within 10% with our data, up to fluxes of $\sim 3.2 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$, while the difference becomes larger for obscured sources ($> 10-20\%$), with both models over predicting the number of sources at all fluxes. In this last case, the Treister et al. (2009) model predictions are generally worse than those of the Gilli et al. (2007) model, by 5-10%. In the hard band instead, model predictions are very similar and in general excellent agreement with our data (differences < 5 up to fluxes of $10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$), for both samples above and below $\text{HR} = -0.2$.

Overall, these discrepancies between data and models are totally expected given that a different spectral model could change source fluxes and sky coverage, and

that the spectral parameters in the Gilli et al. and Treister et al. models are different from those used in this work. Therefore, despite all the underlying assumptions, the differences between observed number counts and phenomenological models are remarkably small (2–5%; see also LaMassa et al. 2013a for a discussion on discrepancies between data and population synthesis models).

2.6 Summary and Conclusions

In this chapter we presented *COSMOS-Legacy*, a 2.2 deg² *Chandra* survey of the COSMOS field. We employed a total of 4.6 Ms of exposure time, including 1.8 Ms already published by E09 plus 2.8 Ms obtained as an X-ray Visionary Project during *Chandra* Cycle 14. The new data comprise 56 overlapping observations which, added to the 36 C-COSMOS pointings, yield a relatively uniform coverage of ~ 150 ks over the whole *Hubble*-covered area. By construction, the survey flux limit is the same of C-COSMOS, computed in three bands using the same approach of P09.

We followed the same procedure used and tested by P09 combining standard CIAO tools for the data reduction, and `PWDetect` and `CMLDetect` for the data analysis, including the source detection and photometry. We also performed aperture photometry for consistency with the E09 and P09 analysis. The analysis was performed on the new *Chandra* data and also on the outer C-COSMOS frame, overlapping with the new observations. Given that the survey properties (exposure, roll angle and background counts) are consistent with C-COSMOS ones, we used the same probability threshold for the source detection corresponding to `DET_ML` = 10.8. At this limit, we detected 2273 sources that were not previously detected in C-COSMOS, by combining detections in the full, soft and hard bands. 385 of these sources were detected in the area overlapping with C-COSMOS: in the same area we have also found 676 of the 694 old detections, while 18 sources were not detected again. The total number of sources in *COSMOS-Legacy* is 4016; 12, 9 and 8 sources with more than 7 counts are expected to be spurious in full, soft and hard band, respectively.

We computed the source number counts in both the soft and hard bands and we find good agreement between our results and other surveys in the literature as listed above. The large number of sources in *COSMOS-Legacy* (20% or more than the sources in other contiguous surveys) allows us to constrain the number of counts at medium fluxes ($\sim 10^{-15}$ erg cm⁻² s⁻¹) with 10% errors and to reduce the uncertainties on the normalization at bright fluxes where discrepancies between different surveys still exist. The combination of *COSMOS-Legacy* with other surveys at fainter and brighter fluxes allows us to cover more than 4 orders of magnitude in flux.

Using the hardness ratio we measure a fraction of obscured sources of $50^{+17}_{-16}\%$, defined as sources with $\text{HR} > -0.2$, corresponding to column density $> 10^{22}$ cm⁻² at all redshifts, despite the uncertainties on the classification due to complex spectral modeling not taken into account in this work (see Wilkes et al. 2009). For the first time, we computed the number counts for obscured and unobscured sources separately using the hardness ratio as an indication for obscuration ($\text{HR} = -0.2$ corresponding to the separation between $>$ and $< 10^{22}$ cm⁻²). The large number of

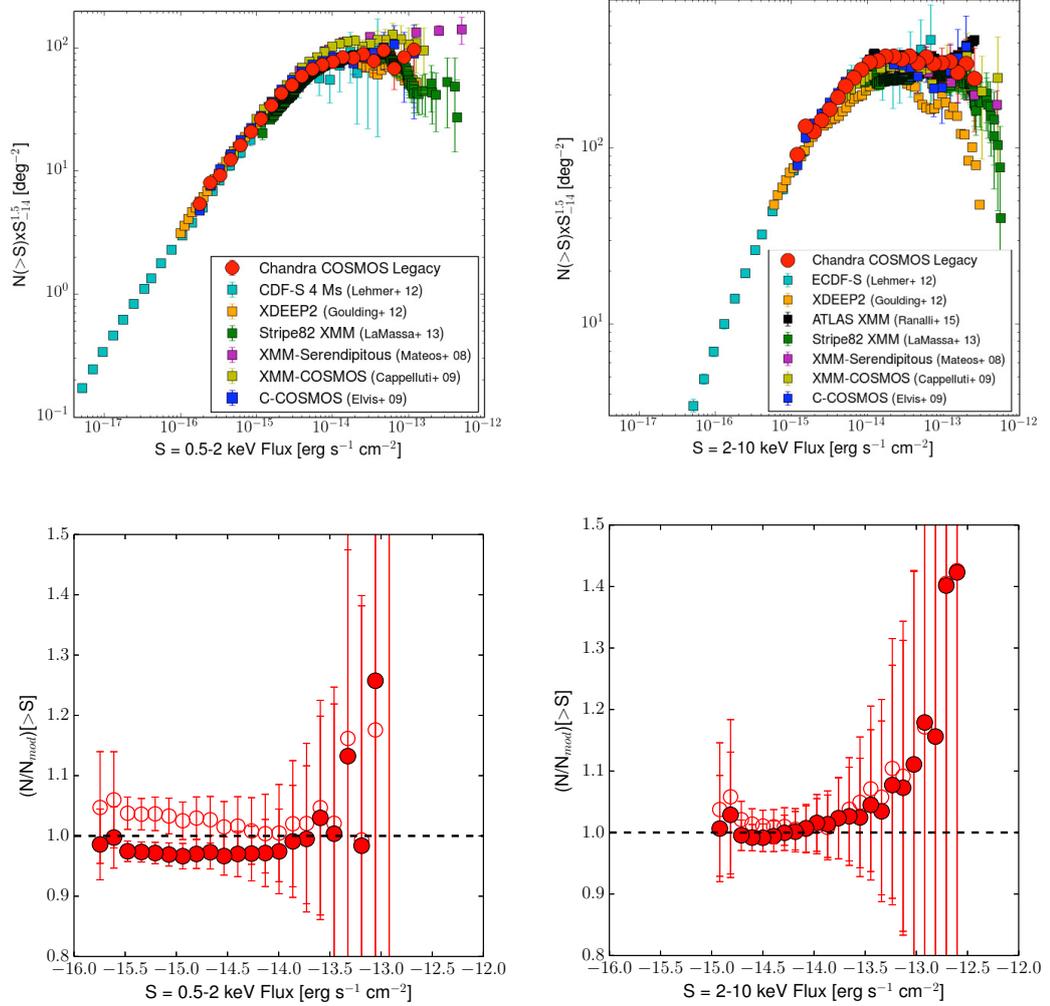


Figure 2.15: Euclidean normalized $\log N$ - $\log S$ curves in soft (top-left) and hard (top-right) bands. The *COSMOS-Legacy* curve for all sources with $\text{DET_ML} > 10.8$ and $\text{SNR} > \text{SNR}_{lim}$ is plotted as red circles. Results from previous works are plotted (see label in the plot). The ratio of *COSMOS-Legacy* number counts to Gilli et al. (2007, red solid) and Treister et al. (2009; red empty) models are plotted in the soft and hard bands (bottom left and bottom right).

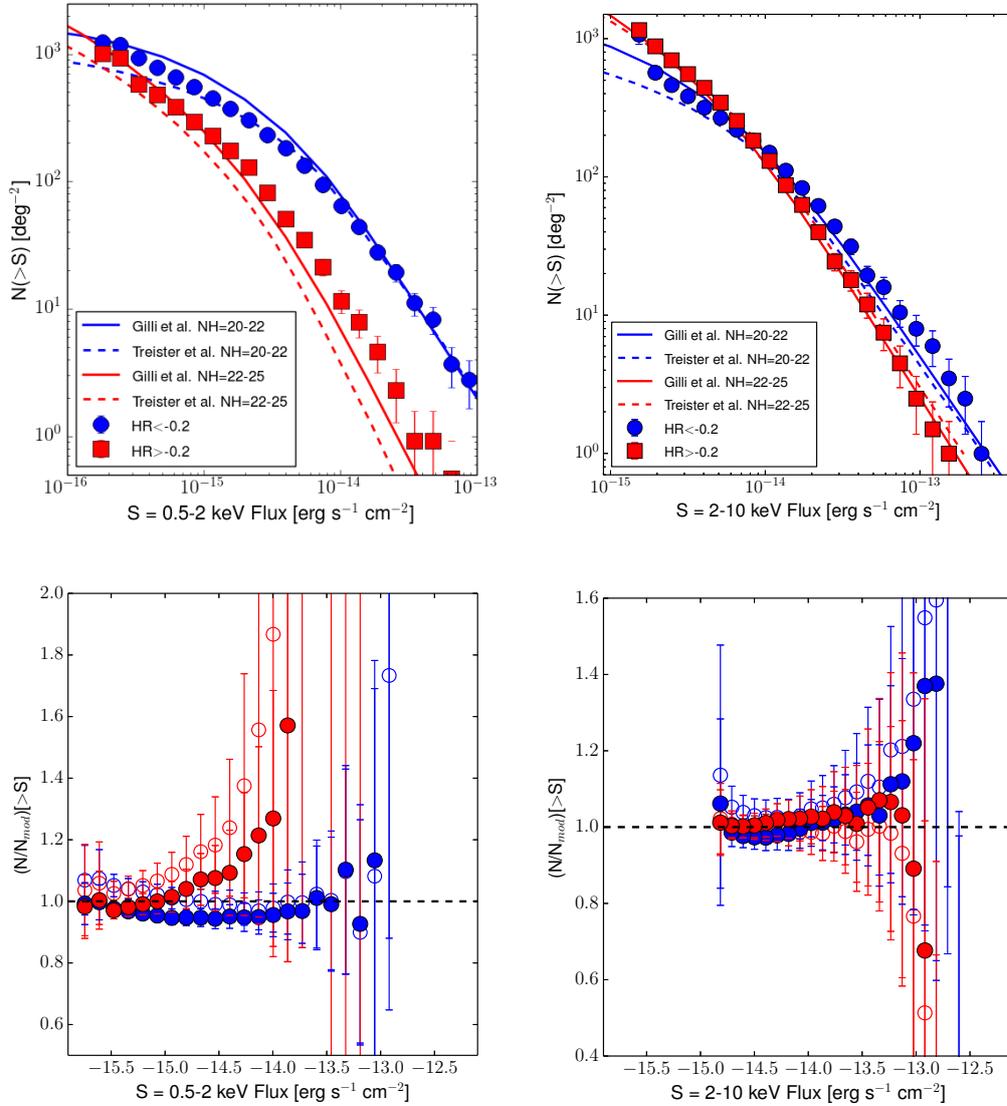


Figure 2.16: Number counts in soft (top-left) and hard (top-right) bands for sources with $\text{HR} > -0.2$ (red squares) and < -0.2 (blue circles) plotted with the Gilli et al. (solid) and Treister et al. (dashed) models with two different column density ranges $>10^{22} \text{ cm}^{-2}$ in red and $<10^{22} \text{ cm}^{-2}$ in blue. The ratio of *COSMOS-Legacy* number counts to Gilli et al. (2007; solid) and Treister et al. (2009; empty) models are plotted in the soft and hard bands (bottom left and bottom right).

sources in each sample (about a thousand or more) allows us to compute the number counts for the two populations and reveals a larger difference in the soft band, while a very small (if not absent) difference in the hard band is observed. Given the large range of luminosities and redshifts probed by *COSMOS-Legacy*, this can be interpreted as a difference in orientation rather than an intrinsic difference due to an evolutionary state between obscured and unobscured sources.

In Figure 2.17, the area–flux parameter space of the most recent *Chandra* and XMM-*Newton* surveys (CDFS 4Ms, Xue et al. 2011; AEGIS-XD, Nandra et al. 2015; XDEEP2-F1, Goulding et al. 2012; C-COSMOS, E09; XMM-COSMOS, Cappelluti et al. 2009; X-Bootes, Murray et al. 2005; XMM-Atlas, Ranalli et al. 2015; Stripe 82, LaMassa et al. 2013a,b; XMM-XXL, Pierre et al. submitted, see also Pierre et al. 2004) is presented. Most surveys lie on a locus (yellow shaded area) determined by our current X-ray telescope capabilities. *COSMOS-Legacy* is exploring a new region off this locus, an additional factor 2-3 deeper at the areas it covers, by using a total exposure time which is unusually large (4.6 Ms total) for that given area flux combination, and preparing for surveys with future facilities. The X-Bootes survey also explores a region off the survey locus, but at brighter fluxes and over a larger area.

In future decades, with facilities like *eROSITA* (Merloni et al. 2012), *Athena* (Nandra et al. 2013, cyan solid line in Figure 2.17) and *X-ray Surveyor* (Vikhlinin et al. 2012, green solid line in Figure 2.17), it will be possible to explore a new region of area-flux parameter space, moving away from the current survey locus towards the bottom right corner of Figure 1.8. For example, *Athena* will perform a multi-tiered survey and given the combination of large effective area and field of view will enable X-ray surveys to be carried out two orders of magnitude faster than XMM-*Newton* and *Chandra* (see Figure 2 of Aird et al. 2013). With a *Chandra*-like resolution over $10'$, *X-ray Surveyor* will be able to cover the same *COSMOS-Legacy* area at the same flux in only 55 ksec (A. Vikhlinin private communication), 80 times faster than *Chandra*.

Thanks to the large area covered at considerable depth, *COSMOS-Legacy* can now address those questions for which a large number of detected X-ray sources at a medium depth with uniform multiwavelength coverage and almost complete redshift information is needed. The excellent positional accuracy allows us to obtain multiwavelength identifications and photometric redshifts for 96% of the sources (see section 3.5.3). Several works on different topics are already in preparation: for example, the X-ray spectral analysis and X-ray variability of the bright sample with a focus on the hunt for obscured sources (Lanzuisi et al. in prep.); the multiwavelength spectral energy distribution fitting with host galaxy properties (mass and star formation rates) for both optically classified as obscured and unobscured sources (Suh et al. in prep.); clustering measurement and dark matter halo mass (Allevato et al. in prep.); a catalog of X-ray extended sources (Finoguenov et al. in prep). Finally, we will discuss the properties of the high-redshift sample, with particular focus on the survey space density, in Chapter 4.

The wide area and the availability of extensive multiwavelength data in the COSMOS field enable us to probe the average X-ray emission of objects not individually detected by *Chandra*, therefore beyond the flux limit, through a stacking analysis. The combined *Chandra COSMOS-Legacy* dataset is now fully implemented in the

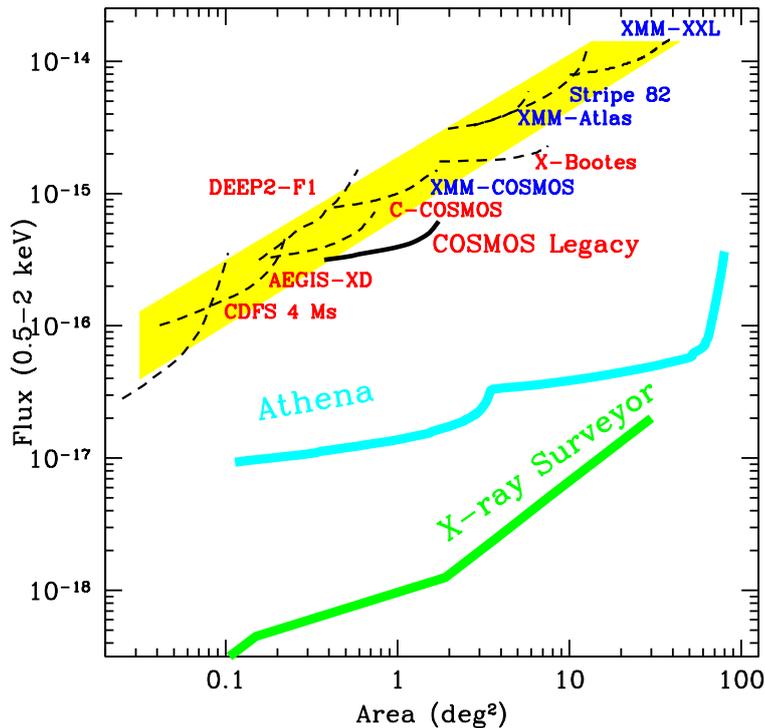


Figure 2.17: Area–flux curves for *Chandra* (red) and *XMM-Newton* (blue) contiguous X-ray surveys. Each survey has been plotted using each sensitivity curve starting from the flux corresponding to the area that is 80% of the maximum area for that survey to the flux corresponding to the 20% of the total area. The plotted surveys are: CDFS 4Ms (Xue et al. 2011), XDEEP2-F1 (Goulding et al. 2011), AEGIS-XD (Nandra et al. 2015), C-COSMOS (E09), XMM-COSMOS (Cappelluti et al. 2009), X-Bootes (Murray et al. 2005), XMM-Atlas (Ranalli et al. 2015), Stripe 82 (LaMassa et al. 2013a,b), XMM-XXL (PI: Pierre; see also Pierre et al. 2004). The survey locus described in the last section is drawn in yellow. The expected area–flux curves for *Athena* (Nandra et al. 2013) and *X-ray Surveyor* (Vikhlinin et al. 2012) are also plotted as cyan and green solid lines, respectively.

web-based *Chandra* stacking tool CSTACK⁶. This enables us to investigate the X-ray properties of differently selected samples, such as optical selected galaxies (e.g. Mezcuca et al. submitted, finding indications of weak AGN activity in low mass non-elliptical galaxies), highly obscured AGN selected using both infrared or radio criteria, and early AGN populations at $z > 5$.

⁶See <http://lambic.astrosen.unam.mx/cstack>. Login as user=guest, password=guest and see the explanatory manual. As of writing this thesis, stacking analyses utilizing the C-COSMOS dataset is publicly available. Analyses involving the whole *Chandra COSMOS-Legacy* dataset is still proprietary and will become public in due course.

2.7 CID-42: a candidate recoiling SMBH in the *Chandra* COSMOS-Legacy survey

Bound binary SMBHs systems can be a byproduct of galaxy major mergers, and these binaries can also eventually merge themselves (e.g. Volonteri et al. 2003, Hopkins et al. 2008, Colpi & Dotti 2009). The coalescence of SMBH binary systems produce strong, anisotropical gravitational wave (GW) radiation: consequently, to conserve linear momentum, the merged SMBH has to recoil (Peres 1962, Bekenstein 1973). When the SMBH recoils from the center of the galaxy, its disk and broad line regions are carried with it, while the external part of the system is left behind, close to the galaxy center (Merritt et al. 2006, Loeb 2007). Since GW recoiling SMBHs are ejected from their host-galaxy center, these events can affect the host-galaxy/SMBH co-evolution, as demonstrated by numerical simulations (Blecha et al. 2011, Sijacki et al. 2011, Guedes et al. 2011). However, the observational research for recoiling SMBHs started only a few years ago (Bonning et al. 2007, Eracleous et al. 2012, see Komossa 2012 for a review) and only few serendipitous discoveries of candidates have been reported so far (Komossa et al. 2008, Shields et al. 2009, Robinson et al. 2010, Jonker et al. 2010, Batcheldor et al. 2010, Steinhardt et al. 2012).

The *Chandra*-COSMOS source CXOC J100043.1+020637 ($z=0.359$), also known as CID-42, is a candidate GW recoiling SMBH, with multiple evidences from both imaging in optical, X-ray and radio (Civano et al. 2010, 2012; hereafter C10 and C12b; Novak et al. 2015; see also Figure 2.18) and optical spectroscopic signatures (C10). All available data are in favour of a recoiling SMBH system, the SMBH being ejected $\sim 1-6$ Myr ago, as shown by detailed modeling presented in Blecha et al. (2013).

C10 showed that both the *Chandra* and the XMM-*Newton* spectra of CID-42 had evidence of a remarkable inverted *P-Cygni* profile, i.e., an absorption feature redshifted with respect to the emission component. This multi-component feature was detected at ~ 6 keV in the rest frame (Figure 2.19, left panel).

The emission feature was consistent with being a neutral iron $K\alpha$ line at the system redshift, with constant flux. The equivalent width of the emission line was however significantly stronger in the *Chandra* spectrum ($EW = 570 \pm 260$ eV) than in the XMM-*Newton* one ($EW = 142^{+143}_{-86}$ eV).

The absorption feature (at ~ 6 keV in the rest frame) was also detected in both XMM-*Newton*-EPIC pn and *Chandra*-ACIS spectra: the line energy centroid slightly changed between different observations, inside the energy range between 5.8 and 6.2 keV ($\Delta E_{rest} \sim 500$ eV; see Figure 8 of C10), with intensity of 350 ± 120 eV. The significance of the redshifted absorption line (see Section 2.7.2 for a more detailed analysis through Monte Carlo Simulations) in the XMM-*Newton* data is 3σ , strengthened by the detection of the line also in *Chandra*-ACIS at 2.2σ (Figure 2.19, left panel).

C10 suggested a possible interpretation for the redshifted absorption line seen in CID-42 as gas (either neutral or ionized iron) infalling into the recoiling SMBH at relativistic velocities ($0.02-0.14c$). However, the combination of low data quality and degeneracy between velocity and ionization state did not allow to make a strong claim on the line physical causes.

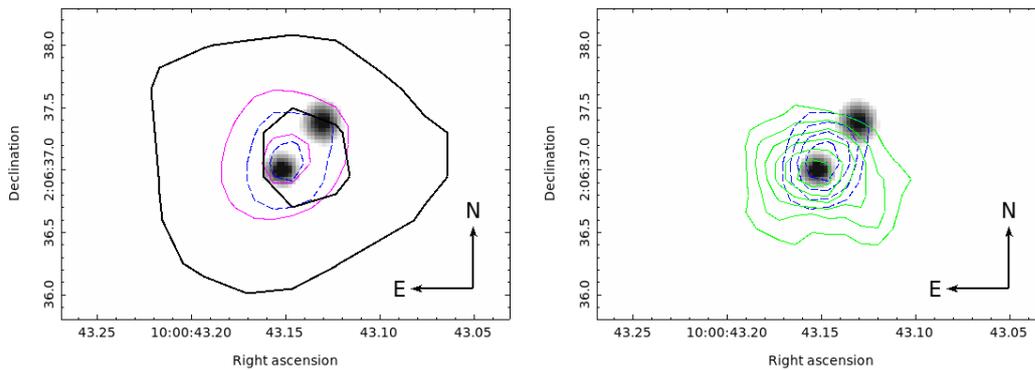


Figure 2.18: Left: HST/ACS gray-scale of the two CID-42 optical components of CID-42. Radio data contours at 3 and 6σ are overlaid. Thick black lines show data from the VLA 1.4 GHz map from Schinnerer et al. (2007; RMS= $10 \mu\text{Jy beam}^{-1}$). Thin magenta lines and dashed blue lines are from the VLA 3 GHz SPW stack map and MSMF map (Novak et al. 2015; RMS= $4.8 \mu\text{Jy beam}^{-1}$ and $4.6 \mu\text{Jy beam}^{-1}$, respectively). Right: HST/ACS gray-scale image of CID-42, with X-ray Chandra High Resolution Camera image contours overlapped in green (Civano et al. 2012b). VLA 3 GHz MSMF map with 1σ steps is also plotted (blue line: 3σ). Image from Novak et al. (2015).

Few cases of objects with redshifted absorption lines are reported in the literature (NGC 3516, Nandra et al. 1999; E1821+643, Yaqoob & Serlemitsos 2005; Mrk 509, Dadina et al. 2005; PG 1211+143, Reeves et al. 2005; Q0056-363, Matt et al. 2005; Ark 120, Nandra et al. 2007; Mrk 335, Longinotti et al. 2007). Interestingly, for the vast majority of these objects the redshifted absorption line was not observed while additional observations were performed (see Vaughan & Uttley 2008 for an extended discussion on the statistical significance of these lines; see also Tombesi et al. 2010 for a general review of candidate redshifted lines in X-ray spectra). A similar statistic suggests that these features must be highly variable, and that their occurrence, or duty cycle, is very low.

2.7.1 Data analysis

In order to perform an accurate characterization of the X-ray absorber (density, velocity, covering factor and ionization state), we requested and obtained a 123 ks long un-interrupted XMM-Newton observation, taken on June 3, 2011 (during revolution 2103, observation ID 0672780101, PI: F. Civano). Standard XMM-Newton SAS tasks *epproc* and *emproc* (SAS 12.0.1) were used to produce calibrated pn and MOS event file. The obtained spectra have ~ 2500 net counts in the 0.5-8 keV band in both the pn and in the MOS1 and MOS2 combined observations, ~ 2200 (2000) net counts in the 0.5-2 keV band in the pn (MOS1 and MOS2 combined) observation and ~ 450 net counts in the 2-8 keV band in both the pn and in the MOS1 and MOS2 combined observations.

During our spectral analysis we modelled simultaneously the source and the background spectra. To do so, we first estimated the XMM-Newton EPIC pn and

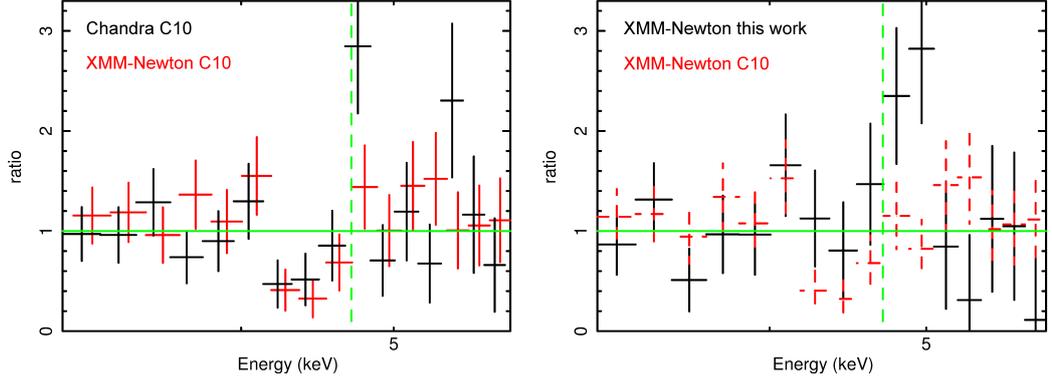


Figure 2.19: Left: data to model ratio in the energy range of the inverted *P-Cygni* profile in EPIC pn (red) and ACIS (black) spectra of CID-42, adapted from C10.

Right: data to model ratio for C10 pn data (red dashed line), compared to the new pn data (black solid). In both panels, the spectra are fitted with only a continuum model. The vertical line corresponds to the 6.4 keV rest frame energy. The spectra have been rebinned in energy with $\Delta E \sim 200$ eV in order to highlight the absorption features.

MOS background spectra, which were fitted over the 0.3-7 keV range using XSPEC version 12.8 (Arnaud 1999). The background model had the following components: two power-law components, a thermal component for the soft part of the spectrum, three Gaussian emission lines to reproduce the features of the pn and MOS backgrounds. This final background best-fit model was then rescaled to the source area.

We fitted the source spectra in the 0.5-7 keV band with an absorbed power-law, with the further contribution of a thermal component (*mekal*). The Galactic column density along the line of sight of $N_{H,Gal} = 2.6 \times 10^{20} \text{ cm}^{-2}$ (Kalberla et al. 2005) was also taken in account. The fitting analysis was made using the modified Cash statistic implemented in XSPEC (*cstat*; Cash 1979), binning the spectra by 1 count per bin. We first separately fit the pn and MOS1+MOS2 spectra, to check the agreement between the different instruments. We then performed a combined analysis, to put tighter constraints on the spectral parameters. The pn and MOS1+MOS2 spectra were first fitted separately to verify the consistency of the fit between instruments, then jointly to more tightly constrain the errors on the spectral parameters.

The spectral analysis results for the joint fit are listed in Table 2.4. The results obtained in C10 are also shown for comparison. The spectral index obtained from the joint fit is $\Gamma = 2.16 \pm 0.08$, slightly steeper than the slope measured in C10. The presence of a thermal component with temperature $kT = 0.18^{+0.02}_{-0.04} \text{ keV}$ is significant at $>5\sigma$, as indicated by the F-test. We only measured an upper limit of $N_H < 6 \times 10^{20} \text{ cm}^{-2}$ for the intrinsic absorption.

In Figure 2.20 we compare the 0.5-10 keV band flux with the fluxes obtained in the previous XMM-Newton and Chandra observations (C10 and C12b). The light curve spans a period of ~ 9 years. The flux reported here is consistent with the flux reported in C10 for the XMM-Newton data and consistent within the error with

Fit Parameter	ACIS-I (C10)	pn (C10)	pn+MOS (here)
Continuum			
Γ	$1.88^{+0.17}_{-0.13}$	$1.95^{+0.07}_{-0.06}$	$2.16^{+0.08}_{-0.08}$
N_H (10^{22} cm $^{-2}$)	<0.2	<0.02	< 0.06
kT (keV)	< 0.11	< 0.13	$0.18^{+0.02}_{-0.04}$
$F_{0.5-2keV}$ (10^{-14} erg cm $^{-2}$ s $^{-1}$)	1.8 ± 0.2	4.8 ± 0.3	5.1 ± 0.2
$F_{2-10keV}$ (10^{-14} erg cm $^{-2}$ s $^{-1}$)	2.9 ± 0.3	6.1 ± 0.4	4.9 ± 0.3
$F_{0.5-10keV}$ (10^{-14} erg cm $^{-2}$ s $^{-1}$)	5.1 ± 0.4	10.9 ± 0.6	10.1 ± 0.5
Emission Line (one Gaussian fit)			
Observed Energy	6.44 ± 0.07	6.60 ± 0.15	$6.62^{+0.12}_{-0.09}$
Line σ (keV)	<0.12	<0.2	<0.6
EW (eV)	570 ± 260	142^{+143}_{-86} eV	593^{+347}_{-390}
Emission Line (double Gaussian fit)			
Observed Energy	$6.70^{+0.11}_{-0.13}$
Line σ (keV)	<0.2
EW (eV)	394^{+278}_{-160}
Observed Energy	$6.41^{+0.37}_{-0.30}$
Line σ (keV)	<0.2
EW (eV)	<500

Table 2.4: Spectral analysis results (continuum and emission line) and fluxes for the joint spectral fitting of pn and MOS1+MOS2 and the values from the ACIS-I and pn spectra analyzed in C10. Errors are at 90% confidence level.

the C12b flux for the *Chandra* HRC observation taken in January 26, 2011. CID-42 is among the 20% most variable sources among the brightest (>1000 counts) 65 sources in a sample of 638 XMM-COSMOS objects (Lanzuisi et al. 2014).

2.7.2 P-Cygni profile analysis

In Figure 2.19, we show the ratio between data and model of the new pn data (black), compared with those presented in C10 (red; the expected energy of the 6.4 keV Fe $K\alpha$ line is also plotted). As shown there, while the iron emission line is still clearly detected, no clear evidence of the redshifted absorption line has been found in the new data.

The upper limit on the absorption feature equivalent width is $EW < 162$ eV, if we fit the line in the same position it was found in C10 (4.5 keV observed frame). This result does not change significantly even if we allow the absorption feature to be observed in the energy range between 4.2 and 4.6 keV, observed frame. In Figure 2.20 (bottom panel), we plot the compilation of measured EW from previous observations and from this work.

In order to extend our analysis and better understand the reason of the disappearing of the redshifted absorption line, we performed Monte-Carlo simulations (10k runs). First of all, we produced simulated spectra using the FAKEIT routine in XSPEC, to estimate the significance of the C10 redshifted absorption line. Our input model was an absorbed power-law with the best-fit parameters obtained in C10 and no emission and absorption lines. The simulated spectra were then fitted first with a simple power-law model, then we added to the model a narrow absorp-

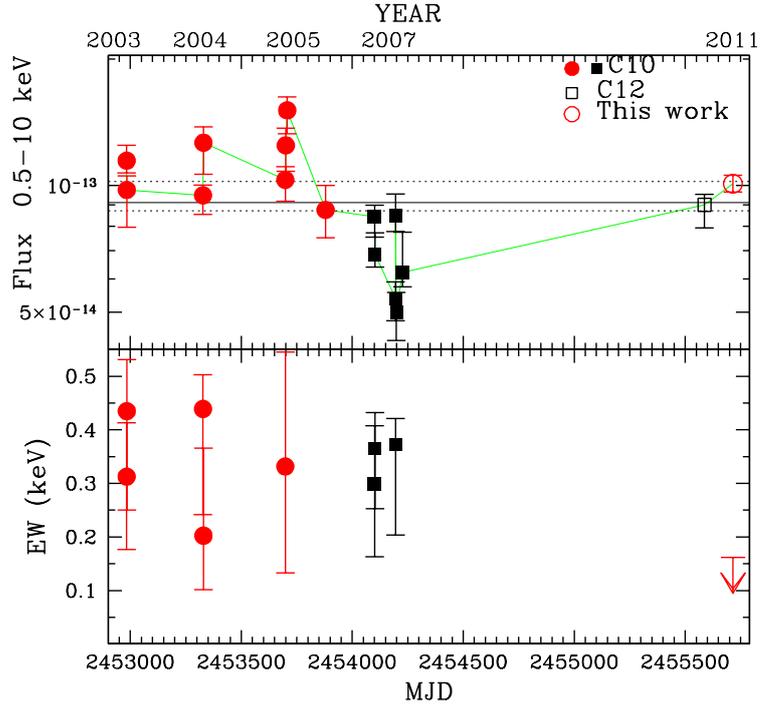


Figure 2.20: *Top*: X-ray full band flux light curve (red=*XMM-Newton* black=*Chandra*) adapted from C10, including the data point from the 2011 *XMM-Newton* data and the C12b *Chandra* HRC data. The reported errors are at 90% confidence level. The mean value and error are reported as horizontal lines. *Bottom*: Absorption line equivalent width measured in different epochs spectra. Figure from Lanzuisi et al. (2013b).

tion Gaussian line. Both the absorption line energy (in the range 2–7 keV) and the line intensity (at negative values) were left free to vary.

Only ~ 30 of these simulations showed a variation $\Delta\text{Cash} > 12$, i.e., the value observed in the C10 *XMM-Newton* spectrum between a fit without and with a line. Consequently, the probability to detect the same feature observed in C10 by chance in the *XMM-Newton* data alone is $\sim 3 \times 10^{-3}$, i.e., the feature is significant at confidence level $\sim 99.7\%$ (3σ). Furthermore, the absorption feature significance is enhanced by the detection also in *Chandra-ACIS* at 2.2σ (Figure 2.19, left panel).

A second kind of simulation using the FAKEIT routine within XSPEC was then performed, to assess whether the apparent disappearance of the absorption line could be justified by low data quality. The simulated spectra have the continuum flux and background level observed in 2011, plus an absorption iron line with the same properties of the one observed in C10 (i.e., same equivalent width, width and observed energy of 4.5 keV). We also performed further simulations, decreasing the intensity of the line from 95% to 5% (in step of 5%) of the C10 value.

The fake spectra were fitted with a power-law model plus an absorption line with a centroid energy fixed at the *XMM-Newton* C10 value, with normalization

free to vary to both positive and negative values.

We measured the line intensity, together with its 90% confidence value, in all the simulated spectra. The probability that the line in the observed data has an intensity comparable with the one observed in C10 is lower than 2.3% with a more stringent limit of 1.2% obtained measuring the fraction of sources with 90% confidence value lower than the one measured in these data. The probability that the line has an intensity (at 90% confidence) comparable to 50% of the C10 is of $\sim 5\%$, and it becomes higher with fainter input lines reaching 20-25%.

2.7.3 Potential causes of the absorption line disappearance

C12b proposed that the most likely explanation for CID-42 is the one of a recoiling SMBH moving out from the galaxy at a velocity of ~ 1300 km/s, an interpretation confirmed also by detailed simulations (Blecha et al. 2013). In the recoiling scenario, the absorption feature is related to absorption by ionized iron, falling into the SMBH with relativistic velocity. While the SMBH is recoiling from the center of the galaxy, it drags with it its close regions, with velocities greater than the recoil velocity⁷.

The equivalent width of the absorption feature observed in C10 implied a high column density of ($N_H \sim 5 \times 10^{23} \text{ cm}^{-2}$), while the negligible absorption in the soft band required the absorber to be highly ionized ($\log \xi \sim 3$). No obscuration is seen in the soft band also in the new observation, therefore the ionization value should have not change significantly. The column density of the absorber should have instead dropped by at least a factor 10 ($N_H < 5 \times 10^{22} \text{ cm}^{-2}$).

If we instead assume that the absorber did not change its density from C10, the ionization parameter should be $\log \xi > 4$ (e.g., the gas has to be extremely ionized). A strong increase of ionization parameter could be explained with increasing flux from the ionizing source, but in CID-42 the flux is actually decreasing (see Fig. 2.20, top). Nonetheless, a delay between the flux burst and the increase of ionization state is expected, and it is related to the distance between the cloud and the central source.

Assuming a velocity law for the gas

$$v(r) = -c \sqrt{\frac{2r_g}{r}} \sqrt{1 - \frac{L}{L_{edd}}} \quad (2.7)$$

(Longinotti et al. 2007) we estimate the distance r of the infalling cloud from the SMBH, $d \sim 100$ -500 gravitational radii. $r_g = GM_{BH}/c^2$ is the black hole gravitational radius, while L/L_{edd} is the ratio between SMBH and Eddington luminosities. We assumed infall velocities in the range $v=0.02$ - $0.14c$ (related to different iron ionization states), and the BH mass and Eddington ratio from C10. A similar distance implies a light travel time of 4 - 20×10^4 s (in the observer frame): consequently, the source ionized flux should have strongly decrease in just few tens of hours, a highly unlikely possibility.

The life time of the infalling gas is estimated to be $\sim (0.3-11) \times 10^6$ s (in the observer frame), i.e., lower than the time between different X-ray observations, but larger than the duration of each observation. In principle, an absorber made

⁷For CID-42, all the material within 10^5 gravitational radii from the SMBH will move with it.

of clumpy and discrete clouds would imply that a different cloud is observed in each observation, and its properties can be unrelated from those of the previously observed cloud. A continuous infalling flow should instead have different density peaks to to produce absorption lines with changing properties.

Finally, Tombesi et al. (2010) proposed that absorption lines observed at energies below 6.4 keV neutral could also be interpreted as blueshifted transitions of helium or hydrogen like ions from elements lighter than iron (Si, S, Ar, Ca), with very high outflowing velocities (0.1-0.5 c). However, a similar scenario would imply velocities of 0.3-0.7 c for the CID-42 outflow, which is very unlikely.

Chapter 3

The *Chandra COSMOS-Legacy* optical/infrared counterparts

In this chapter, we present the catalog of optical and infrared counterparts of new *Chandra COSMOS-Legacy* sources (see previous chapter): this work will also be published in Marchesi et al. (submitted). We then describe and discuss several X-ray and optical/IR photometric and spectroscopic properties of the sources in the whole survey (i.e., both from the new dataset and the old C-COSMOS one), with a particular focus on the different ways to estimate source obscuration using the optical/IR rather than the X-ray information. The chapter is organized as follows: in section 3.1 we describe the X-ray, optical and infrared catalogs used in this work, in Section 3.2 we describe the cross-catalog identification technique, while in Section 3.3 we show the results obtained in the identification process, and in Section 3.4 we show some basic properties of the different types of optical counterparts. In Section 3.5 the spectroscopic and photometric redshifts of the survey are described, together with the spectral and SED-based classification, in Section 3.6 we analyze the relations between X-ray and optical/IR properties and in Section 3.7 we summarize the main results of this chapter.

Throughout this work we assume a cosmology with $H_0=70$ km s⁻¹ Mpc⁻¹, $\Omega_M=0.29$ and $\Omega_\Lambda=0.71$. The AB magnitude system is used if not otherwise stated.

3.1 Identification datasets

The X-ray catalog used in this work is obtained from the *Chandra* COSMOS Legacy survey, which properties have extensively been described in Chapter 2. In this section, we refer to the subsample of the catalog which contains 2273 new point-like X-ray sources, not previously detected in C-COSMOS, detected down to a maximum likelihood threshold DET_ML=10.8 in at least one band (0.5-2, 2-7 or 0.5-7 keV), corresponding to a Poisson probability of $P \simeq 5 \times 10^{-5}$ that a detected source is actually a background fluctuation.

The flux limits of the survey at 20% of the area of the whole survey are 1.3×10^{-15} erg s⁻¹cm⁻² in the full band (0.5-10 keV), 3.2×10^{-16} erg s⁻¹cm⁻² in the soft band (0.5-2 keV) and 2.1×10^{-15} erg s⁻¹cm⁻² in the hard band (2-10 keV). The full and hard band fluxes were extrapolated from net counts measured in 0.5-7 and 2-7 keV,

respectively, assuming a power law with a slope of $\Gamma=1.4$ (not only for consistency with the work done in C-COSMOS, but also because this slope well represents a distribution of both obscured and unobscured AGN, being the X-ray background slope, see, e.g., Markevitch et al. 2003). The number of sources with $\text{DET_ML}>10.8$ in at least one band, for each combination of bands, is reported in Table 2.1

We identify the X-ray sources searching for counterparts in three different bands:

1. i band ($\sim 7600 \text{ \AA}$), using the Subaru photometric catalog (Capak et al. 2007). Given that the Subaru catalog is saturated at magnitudes brighter than $i_{AB}=20$, we completed our i -band sample using information from the Canada-France-Hawaii Telescope (CFHT; McCracken et al. 2010) and from the SDSS catalog (see Section 3.3.2 for further details on the positional match between the i -band CFHT and SDSS sources and the sources detected in K or $3.6 \mu\text{m}$ band). In the analysis of the X-ray, optical and IR properties of the sample described in Section 3.6, we used the Subaru magnitude; if the Subaru magnitude was not available, we used the CFHT magnitude, and we used the SDSS magnitude only for those sources with no Subaru or CFHT magnitude. Sources with only SDSS information are mainly very bright sources saturated in Subaru and CFHT catalogs. The final optical catalog contains about 870,000 sources at a signal-to-noise ratio (SNR) >5 , covering a range in magnitude between $i \simeq 12$ and $i \simeq 27$. From now on we refer to this catalog as the ‘‘optical catalog’’.
2. K_S band ($2.15 \mu\text{m}$), using the UltraVISTA information from the Laigle et al. (submitted) catalog, not available at the time of C12, and the CFHT catalog. The UltraVISTA catalog has been obtained detecting and selecting objects using the ultra-deep chi-squared combination of $YJHK_S$ and z^{++} images. This catalog, although not K -selected, is sensitive to redder wavelengths than the Subaru i -band catalog, and it is therefore complementary to it. The catalog contains $\simeq 415,000$ sources detected at $\text{SNR}>5$ to a K_S magnitude limit of 26, and covers an area of $\simeq 2.0 \text{ deg}^2$, while the CFHT catalog contains $\simeq 320,000$ sources detected at $\text{SNR}>5$ to a magnitude limit of 24.5, and covers an area of $\simeq 2.2 \text{ deg}^2$. In the analysis of the X-ray, optical and IR properties of the sample described in Section 3.6 we used the CFHT information only for sources with no secure UltraVISTA counterpart available. The *Chandra* COSMOS Legacy survey area is not completely covered by the K -band catalog: 27 X-ray sources ($\simeq 1\%$) are in fact outside the field of view of both the UltraVISTA and the CFHT surveys.
3. $3.6 \mu\text{m}$, using the *Spitzer* IRAC catalog from Sanders et al. (2007; hereafter we refer to this catalog as the Sanders catalog) and the SPLASH IRAC magnitude from the Laigle et al. (submitted) catalog (hereafter ‘‘SPLASH catalog’’). It is worth noticing that the SPLASH catalog, unlike the Sanders catalog, is not a $3.6 \mu\text{m}$ -selected catalog. The $3.6 \mu\text{m}$ SPLASH magnitude has been obtained performing aperture photometry at the position where the source has been detected in the combined $YJHK_S$ and z^{++} image. Nonetheless, we used the SPLASH information because it reaches more than 1.5 magnitudes deeper than the Sanders catalog, with a significantly smaller photometric error. The SPLASH catalog contains $\simeq 350,000$ sources with $\text{SNR}>5$, with a

magnitude limit of 26.0 (i.e., $\simeq 0.15 \mu\text{Jy}$), and covers an area of $\simeq 2.4 \text{ deg}^2$: 22 *Chandra COSMOS-Legacy* sources lie outside the field of view of this catalog. The Sanders catalog contains instead $\simeq 330,000$ sources at $3.6 \mu\text{m}$ to a magnitude limit of 24.5 (i.e., $\simeq 0.6 \mu\text{Jy}$) at $\text{SNR} > 5$ and covers the whole *Chandra COSMOS-Legacy* field. In the analysis of the X-ray, optical and IR properties of the sample described in Section 3.6 we used the Sanders information only for sources with no secure SPLASH counterpart.

In the final part of the identification process we also made use of the Advanced Camera for Surveys (*ACS*)/Hubble Space Telescope (*HST*) images of the COSMOS field (Scoville et al. 2007b; Koekemoer et al. 2007) to visually check our identifications, taking advantage of the ACS PSF, of the accuracy of the positions, and of the depth of the observations ($I_{F814W} \simeq 27.8 \text{ AB mag}$, 5σ for an optimally extracted point source). The *ACS/HST* survey covers only the central $\simeq 1.5 \text{ deg}^2$ of the COSMOS field, therefore only $\simeq 70\%$ of the *Chandra COSMOS-Legacy* sources were actually imaged with *ACS/HST*: for the remaining part, we used the *i*-band Subaru images.

We report in Table 3.1 the limiting magnitudes at $\text{SNR} > 5$ for all the catalogs used in our identification process.

Catalog	Mag _{lim} (AB)
<i>i</i> Subaru	27.4
<i>i</i> CFHT	25.1
<i>i</i> SDSS	24.6
<i>K</i> UltraVISTA	26.0
<i>K</i> CFHT	24.0
$3.6 \mu\text{m}$ SPLASH	26.0
$3.6 \mu\text{m}$ Sanders	24.5

Table 3.1: Catalogs used to find Legacy counterparts and their magnitude limit at $\text{SNR} > 5$.

3.2 X-ray source identification method

3.2.1 Method

Following the procedure of Brusa et al. (2005), we used the likelihood ratio (LR) technique adopted in C12 and first developed by Sutherland & Saunders (1992). This procedure was applied first to the XMM-COSMOS catalog (Brusa et al. 2007, hereafter B07; B10) with a percentage of “reliable identifications” greater than 80%, and later on C-COSMOS with a percentage of “reliable identifications” of $\simeq 96\%$. This technique takes into account, for each possible counterpart, the probability that it is a real or a spurious identification, using both the separation between the optical and the X-ray positions, and, as a prior, the information on the counterpart magnitude with respect to the overall magnitude distribution of sources in the field, thus making this method much more statistically accurate than one based on a positional match only.

The LR is defined as the ratio between the probability that an optical or infrared source is the correct identification and the corresponding probability for a background, unrelated object:

$$LR = \frac{q(m)f(r)}{n(m)} \quad (3.1)$$

where m is the magnitude and r the positional offset from the X-ray source position of the optical or infrared candidate counterpart.

$n(m)$ is the density of background objects with magnitude m : we computed the distribution of the local background objects using the objects within a $5''$ - $30''$ annulus around each X-ray source. The $5''$ inner radius was used in order to avoid the presence of true counterparts in the background distribution, while we chose a $30''$ outer radius to avoid true counterparts of other X-ray sources. In the case of X-ray pairs the outer radius could contain the counterpart of a nearby X-ray source, but every annulus contains a number of background sources large enough (~ 80 sources in i -, ~ 70 in K - UltraVISTA and ~ 45 in the $3.6 \mu\text{m}$ band SPLASH catalog, respectively) to avoid significant effects of contamination.

$q(m)$ is the expected distribution function (normalized to 1) for the magnitude, m , of the real optical counterpart candidates. To compute $q(m)$ we first assumed an universal optical/infrared magnitude distribution for all X-ray sources, thus neglecting any influence of the X-ray flux on $q(m)$. Then we computed $q'(m)$ as the number of sources with magnitude m within $1''$ of the X-ray source, minus the expected number of background sources with magnitude m in a $1''$ circle. The $1''$ radius maximizes the statistical significance of the over density around X-ray sources: a smaller radius would give a higher Poissonian noise, while a larger radius would increase the number of background sources. We report in Figure 3.1 the observed magnitude distribution of the objects in the three catalogs (i , K , and $3.6 \mu\text{m}$) within a radius of $1''$ around each X-ray source (black solid histogram), together with the expected distribution of background objects in the same area (black dashed histogram). $q'(m)$ is, as we said, the difference of these two curves and is plotted in red. The background distribution has a peak at low magnitudes in the i band because the optical catalog has a magnitude limit ~ 3 - 4 magnitudes fainter than the K and $3.6 \mu\text{m}$ catalogs, but sources with $i_{AB} > 26$ are, on average, less likely to be reliable counterparts of X-ray sources. Finally, we normalized $q'(m)$ in order to have $q(m) = \text{const} \times q'(m)$ such that $\int_{-\infty}^{+\infty} q(m) dm = 1$. The normalization value const is here assumed 0.92, slightly larger than in C12, where it was $\text{const} = 0.9$. This normalization choice is the best trade-off between completeness and reliability, i.e., it allows us to find a larger number of counterparts without significantly increasing the number of expected spurious detections.

Finally, $f(r)$ is the probability distribution function of the positional errors, assumed to be a two-dimensional Gaussian, with $\sigma = \sqrt{\sigma_X^2 + \sigma_{opt}^2}$. σ_X is the X-ray positional uncertainty, computed as described in Section 2.3.1, while σ_{opt} is the positional uncertainty in the optical/IR band. We adopted the same optical positional uncertainties of C12, i.e., $0.2''$ for the K band (McCracken et al. 2010), $0.3''$ for the i band (Capak et al. 2007) and $0.5''$ for the $3.6 \mu\text{m}$ band (Sanders et al. 2007).

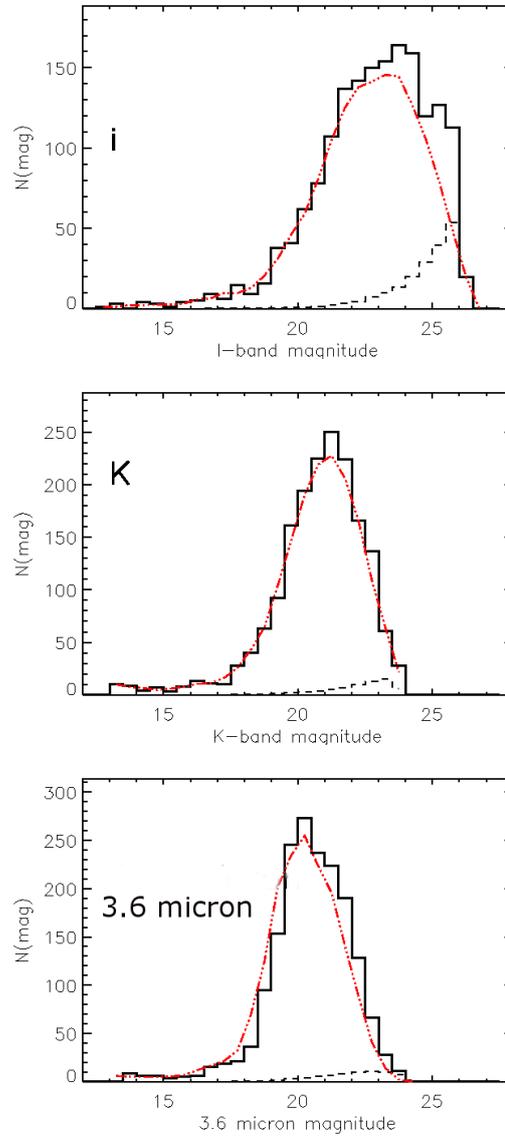


Figure 3.1: Observed AB magnitude distribution of all the i -band, K -band and $3.6 \mu\text{m}$ band (from top to bottom) objects within $1''$ around the Legacy sources (solid black histogram) together with the expected distribution of background objects ($n(m)$, black dashed histogram). The difference between the two distributions (red dashed-dotted line) is the expected magnitude distribution of counterparts ($q'(m)$, smoothed for plotting purposes). There is an excellent agreement between these plots and those of C12 for C-COSMOS.

3.2.2 Choice of the Likelihood Ratio threshold

The LR ratio technique is based on three parameters that define the whole sample: the likelihood ratio threshold (L_{th}), i.e. the minimum value of likelihood ratio that a possible match needs to be accepted, the reliability (R) and the completeness (C).

A fundamental step in the optical counterparts identification is the choice of the best likelihood threshold value (L_{th}) for LR, in order to make a distinction between real and spurious identifications. L_{th} should not be too high, otherwise we would miss too many real identifications and consequently reduce the sample completeness, but L_{th} has also to be high enough to keep the number of spurious identifications low and the reliability of the identification high.

Reliability describes the possibility of having multiple candidate counterparts for the same X-ray source. For a given optical object j , the reliability R_j of being the correct counterpart is

$$R_j = \frac{(LR)_j}{\sum_i (LR)_i + (1 - Q)}, \quad (3.2)$$

where the sum is over the set of all optical candidate counterparts and $Q = \int_m q(m) dm$ is normalized in order to be equal to the ratio between the number of X-ray sources identified in the given optical/infrared band and the total number of sources in the X-ray sample. The reliability R_k for each X-ray source is the sum of the reliabilities R_j of all the possible counterparts of the k -th X-ray source and it is by definition equal to 1. The reliability parameter (R) for the whole sample, instead, is defined as the ratio between the sum of all the reliabilities of the candidate counterparts and the total number of sources with $LR > L_{th}$, i.e. $R = N_{ID} / N_{LR > L_{th}}$.

The completeness parameter (C) of the total sample is defined as the ratio between the sum of the reliability of all the sources identified as possible counterparts and the total number of X-ray sources ($C = N_{ID} / N_X$).

In C12 and in B07, L_{th} was defined as the likelihood ratio where the quantity $(C + R)/2$ is maximized. In the *Chandra COSMOS-Legacy* survey $(C+R)/2$ is almost flat at $L_{th} \geq 0.5$, as can be seen in Figure 3.2, where we plot C , R and $(C+R)/2$ versus L_{th} for the optical catalog, so we select a L_{th} value of 0.5 for both i and K -bands. Given the lower spatial resolution of the $3.6 \mu\text{m}$ data, we chose a slightly higher threshold $L_{th} = 0.7$ in this band, to reduce the number of spurious identifications. The corresponding sample completeness and reliability for the catalogs in the three bands are shown in Table 3.2: as a general trend, both C and R grow moving from optical to infrared, due to the stronger relation of K or $3.6 \mu\text{m}$ magnitudes with the X-ray flux (Mainieri et al. 2002, Brusa et al. 2005).

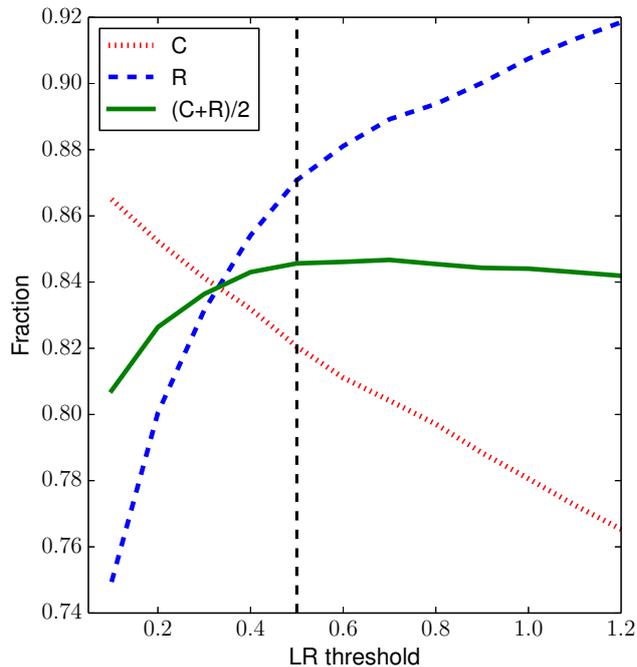


Figure 3.2: Completeness (C , red dotted line), reliability (R , blue dashed line) and $(C+R)/2$ (green solid line) at given values of L_{th} matching the optical catalog with new Legacy sources. The dashed black line shows the selected threshold in this band, $L_{th}=0.5$

As a final remark, it is worth noticing that the values of C and R we obtained for the new *Chandra* COSMOS Legacy dataset are all in good agreement with those obtained for C-COSMOS ($C=0.85$ and $R=0.88$ for i , $C=0.90$ and $R=0.92$ for K , and $C=0.96$ and $R=0.96$ for $3.6 \mu\text{m}$), and are higher than those of XMM-COSMOS because of the better *Chandra* positional accuracy (angular resolution of $\simeq 0.5''$ and $\simeq 6''$ for *Chandra* and XMM-*Newton* full width half maximum, FWHM, respectively).

Band	C	R	LR_{th}
i	0.82	0.87	0.5
K	0.86	0.93	0.5
$3.6 \mu\text{m}$	0.92	0.97	0.7

Table 3.2: Completeness (C) and Reliability (R) for each optical/IR band.

3.3 X-ray source identification results

In this section we show the procedure adopted to define the final counterparts after performing the likelihood ratio analysis. As in C12 and in XMM-COSMOS (B07, B10), the X-ray sources have been divided into four classes, based on their

counterparts associations:

1. Secure. Sources with only one counterpart with $LR > LR_{th}$. The vast majority of counterparts belongs to this class. 2214 of the 2273 new *Chandra COSMOS-Legacy* sources ($\simeq 97\%$) have been classified secure after the whole identification procedure (see Table 3.4).

2. Ambiguous. Sources with more than one counterpart above the threshold. 24 of the 2273 new *Chandra COSMOS-Legacy* sources have been classified as ambiguous after the whole identification procedure.

3. Subthreshold. Sources with one or more possible counterparts with $LR < LR_{th}$ within $5''$ from the X-ray centroid. 4 of the 2273 new *Chandra COSMOS-Legacy* sources have been classified as subthreshold after the whole identification procedure.

4. Unidentified. Sources with no counterpart, even below the threshold, within $5''$ from the X-ray centroid. 31 of the 2273 new *Chandra COSMOS-Legacy* sources have been classified as unidentified after the whole identification procedure.

We show a few examples of objects belonging to each of these classes in Figure 3.3.

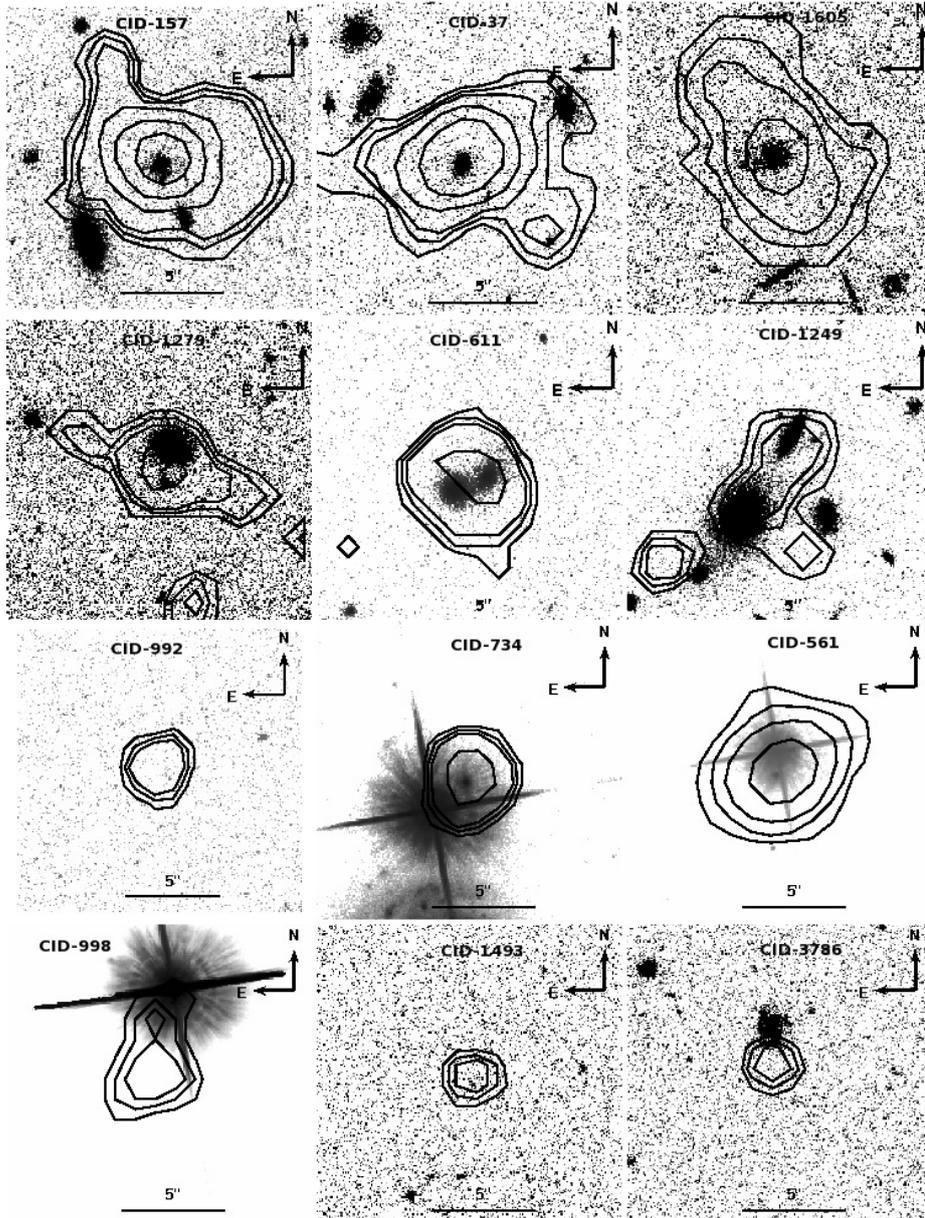


Figure 3.3: Examples of different classes of optical identification: from top to bottom, ACS images of secure, ambiguous, unidentified and subthreshold counterparts are shown. The X-ray contours are overlaid. Ambiguous counterparts have two optical sources inside the X-ray contour. Unidentified sources: in the CID-992 position no optical emission is detected, while CID-734 and CID-561 are very close to a bright star and is not possible to estimate their photometry. Subthreshold: CID-998 X-ray centroid is distant $\sim 5''$ from a bright star, CID-1493 candidate counterpart is a faint optical source with $i_{AB}=26.4$ and CID-3786 candidate counterparts is distant $\sim 3''$ from the X-ray centroid. Taken from C12.

3.3.1 Identification rates

First of all, we run the LR technique with the K -band catalogs, using both the UltraVISTA and the CFHT catalogs we described in Section 3.1: the positional error for the K -band sources has been fixed to $0.2''$, as in C12. We first matched our sources with those in the UltraVISTA area, assuming $L_{th}=0.5$, and we obtained 1690 counterparts with $LR > L_{th}$, while another 117 sources have a counterpart with $LR < L_{th}$. 583 *Chandra COSMOS-Legacy* sources have therefore no secure UltraVISTA counterpart (117 sources with a counterpart with $LR < L_{th}$ and 466 with no UltraVISTA counterpart). In the CFHT catalog, 379 of these 583 sources have at least one counterpart with $LR > L_{th}$: as a final summary, 2069 sources (92.2% of the X-ray sample inside the composite K -band field of view) have at least one counterpart with $LR > L_{th}$ in the K -band.

We then run the LR technique with the i -band Subaru catalog we described in Section 3.1. The adopted positional error for the i -band sources is $0.3''$, as in C12. At a L_{th} value of 0.5, there are 1594 Legacy sources (70.1%) with secure or ambiguous Subaru i -band counterpart with $SNR > 5$ and $LR > L_{th}$, while another 69 sources (3.0% of the whole sample) have one or more counterparts with $LR < L_{th}$.

Finally, we matched our X-ray catalog with the $3.6 \mu\text{m}$ catalog: the positional error for the $3.6 \mu\text{m}$ sources has been fixed to $0.5''$, as in C12. We first matched the X-ray catalog with the SPLASH catalog: at a L_{th} value of 0.7, there are 2046 Legacy sources with at least one SPLASH counterpart with $SNR > 5$ and $LR > L_{th}$ (91.1% of 2246 X-ray sources inside the SPLASH field of view), while another 41 sources (1.8%) have one or more counterparts with $LR < L_{th}$. 227 *Chandra COSMOS-Legacy* sources have therefore no secure SPLASH counterpart (41 sources with a counterpart with $LR < L_{th}$ and 186 with no UltraVISTA counterpart). We then matched these 227 sources, with the Sanders catalog, and we found another 125 sources with $LR > L_{th}$. Therefore, combining the two $3.6 \mu\text{m}$ catalogs 2171 sources (95.5% of the whole sample) have at least one counterpart with $LR > L_{th}$.

The identification rates in all bands are in very good agreement with those reported in C12.

3.3.2 Counterparts with $2 < SNR < 5$

In order to complete our identification of optical counterparts, we looked for i and K -band counterparts with $2 < SNR < 5$; we did not perform this analysis in the $3.6 \mu\text{m}$ band, due to its lower spatial resolution. There are 157 X-ray sources with no counterpart with $SNR > 5$ in i -band but with at least one counterpart with $2 < SNR < 5$ in i -band. Of these sources, 148 have at least one counterpart with $LR > L_{th}$, while the other 9 have $LR < L_{th}$. There are also 18 X-ray sources with no counterpart with $SNR > 5$ in K -band but at least one counterpart with $2 < SNR < 5$ and $LR > L_{th}$ in the composite UltraVISTA/CFHT K -band, and one source with one counterpart with $2 < SNR < 5$ and $LR < L_{th}$ in the composite UltraVISTA/CFHT K -band.

To complete our i -band catalog, especially at $i_{AB} < 20$, where the Subaru catalog is saturated, we also matched our K and $3.6 \mu\text{m}$ secure counterparts with the CFHT and SDSS i -band catalogs, with maximum separation $d_{ik}=1''$: we found i -band magnitude for 301 X-ray sources (13.2% of the whole sample).

We report in Table 3.3 the number of counterparts in the i , K and $3.6 \mu\text{m}$ bands, first using only sources with $\text{SNR} > 5$, then introducing also sources with $2 < \text{SNR} < 5$. As can be seen, the fraction of sources with a secure counterpart is excellent in every band (79.8% in the i -band, 85.1% in the K -band and 90.1% in the $3.6 \mu\text{m}$ band), but the number of ambiguous sources, i.e. of sources with more than one possible counterpart in an optical or IR band, is significant, especially in the i and K -bands, where $\simeq 9\%$ and $\simeq 7\%$ of the *Chandra COSMOS-Legacy* sources are ambiguous. In the next section, we describe the approach chosen to significantly reduce the number of ambiguous counterparts.

3.3.3 Solving the cases of ambiguous sources

As previously explained, $\simeq 8\%$ of X-ray sources have been flagged as “ambiguous” in both i and K -band. We developed the following procedure to choose the correct counterpart: the main assumption is to use secure counterparts in one band to solve ambiguities in the other one. We started by matching i and K -band counterparts, and then we introduced those in the $3.6 \mu\text{m}$. For each source we run a four different checks: if one was not satisfied, we moved to the following one.

1. Counterparts in i and K -band have $R > 0.9$ in both bands. We kept these counterparts as the good ones and we rejected any other counterpart of the same X-ray source. The largest part of ambiguities ($\simeq 50\%$) is solved in this first step.
2. There is a counterpart with $R > 0.9$ in one band and the distance between this counterpart and only one counterpart in the other band is $d_{ik} < 1''$. We kept these counterparts as the good ones and we reject any other counterpart of the X-ray source. Other $\simeq 25\%$ of ambiguities is solved in this step.
3. The two counterparts with largest R have $d_{ik} < 1''$. We kept these counterparts as the good ones and we rejected any other counterpart of the same X-ray source. After this step, less than 15% of the original ambiguous identifications are still ambiguous.
4. There is a secure $3.6 \mu\text{m}$ counterpart within $1''$ from the X-ray source and one of the counterparts in i or K -band have distance from the $3.6 \mu\text{m}$ counterpart $< 1''$.

The number of $3.6 \mu\text{m}$ ambiguous identifications is lower than in the i and K -band ones, because of the *Spitzer* lower spatial resolution. For this reason, to solve ambiguities in the IRAC band we adopted a simplified procedure, where we first kept, if present, the sources with $R > 0.9$ and rejected the other candidates. Then, for the smaller fraction of sources still ambiguous ($\simeq 15$ sources), we looked for a secure counterpart in the optical or K -band within $1''$. With this procedure, no counterpart in the $3.6 \mu\text{m}$ band is flagged as ambiguous.

During the analysis of ambiguous sources, we did not use deblending techniques.

3.3.4 Final results for optical counterparts

We finally performed a complete visual check of all the X-ray sources and their counterparts; we found a further group of visually good counterparts ($\simeq 2\%$ of all the secure counterparts in the optical catalog, and $\simeq 1\%$ of all the secure counterparts in the K -band catalog), which were not previously found mainly because they had $\text{SNR} < 2$. All these new counterparts have separation from the X-ray centroid smaller than $1''$ and already have a counterpart detected with the LR ratio technique in at least one of the other two optical/IR bands.

We report in Table 3.4 the final number of counterparts in the i , K and $3.6 \mu\text{m}$ bands, after the resolution of ambiguous counterparts and the visual inspection. 2214 sources (97.4%) have now a secure counterpart, i.e., one counterpart above $\text{LR} > L_{th}$ with all the possible others above threshold rejected after our procedure and visual inspection: this result is comparable with the one obtained in CDF-S (96.8%, Xue et al. 2011) and better than the one in Stripe 82 ($\simeq 80\%$ in the optical SDSS band, $\simeq 59\%$ in the UKIDSS near-IR band and $\simeq 65\%$ in the WISE $3.6 \mu\text{m}$ band, LaMassa private communication). Other 24 sources (1.1%) have been instead classified as ambiguous, and only four sources are classified as subthreshold. Finally, 31 sources (1.4%) have no counterpart in any of the optical or infrared bands. These sources are candidate obscured or high- z AGN, or both; however, it is also worth noticing that a fraction of 0.3% of *Chandra COSMOS-Legacy* sources (i.e. $\simeq 12$ in full, 9 in soft and 8 in hard band, assuming a threshold of 7 net counts) is expected to be spurious at the likelihood threshold used in the X-ray catalog (see section 2.3.1).

We also point out that the fraction of counterparts we found is consistent with the one obtained by Hsu et al. (2014, $\simeq 96\%$) using a slightly different matching method, based on Bayesian statistics, which also takes into account both the magnitude and the source position, as the LR ratio technique, and in addition works simultaneously on multiple bands. We decided not to use this technique for consistency with the C-COSMOS analysis and also because the Hsu et al. (2014) method, although used on the CDF-S, becomes significantly more effective than the one we used only on very large area surveys, with millions of potential counterparts, a significant fraction of which with non-negligible positional error and without homogenous multiwavelength coverage.

Class	i_p SNR>5	i_p SNR>2	i_{other}	i_{whole} SNR>2	$f_{i,whole}$ SNR>2	K SNR>5	K SNR>2	f_K SNR>2	$3.6 \mu\text{m}$ SNR>5	$3.6 \mu\text{m}$ SNR>2	$f_{3.6}$ SNR>2
Secure	1465	1581	232	1813	79.8%	1923	1935	85.1%	2049	2049	90.1%
Ambiguous	129	161	40	201	8.8%	148	154	6.8%	125	125	5.5%
Subthreshold	69	78	29	107	4.7%	53	54	2.4%	37	37	1.6%
Unidentified	610	453	–	152	6.7%	149	130	5.7%	62	62	2.7%

Table 3.3: Number of X-ray sources identified in each band and in total, for counterparts with $\text{SNR} > 5$ and adding counterparts with $2 < \text{SNR} < 5$, and fraction f of sources with respect to the whole survey, after the contribution of sources with $2 < \text{SNR} < 5$ has been taken in account. i_p identifies sources with Subaru i -band magnitude, i_{other} identifies sources with CFHT or SDSS i -band magnitude and i_{whole} summarizes all sources with i -band magnitude.

Class	i	f_i	K	f_K	3.6 μm	f_K	Total	f_{total}
Secure	2100	92.4%	2119	93.2%	2171	95.6%	2214	97.4%
Ambiguous	17	0.7%	9	0.4%	0	0%	24	1.1%
Subthreshold	92	4.0%	28	1.3%	36	1.6%	4	0.1%
Unidentified	64	2.8%	117	5.1%	66	2.7%	31	1.4%

Table 3.4: Final number of X-ray sources identified in each band and in total. “Total” is the number of sources with an identification in one or more bands.

3.3.5 Sources in C-COSMOS with updated optical counterpart

676 of the 1743 C-COSMOS sources have been observed again during the *Chandra COSMOS-Legacy* observations, thus having now double *Chandra* exposure and therefore improved positional accuracy err_{pos} , given that $err_{pos} \propto C_S^{-0.5}$, where C_S are the source net counts (see section 2.3.1 for further details). We performed the same LR technique we used on the 2273 new *Chandra COSMOS-Legacy* sources to check if the potential slight change in the X-ray position of the source due to the larger exposure and/or the use of different catalogs of optical/IR counterparts with respect to C12 implied the identification of a different (or new) optical/IR counterpart. We found that 9 (1.3%) sources have a different optical/IR counterpart, while 6 (1%) sources that had no optical counterpart in C-COSMOS now have a secure one. For all these sources, a new photometric redshift (see Section 3.5.2) has also been computed.

We also run the LR ratio identification procedure over the whole C-COSMOS sample with the new UltraVISTA and SPLASH information, that were not available at the time of C12. We found 52 sources (3% of the whole C-COSMOS sample) with no CFHT K -band information but with a secure UltraVISTA counterpart. We also found 49 sources (2.8% of the whole C-COSMOS sample) with no 3.6 μm IRAC information from the Sanders catalog, but with a secure SPLASH counterpart.

We report this updated information, together with the newly available redshifts, in the new catalog of optical counterparts of the whole *Chandra COSMOS-Legacy* survey.

3.3.6 Sources in XMM-COSMOS with updated optical counterpart

866 new *Chandra COSMOS-Legacy* sources have a counterpart in the XMM-COSMOS catalog (Cappelluti et al. 2009) within $10''$. 104 of these sources have a different optical counterpart than in the XMM-COSMOS optical catalog (B10), mainly because of the better angular resolution of *Chandra* compared to XMM-*Newton*, but also because we used different optical catalogs; also, a significant fraction of these sources was flagged as ambiguous in B10. It is also worth noticing that 36 of these *Chandra* sources are actually part of pairs of counterparts of the same XMM-*Newton* source (once again because of the better *Chandra* angular resolution).

We report the XMM-COSMOS identification number of all *Chandra COSMOS-Legacy* sources in the new catalog we developed, available in the COSMOS repository and online; a summary of the catalog is also reported in Appendix C.

3.3.7 Radio-detected counterparts

We matched the *Chandra COSMOS-Legacy* catalog with the JVLA-COSMOS 3 GHz catalog (RMS=4.8 $\mu\text{Jy beam}^{-1}$; Novak et al., in prep.), using a maximum separation distance of 3". 1308 of the 4016 (32.4%) *Chandra COSMOS-Legacy* sources have a radio counterpart down to SNR=5. We tested if the radio detected sources show different optical/IR and X-ray properties than those non-radio detected; we found that radio-detected sources are brighter in optical and IR, they are at lower redshifts, and a larger fraction of them is obscured (and this last result is confirmed by the HR, which is slightly higher for the *Chandra COSMOS-Legacy* sources with radio counterpart). Although beyond the purpose of this thesis, it is worth noticing that the COSMOS field was already analyzed in the radio at 1.4 GHz (RMS= 10 $\mu\text{Jy beam}^{-1}$; Schinnerer et al. 2007), and a joint analysis of sources detected at both frequencies could allow us to estimate the radio spectral index α .

We also estimated the radio emission for optically selected obscured and unobscured AGN: 284 Type 1 AGN (29% of the whole sample of Type 1 AGN) have a radio counterpart, with mean (median) radio flux 3.18×10^{-4} (2.77×10^{-5}) Jy at 3 GHz, while 994 Type 2 AGN (36% of the whole sample of Type 2 AGN) have mean (median) radio flux 3.19×10^{-4} (3.19×10^{-5}) Jy at 3 GHz. On the basis of a KS-test we cannot reject the hypothesis that the two distributions are different with significance >95%.

Finally, we estimated the radio loudness of the sources using the ratio $\log(R_X) = \log(\nu L_\nu(5 \text{ GHz})/L_X(2-10 \text{ keV}))$ (Terashima & Wilson 2003), where $L_X(2-10 \text{ keV})$ is computed rest-frame, with no absorption correction: we found that the mean (median) $\log(R_X)$ is -3.32 ± 0.68 (-3.37). Moreover, 733 of the 848 ($\simeq 86\%$) X-ray sources with a radio detection that have an optical/IR SED consistent with a galaxy one (see section 3.5.5 for further details) have $\log(R_X) > -3.8$, a value that can be explained only assuming the presence of an AGN (Mezcua et al. 2013). We will further discuss in sections 3.6.1 and 3.6.4, the fact that a significant fraction of sources where the galaxy contribution is more significant in the optical/IR are actually revealed as AGN thanks to the X-ray information.

3.4 X-ray to optical positional offset and optical properties versus identification class

We present in Figure 3.4 the X-ray to optical/IR separation and the optical/IR magnitude, in *i* (Figure 3.4a, cyan), *K* (Figure 3.4b, orange) and 3.6 μm band (Figure 3.4c, red). More than 90% of the secure counterparts have a distance from the X-ray source smaller than 1": the mean (median) value of the distance from the X-ray source is $0.70 \pm 0.50''$ ($0.59''$) for the *i*-band counterparts, $0.67 \pm 0.48''$ ($0.56''$) for the *K*-band counterparts and $0.69 \pm 0.49''$ ($0.58''$) for the 3.6 μm counterparts, a result in agreement with the one obtained during the astrometric correction process of the X-ray observations described in section 2.2.1. The distribution is instead wider for subthreshold sources, where the mean (median) distance from the X-ray source is $1.24 \pm 0.70''$ ($1.17''$) in *i*-band, $1.64 \pm 0.67''$ ($1.56''$) in *K*-band and $1.60 \pm 0.96''$ ($1.57''$) in 3.6 μm band. Moreover, subthreshold counterparts are on average 1.5-2 magnitudes fainter than the secure counterparts (see Table 3.5 for a summary).

Both the fainter magnitudes and the larger X-ray to optical/IR separations are consistent with the subthreshold counterparts being less reliable than the secure ones (see also B10).

Band	Secure		Subthreshold	
	mean	median	mean	median
<i>i</i>	22.8	23.0	24.5	25.2
<i>K</i>	20.9	21.0	22.1	22.8
3.6 μm IRAC	20.4	20.5	22.4	22.8

Table 3.5: Mean and median magnitude values for secure and subthreshold counterparts in each of the three bands used in our analysis.

We also analyzed the distribution of the distance between optical and infrared counterparts for the same X-ray source: for secure counterparts (in both bands), the mean (median) distance between *i* and *K* counterparts is $0.17 \pm 0.38''$ ($0.07''$) and that between *i* and 3.6 μm counterparts is $0.17 \pm 0.41''$ ($0.07''$). We do not report the distance between secure *K* and 3.6 μm counterparts because the vast majority of them come from the same catalog (Laigle et al. submitted), which contains both the UltraVISTA and the SPLASH magnitude information, and have therefore the same right ascension and declination. This small value in the separation between optical and *K*/IR counterparts is consistent with the fact that secure counterparts in different bands are actually the same source.

We studied the distribution of the difference between X-optical distances of the closest and the second closest possible counterpart of “ambiguous” identifications, and the distribution of the difference $|\text{Mag}_2 - \text{Mag}_1|$, where Mag_1 and Mag_2 are the magnitudes of the “ambiguous” identifications: here we define as “ambiguous” only the sources with no secure counterpart after running the procedure described in Section 3.3.3. As expected, for more than 75% of the ambiguous counterparts the difference between the distances of the two candidate counterparts from the X-ray source is smaller than $1''$, i.e. comparable with *Chandra* resolution ($\simeq 0.5''$). Similarly, for more than 70% of the ambiguous identifications the difference in magnitude between the two candidate counterparts is < 1 mag, therefore not allowing to select one of the sources as a secure counterpart.

3.5 Spectroscopic and photometric redshift distribution

3.5.1 Spectroscopic redshifts

We cross-correlated our optical counterparts with the master spectroscopic catalog available within the COSMOS collaboration (Salvato et al. in prep.), which contains $\simeq 80,000$ spectroscopic redshifts. The catalog includes redshifts from SDSS (DR12), VIMOS (zCOSMOS: Lilly et al. 2007, Lilly et al. 2009; VUDS: Le Fevre et al. 2015), MOSFIRE (Scoville et al. in prep.; MOSDEF: Kriek et al. 2015), several years of DEIMOS observations from multiple observing programs (e.g. Kartaltepe et al. 2010, Hasinger et al. in prep.), IMACS (Trump et al. 2007, 2009a), Gemini-S (Balogh et al. in prep.), FORS2 (Comparat et al. 2015), FMOS (Silverman et al.

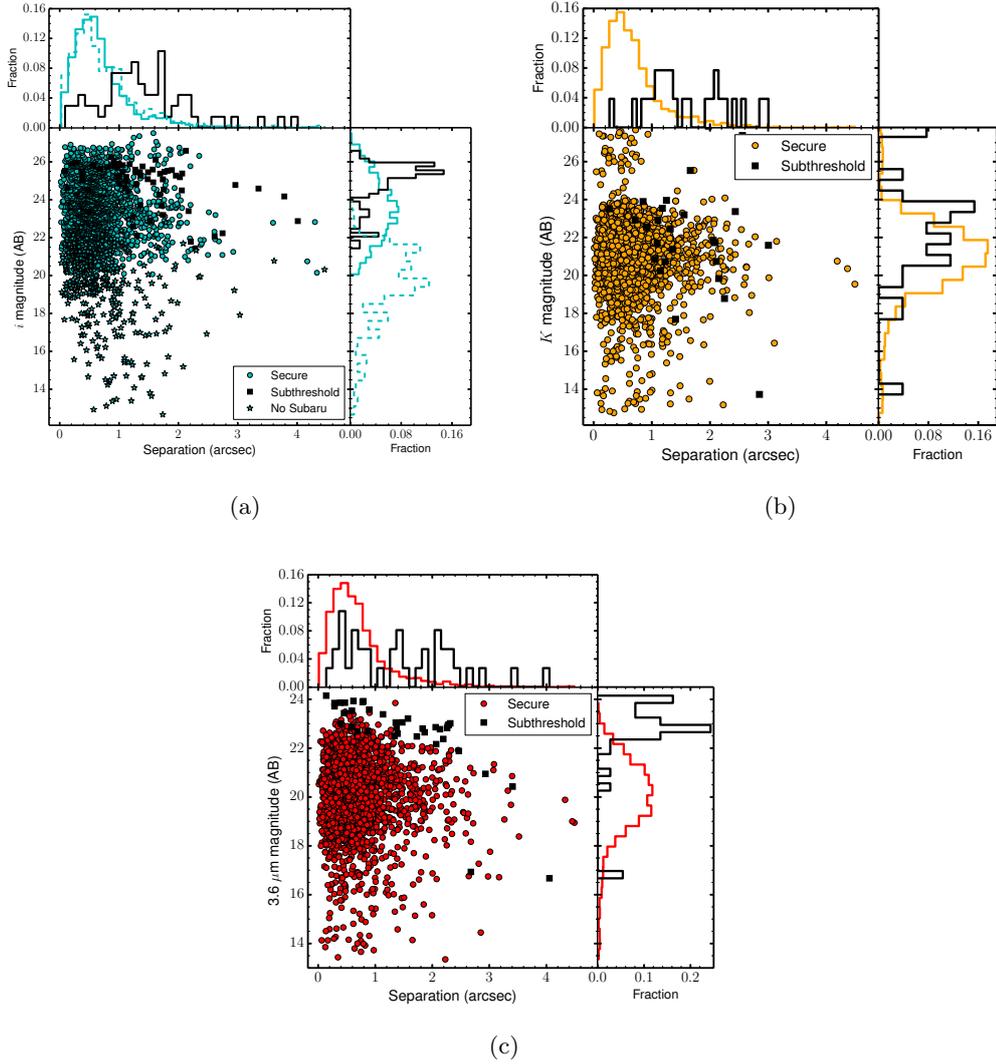


Figure 3.4: Separation between X-ray and optical (IR) positions for (a) *i*-band, (b) *K*-band, and (c) 3.6 μm bands. Secure counterparts are shown in cyan (*i*-band), orange (*K*-band) and red (3.6 μm band), while sub-threshold counterparts are shown in black. Sources with *i*-band magnitude from CFHT or SDSS are plotted as cyan stars. Histograms of separation and magnitude are shown in each of the three plots. Histogram of separation and magnitude for *i*-band sources with CFHT or SDSS information are showed with a dashed line in (a).

2015) and PRIMUS (Coil et al. 2011), plus a negligible number of sources provided by other smaller contributions.

The redshift confidence from the various contributors has been translated into the classification as defined in zCOSMOS: 688 of the 2273 *Chandra COSMOS-Legacy* sources have a reliable spectroscopic redshift, i.e. with confidence ≥ 3 (spectroscopic accuracy $> 99.5\%$, estimated using those objects observed more than once, and verifying if their redshifts were in agreement). However, our sample contains

also 188 sources with a less reliable spectroscopic redshift (spectroscopic accuracy $<99.5\%$) but with the photometric redshift z_{phot} specifically provided for this catalog (see Section 3.5.2) such that $\Delta z = \frac{|z_{spec} - z_{phot}|}{1 + z_{spec}} < 0.1$. For these sources, we adopted as final value the spectroscopic redshift one. In summary, we provide a spectroscopic redshift for 876 sources ($\simeq 39\%$ of the sample).

3.5.2 Photometric redshifts

For 1306 sources, we can provide only photometric redshifts. Photometric redshifts have been produced following the same procedure described in detail in Salvato et al. (2011), without any further training sample. Depending on the X-ray flux of the sources and on the morphological and photometric (e.g variability) properties of the counterpart, specific priors and libraries of templates (including galaxies, AGN/galaxy hybrids, AGN and QSOs) have been adopted, and the best fit has been found through a χ^2 minimization, using the publicly available code LePhare (Arnouts et al. 1999, Ilbert et al. 2006). Using the secure spectroscopic subsample as reference (i.e. only those sources with spectroscopic accuracy $>99.5\%$), we found an accuracy of $\sigma_{\Delta z / (1 + z_{spec})} = 0.03$, with a fraction of outliers $<8\%$ (54 of 688), consistent with what was already found for C-COSMOS (78 outliers out of 1020 secure spec-z), using a different spectroscopic sample. Breaking down the sample (see Figure 3.5), for the 474 sources that are brighter than $i_{AB} = 22.5$ the accuracy, estimated using the normalized median absolute deviation $\sigma_{NMAD} = 1.48 \times \text{median}(\|z_{spec} - z_{phot}\| / (1 + z_{spec}))$, is $\sigma_{NMAD} = 0.012$ with 5.7% of outliers. For the fainter sample of 176 sources, where the number of the available photometric bands decreases and the photometric errors increase, the accuracy decreases by a factor of ~ 3 ($\sigma_{NMAD} = 0.034$) and the number of outliers increases by the same factor (15.3%). The whole sample has $\sigma_{NMAD} = 0.018$ with 7.9% of outliers.

The photo-z computation provides for each source a probability distribution function (Pdz), which gives the probability of a source to be at a given redshift bin: the nominal photo-z value is actually the maximum of this Pdz. The integrated area of the Pdz on all redshift bins is by definition equal to 1. At all redshifts, the agreement between the distribution of the nominal values of the photometric redshifts and the average distribution of the Pdz is very good. However, using the Pdz instead of just the photo-z nominal value allows to perform a much more statistically thorough analysis (Georgakakis et al. 2014). *Chandra COSMOS-Legacy* Pdz-s are already being used in the AGN clustering estimate at high redshift (Allevato et al. in prep.) and in the space density computation at $z > 3$, which we extensively describe in chapter 4.

3.5.3 Redshift summary

From now on, we will talk about the whole *Chandra COSMOS-Legacy* survey, i.e., of both the new dataset described so far, together with the old C-COSMOS sources.

The total number of new *Chandra COSMOS-Legacy* sources with a redshift, either spectroscopic or photometric, is 2182, i.e. 96% of the entire sample. In C-COSMOS, 1695 of the 1743 X-ray sources have a redshift (i.e. 97.3%), 1201 of

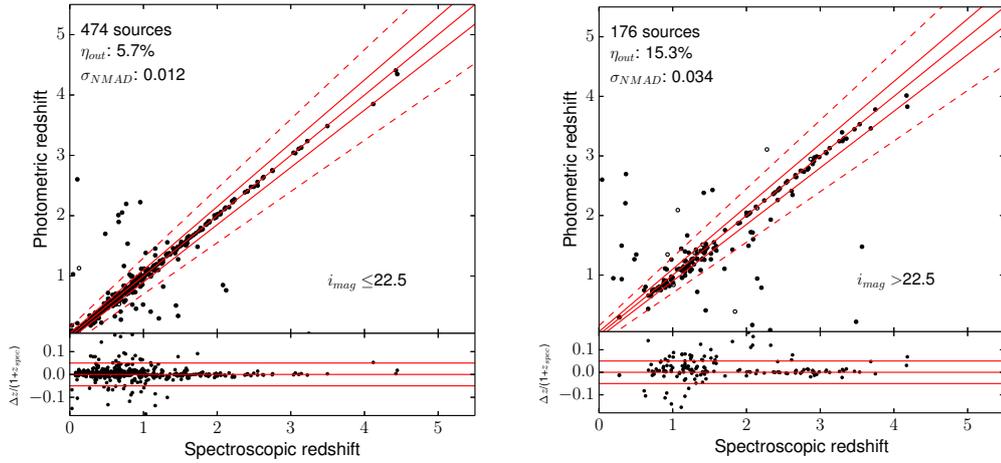


Figure 3.5: Photometric redshifts compared to the secure spectroscopic redshifts, for sources brighter (left) and fainter (right) than $i_{AB}=22.5$. Open circles represent sources for which there is at least a second significant peak in the redshift probability distribution. Red solid lines correspond to $z_{\text{phot}} = z_{\text{spec}}$ and $z_{\text{phot}} = z_{\text{spec}} \pm 0.05 \times (1 + z_{\text{spec}})$, respectively. The dotted lines limit the locus where $z_{\text{phot}} = z_{\text{spec}} \pm 0.15 \times (1 + z_{\text{spec}})$. Photo- z computed for the fainter sources are significantly worse in terms of both dispersion and fraction of outliers.

which have a reliable spectroscopic redshift, either secure (1022) or in agreement with the photo- z (179). With respect to the C12 catalog, we added new reliable spectroscopic redshift information to 286 C-COSMOS sources.

Summarizing, 3877 of the 4016 X-ray sources in the whole *Chandra COSMOS-Legacy* field have a redshift, i.e. $\simeq 96.5\%$ of the whole sample. We have a reliable spectroscopic redshift for 2078 of these sources (51.7% of the whole sample). As a comparison, $\simeq 91\%$ of the 740 sources in CDF-S (Xue et al. 2011) have either a spectroscopic or photometric redshift, and $\simeq 46\%$ have a reliable spectroscopic redshift, while $\simeq 30\%$ of the sources in Stripe 82 (LaMassa et al. 2013a and submitted) have a reliable spectroscopic redshift.

In Figure 3.6 we show the whole *Chandra COSMOS-Legacy* survey spectroscopic completeness: the completeness is $\geq 80\%$ up to a i -band magnitude (AB) of 21.5, then there is a linear decline in the completeness value, which is $\simeq 70\%$ at $i_{AB}=22.5$, $\simeq 50\%$ at $i_{AB}=23.5$, finally dropping below 25% only for sources fainter than $i_{AB}=24.5$. The relatively low completeness ($\simeq 50\%$) at bright magnitudes ($i_{AB}<16$) is due to the fact that most of the sources in this magnitude range are stars for which no spectrum was taken.

In Figure 3.7 we show the spatial distribution of the sources with spectroscopic redshift (black circles) over the whole *Chandra COSMOS-Legacy* area (red solid polygon): as can be seen, the spectroscopic follow-up of the *Chandra COSMOS-Legacy* sources has so far been focused mainly on the central C-COSMOS area (green solid line), while a significant fraction of sources in the external part of the *Chandra COSMOS-Legacy* field has not been observed yet. Therefore, the spectroscopic

completeness value of the whole survey will easily grow in the coming years, thanks to a dedicated program with *Keck*-DEIMOS (P.I.: G. Hasinger).

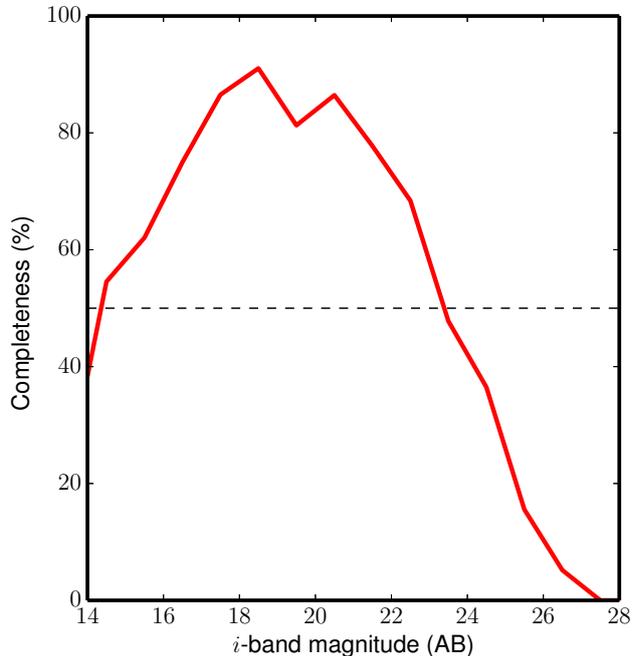


Figure 3.6: Spectroscopic completeness of the *Chandra COSMOS-Legacy* survey as a function of i_{AB} (red solid line). 50% completeness (black dashed line) is also plotted.

The redshift distribution of all *Chandra COSMOS-Legacy* sources with a redshift is plotted in Figure 3.8 (red solid line). The shape of the distribution is consistent with that of C-COSMOS (blue dotted line) and peaks at $z=1-2$. Many spikes are visible in the distribution (see, e.g., $z \simeq 1$, $z \simeq 1.3$), and these features are linked to large-scale structures in the COSMOS field (Gilli et al. 2009). The evidence of the most prominent spikes linked to the large-scale structures remains also when using only reliable spectroscopic redshifts (black dashed line).

Sources without optical identification

80 sources in the whole survey have no optical counterpart and lie inside the optical/IR field of view. We further analyzed these sources, because some of them could be obscured and/or high-redshift AGN (Koekemoer et al. 2004). We visually inspected all these objects, using both X-ray and optical/IR images, and we found that about 50% of the sources have no optical counterpart because of bad optical imaging, or because the possible counterpart is close to a very bright object (star or extended galaxy) and it is therefore undetected.

After this visual check, there are still 43 sources without an optical counterpart, but with a K -band or $3.6 \mu\text{m}$ IRAC counterpart, or with no counterpart at all.

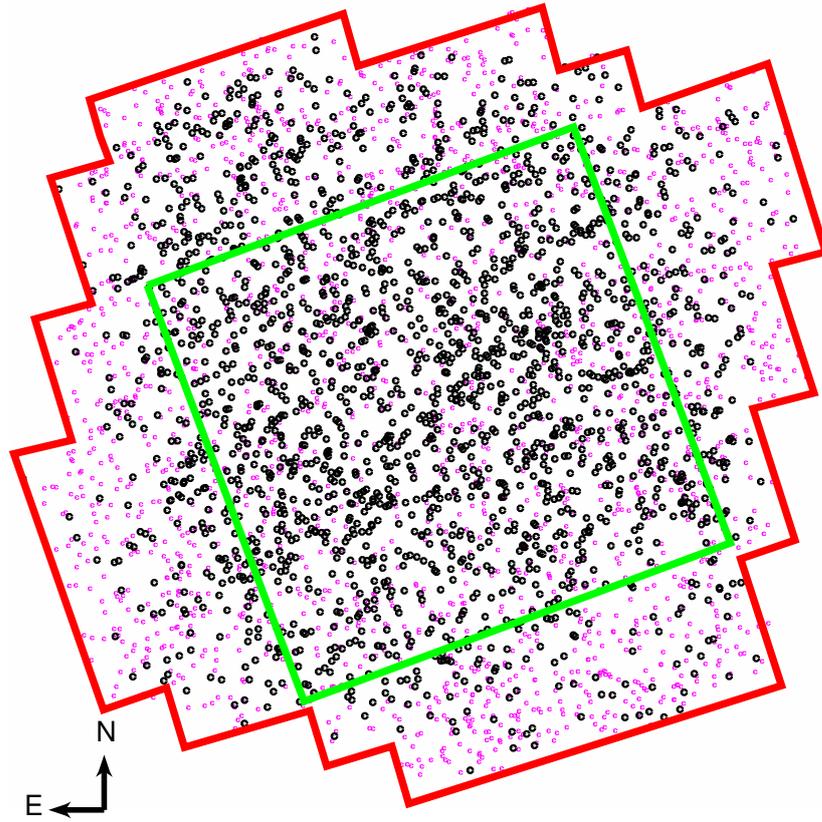


Figure 3.7: Sources with (black circles) and without (magenta circles) spectroscopic redshift in the *Chandra COSMOS-Legacy* area (red solid line). The C-COSMOS area is also plotted (green solid line). A significant fraction of sources in the external part of the field has not been spectroscopically followed-up yet.

19 of these sources have both a K -band and a $3.6 \mu\text{m}$ IRAC counterpart, 7 have only a K -band counterpart, 7 have only a $3.6 \mu\text{m}$ IRAC counterpart and 10 have no counterpart at all. Nine of these sources have no soft band detection, thus suggesting high obscuration rather than high redshift.

3.5.4 High-redshift sample

Chandra COSMOS-Legacy is also the X-ray survey on a single contiguous field with the highest number of high redshift sources: in the whole field there are 174 sources with $z \geq 3$ (78 of which have reliable spec- z), 27 sources with $z > 4$ (11 with reliable spec- z), 9 sources at $z > 5$ (2 with reliable spec- z) and 4 sources (3 of which are new, all 4 are photo- z) at $z > 6$. The source with the highest spectroscopic redshift, $z=5.3$, lies in a proto-cluster, where it is also the only X-ray source detected (Capak et al. 2011; Kalfountzou et al. in prep.). A detailed discussion of the sources at $z \geq 3$, together with an extended analysis of the space density of the X-ray sources in this redshift range, is presented in chapter 4.

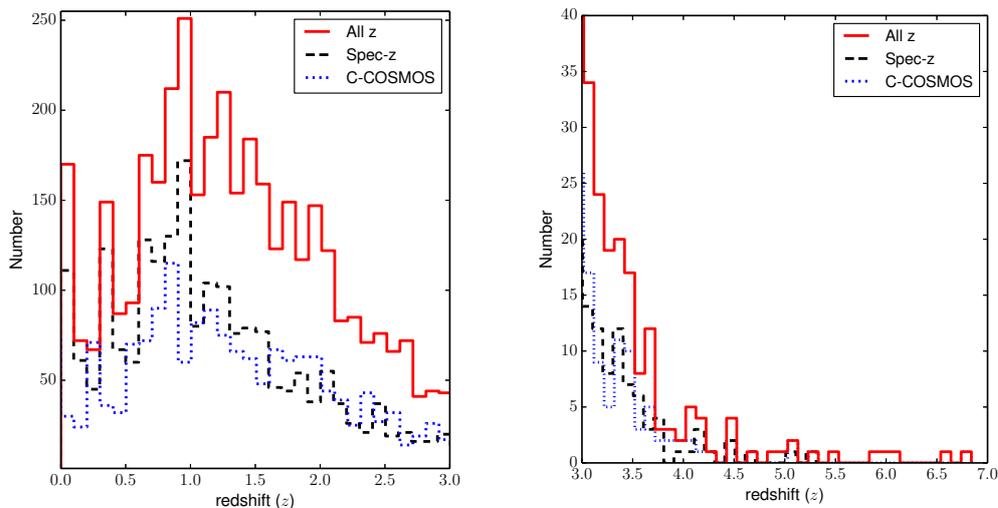


Figure 3.8: Redshift distribution of the whole *Chandra COSMOS-Legacy* (red solid line), of the sources with reliable spectroscopic redshifts (black dashed line), and of C-COSMOS spec+photo-z (blue dotted line) for the redshift range $z=0-3$ (left) and $z=3-7$ (right).

3.5.5 Spectroscopic and photometric types

We report in Table 3.6 the characterization of the sources by spectroscopic type (when available) for the new *Chandra COSMOS-Legacy* sources and for those in C-COSMOS. In the same table we also show how sources have been divided on the basis of the template which best fits the SED of the sources.

In the whole survey, there are 1692 sources with a reliable spectroscopic redshift and a spectral type information; 654 of these are new sources. Of these 1692 sources, 611 (36% of the spectroscopic sample with spectral type information) show evidence of at least one broad (i.e. with $\text{FWHM} > 2000 \text{ km s}^{-1}$) line in their spectra (BLAGN). There are 995 sources (59% of the spectroscopic sample with spectral type information) with only narrow emission lines or absorption lines. These objects are defined as “non broad-line AGN” (non-BLAGN). We do not make a further separation between star-forming galaxies and Type 2 AGN on the basis of the source spectra, because the large majority of these sources have low SNR spectra (mainly obtained just to determine the redshift) or are in an observed wavelength range which does not allow to use optical emission line diagnostic diagrams to disentangle in Type 2 AGN and star-forming galaxies.

Finally, the sample contains 86 spectroscopically identified stars (5% of the spectroscopic sample; see Wright et al. (2010) for a detailed analysis of the stars detected in C-COSMOS).

It is worth noticing that $\simeq 58\%$ of sources in the whole sample are still without spectroscopic type, thus the fractions of different spectral types may be not representative of the complete sample.

3855 sources (96.0% of the whole sample) have a photometric SED template

information. The largest part (64%) of these sources are fitted with a non-active galaxy, 9% are fitted with an obscured AGN template and 23% by a template with contribution by unobscured AGN. Finally, 121 sources, 3% of the whole sample, have been identified as stars on the basis of the photometric template.

We compared the spectroscopic and photometric classifications and we found that 82% of the sources with BLAGN spectral type have been fitted with an unobscured AGN template, while 97% of the non-BLAGN are fitted with either a galaxy template (74%) or with an obscured AGN template (23%). The lower agreement for BLAGN is not surprising, given that BLAGN SEDs can be contaminated by stellar light; this is particularly true for low-luminosity AGN (Luo et al. 2010; Elvis et al. 2012; Hao et al. 2014). Finally, 81 of the 86 spectroscopically identified stars (94%) are also photometric stars. As a general assumption, we use the spectroscopic type when available and if not the photometric one. In the following part of this work, we refer to BLAGN or unobscured sources as “Type 1”, and to non-BLAGN or obscured sources as “Type 2”.

It is worth noticing that in XMM-COSMOS (B10) there were $\simeq 50\%$ Type 1 sources and $\simeq 50\%$ Type 2 sources: *Chandra COSMOS-Legacy* reaches a flux limit three times deeper than XMM-COSMOS and therefore samples a larger fraction of obscured objects.

3.5.6 X-ray luminosity

In Figure 3.9 we show the X-ray luminosity versus redshift, in both soft (left, 2698 out of 4016 sources) and hard (right, 2354 sources) bands, for sources with $z > 0$ and $\text{DET_ML} > 10.8$. We converted fluxes into luminosities using the best redshift available, i.e. the spectroscopic one when available and the photometric redshift for the remaining sources; we used an X-ray spectral index of $\Gamma = 1.4$, to compute K -corrected luminosities. We did not apply any obscuration correction. In Figure 3.9, right, we also plot the $z-L_{2-10\text{keV}}$ curve of the knee of the AGN luminosity function (black dashed line), computed following the Flexible Double Power-Law (FDPL) model from Aird et al. (2015):

$$\log L^*(z) = 43.53 + 1.23 \times x + 3.35 \times x^2 - 4.08 \times x^3, \quad (3.3)$$

where $x = \log(1+z)$. As can be seen, we are able to sample with excellent statistics the luminosity range below the knee of the luminosity function, up to redshift $z \simeq 4$.

26% and 13% of the sources in the soft and hard band, respectively, have luminosities $L_X < 10^{42}$ erg s $^{-1}$, i.e. lower than the threshold which is conventionally used to separate clear AGN from galaxies with no or low nuclear emission, low-luminosity AGN or very obscured AGN (see, e.g., Basu-Zych et al. 2013; Kim & Fabbiano 2014; Civano et al. 2014; Paggi et al. submitted). This fraction, although not negligible, is significantly lower than the fraction of sources that have been fitted with a galaxy SED template (66% of all the sources). Therefore, the majority of sources fitted with a galaxy template are actually more likely to be obscured AGN rather than normal and starburst galaxies.

In Figure 3.10 we show the 0.5-2 keV rest-frame luminosity versus redshift distribution for *Chandra COSMOS-Legacy* (red circles), CDF-S 4 Ms (Xue et al. 2011) and Stripe 82 (LaMassa et al. 2013a and submitted). *Chandra COSMOS-Legacy*

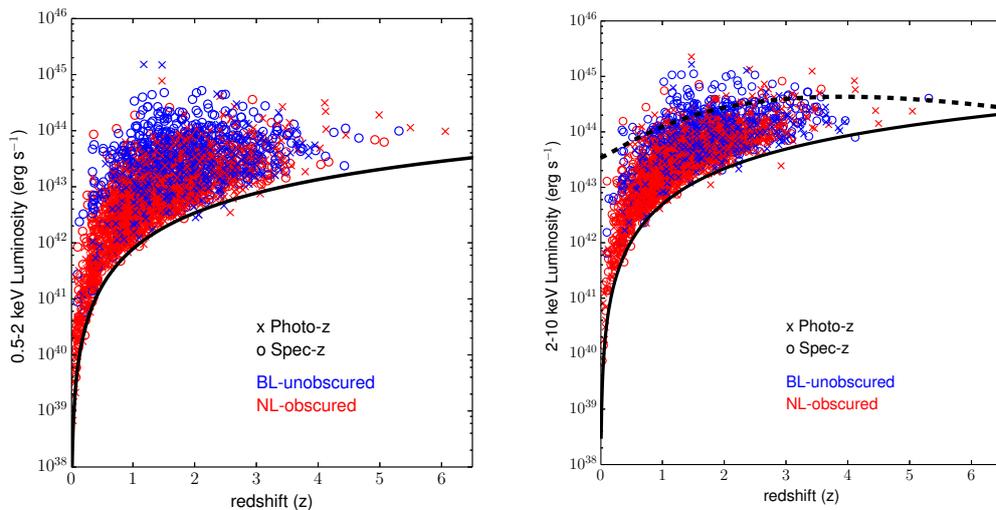


Figure 3.9: Rest-frame luminosity versus redshift in soft (0.5-2 keV, left) and hard (2-10 keV, right). Spectroscopic type (open circles) is plotted when available, otherwise photometric information (cross) is shown. Blue sources are Type 1 AGN; red are Type 2 AGN. We also plotted the survey flux limit (black solid line) and the L^* curve as function of redshift from Aird et al. (2015, black dashed line).

is an excellent bridge between deep pencil-beam surveys like CDF-S and large area surveys like Stripe 82. On the one hand, the CDF-S survey samples a significant fraction of sources with low X-ray luminosity, therefore more likely to be either star-forming galaxies or very obscured AGN. On the other hand, Stripe 82 main goal, is to find very bright and rare AGN, for which the coverage of large areas is required.

Another way to see the complementarity between different surveys is to look at the luminosity distribution in 0.5-2 keV, which is shown in Figure 3.11. *Chandra COSMOS-Legacy* luminosity distribution in the soft band peaks at $L_X \simeq 10^{43}$ erg s^{-1} , while more than 50% of the sources in CDF-S 4 Ms have $L_X < 10^{42}$ erg s^{-1} . Finally Stripe 82 0.5-2 keV luminosity distribution peaks at $L_X \simeq 2 \times 10^{44}$ erg s^{-1} .

In Figure 3.12 we show the hard band luminosity distribution: in this energy range, the *Chandra COSMOS-Legacy* distribution peaks is at $L_X \simeq 9 \times 10^{43}$ erg s^{-1} . The complementarity between surveys (Figure 3.12, right panel) is still present, although less stronger than in the 0.5-2 keV band. Here CDF-S 4 Ms peaks at $L_X \simeq 10^{43}$ erg s^{-1} , while Stripe 82 peaks at $L_X \simeq 5 \times 10^{44}$ erg s^{-1} .

In Figures 3.11 and 3.12 (left panels) we also plot the luminosity distribution of XMM-COSMOS in the 0.5-2 keV and 2-10 keV bands, respectively (B10, orange solid line): as can be seen, XMM-COSMOS already sampled the high luminosity distribution in the COSMOS field, while *Chandra COSMOS-Legacy* statistics is significantly better moving towards lower luminosities (i.e., $L_X \leq 5 \times 10^{43}$ erg s^{-1} in soft and $L_X \leq 10^{44}$ erg s^{-1} in hard band, respectively).

Chandra COSMOS-Legacy covers with an excellent statistics the range of redshift $1 \leq z \leq 3$, i.e. at the peak of the AGN activity and the following period, where

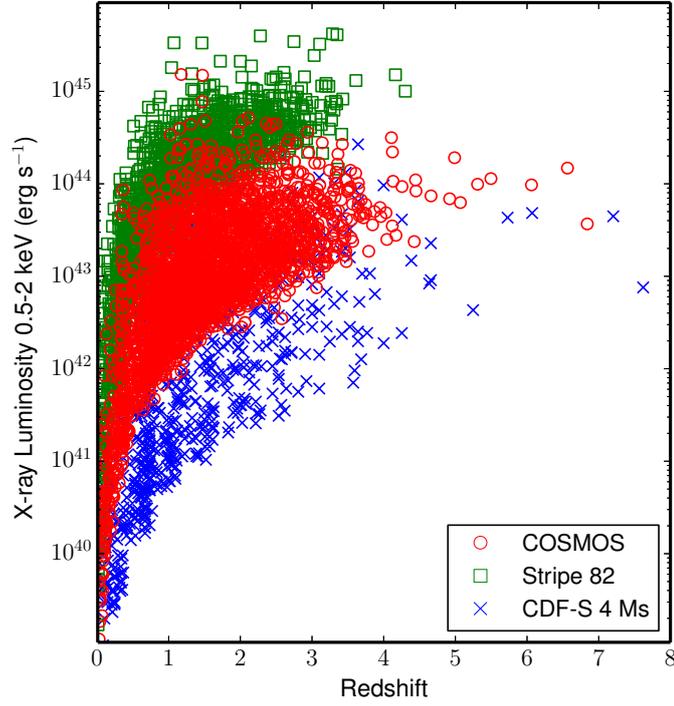


Figure 3.10: Rest-frame luminosity versus redshift in soft 0.5-2 keV band for *Chandra COSMOS-Legacy* (red circles), Stripe 82 (green squares) and CDF-S 4 Ms (blue crosses).

the sources span about two orders of magnitude in luminosity ($10^{42.5}$ - $10^{44.5}$ erg s $^{-1}$): in the redshift range $z=[1-2]$ there are 1572 sources, while in the range $z=[2-3]$ there are 713 sources.

	N_{new}	$\%_{new}$	N_{CCosm}	$\%_{CCosm}$	N_{all}	$\%_{all}$
Spectroscopic redshifts						
Broad line	239	36	372	36	611	36
Not broad line	387	59	608	59	995	59
Star	28	4	58	6	86	5
Photometric redshifts						
Unobscured AGN template	445	21	449	27	894	23
Obscured AGN template	261	12	104	6	365	9
Galaxy template	1398	65	1077	64	2475	64
Star template	61	3	60	4	121	3
Visually selected star	8		0		8	

Table 3.6: Number of X-ray sources divided by spectral or photometric type. N_{new} is the number of sources from the new survey, N_{CCosm} is the number of sources from C-COSMOS and N_{all} is the sum of the previous two values. The fraction is measured on the total number of sources with spectroscopic or SED template best fitting information.

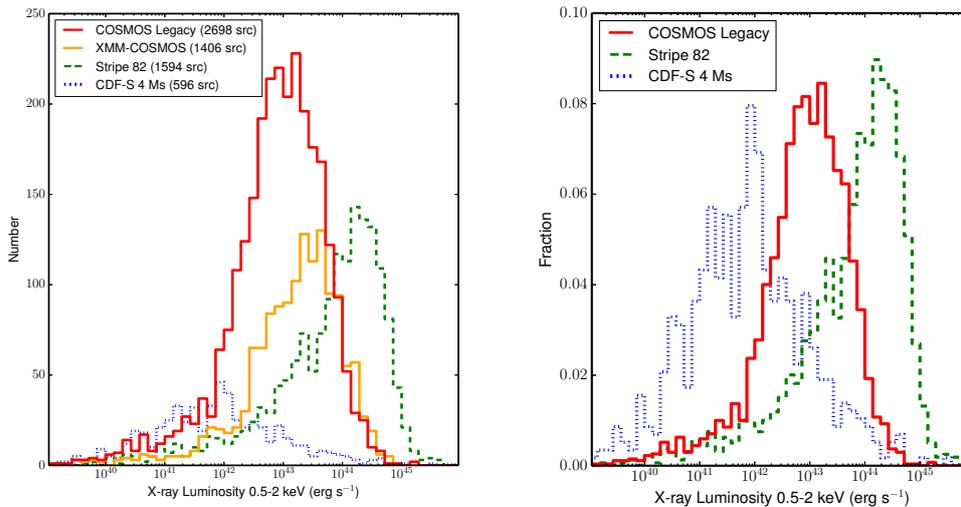


Figure 3.11: 0.5-2 keV rest-frame luminosity distribution (left panel), for all sources in *Chandra COSMOS-Legacy* with $z > 0$ (spectroscopic or photometric) and $\text{DET_ML} > 10.8$ in the given band (red solid line), XMM-COSMOS (orange solid line), CDF-S (blue dotted line) and Stripe 82 (green dashed line). We also plot (right panel) 0.5-2 keV rest-frame luminosity normalized distribution, to show the complementarity between surveys with different area-depth.

3.6 X-ray, optical and infrared properties of *Chandra COSMOS-Legacy* sources

3.6.1 Redshift Evolution of hardness ratio

Through unbinned statistics and careful background modelization, the minimum number of counts required for the X-ray spectral analysis is set only by the maximum relative error that one wants to allow. However, assuming a threshold of 70 net counts (Lanzuisi et al. 2013a), there are only $\simeq 950$ of the 4016 sources in our survey (i.e. $\simeq 24\%$) that fulfill this requirement. Nevertheless, it is possible to use the Bayesian estimate of Hardness Ratios (BEHR) method (Park et al. 2006) to derive a rough estimate of the X-ray spectral shape and therefore of the source nuclear obscuration. The hardness ratio (HR) of the source is defined as the ratio $\frac{H-S}{H+S}$, where H and S are the net counts of the source in the hard (2-7 keV) and in the soft (0.5-2 keV) band, respectively: an extended description of the procedure adopted to compute HR is reported in section 2.3.1. BEHR is particularly effective in the low count regime, because it does not need a detection in both bands to work and it runs Markov chain Monte Carlo calculation to compute errors.

To separate unobscured and obscured sources, we adopted a redshift dependent HR threshold (HR_{th}), computed assuming a typical obscured AGN spectrum, with a power-law with $\Gamma = 1.4$: consequently, we consider sources with $\text{HR} > \text{HR}_{th}$ as obscured. For sources with no redshift information, we used $\text{HR}_{th} = -0.2$, i.e. the mean HR value of our redshift-dependent curve. 1993 sources in *Chandra COSMOS-*

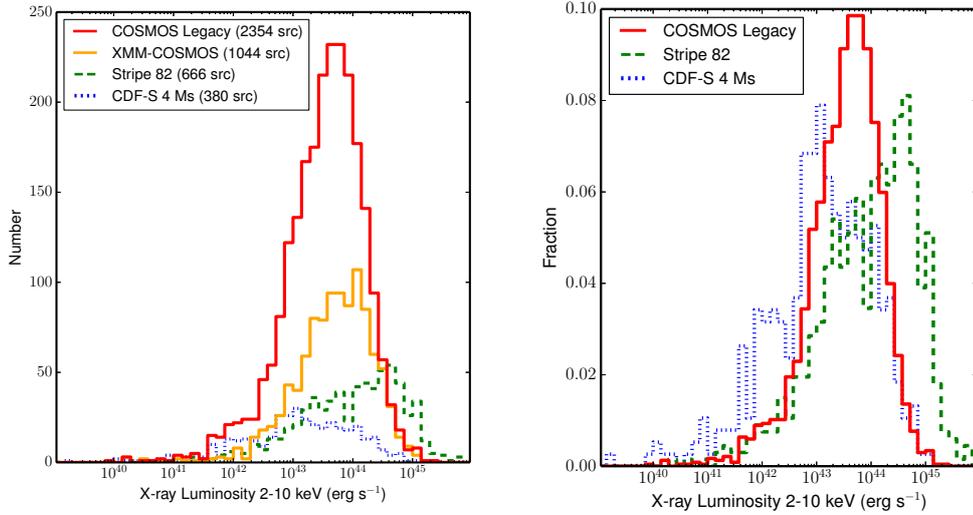


Figure 3.12: 2-10 keV rest-frame luminosity distribution (left panel), for all sources in *Chandra COSMOS-Legacy* with $z > 0$ (spectroscopic or photometric) and $\text{DET_ML} > 10.8$ in the given band (red solid line), XMM-COSMOS (orange solid line), CDF-S (blue dotted line) and Stripe 82 (green dashed line). We also plot (right panel) 2-10 keV rest-frame luminosity normalized distribution.

Legacy ($\simeq 49.6\%$ of the whole sample) have $\text{HR} > \text{HR}_{th}$, including both nominal values and 90% significance lower limits. We point out that such a value should be treated as a lower limit on the obscuration of the AGN population in COSMOS, particularly for those sources at high redshift and low-luminosity. There are in fact two main caveats involved in the use of the HR threshold: (i) the soft appearance of a fraction of Compton Thick sources at high redshift (Brightman et al. 2014), where we observe the intrinsic hard band emission in the soft band; (ii) a fraction of more obscured sources (at a given intrinsic flux) have flux below the flux limit of the survey and is therefore missed (Wilkes et al. 2013).

In Figure 3.13, we show the HR distribution for optically classified Type 1 (blue) and Type 2 (red) sources: spectral types are used when available, and the best-fit SED template model for the remaining sources. The mean (median) HR is $\text{HR} = -0.26 \pm 0.32$ (-0.3) for Type 1 sources and $\text{HR} = -0.03 \pm 0.46$ (-0.10) for Type 2 sources, taking in account in the computation also the 371 lower limits and the 616 upper limits (shown in Figure 3.13 as dashed lines). The hypothesis that the two distributions are actually the same is rejected on the basis of a Kolmogorov-Smirnov (KS) test, with a probability $> 99.998\%$. A similar result was already shown in B10 in XMM-COSMOS: we found that the values do not change significantly if we use only a subsample with flux $f_{0.5-10} < 5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, i.e. in the range where *Chandra COSMOS-Legacy* statistics is significantly larger than the XMM-COSMOS one, and are therefore not dominated by the brightest sources.

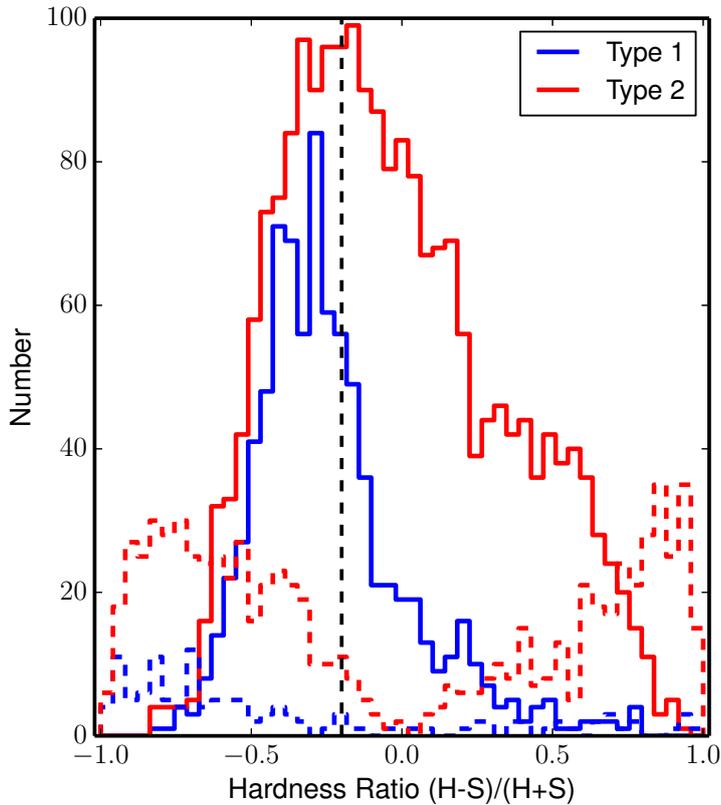


Figure 3.13: Average HR distribution for optically classified Type 1 (blue) and Type 2 (red) sources. Upper (towards the left) and lower (towards the right) limits are plotted as dashed lines. The black dashed line at $HR = -0.2$ marks the average HR_{th} , computed assuming a typical obscured AGN spectrum, with a power-law with $\Gamma = 1.4$.

Finally, we studied the behavior with redshift of the HR: we show the result in Figure 3.14, where once again we divide our sample in Type 1 (blue) and Type 2 (red) sources, on the basis of the optical classification. We also show three curves of different column density ($N_H = 10^{21}, 10^{22}$ and 10^{23} cm^{-2} , dotted, dashed and solid line, respectively), obtained assuming a power-law spectrum with $\Gamma = 1.4$ (black) and $\Gamma = 1.8$ (green). As can be seen, the average HR of Type 2 lies above the $N_H = 10^{22} \text{ cm}^{-2}$ curve at all redshifts, regardless of the assumed Γ , while the average HR of Type 1 sources is generally below the $N_H = 10^{21} \text{ cm}^{-2}$ curve computed assuming $\Gamma = 1.4$. However, the large dispersion in the HR distribution, at any redshift, does not allow to claim that the optically classified Type 1 and Type 2 sources lie in two different regions of the HR versus redshift diagram. Such a dispersion (in Figure 3.14 we show the 68% dispersion) is particularly large for Type 2 sources ($\sigma > 0.3$ at $z < 3$), where it is at least partially due to the fact that a significant fraction of sources with a galaxy best-fit SED template are actually objects where the galaxy optical contribution is dominant, and it is therefore not possible to

correctly classify the AGN; in the X-ray, instead, the AGN contribution is almost unbiased even at the *Chandra COSMOS-Legacy* flux limit. We discuss further the different information obtained using the HR as an obscuration indicator, instead of the optical classification, in Section 3.6.4.

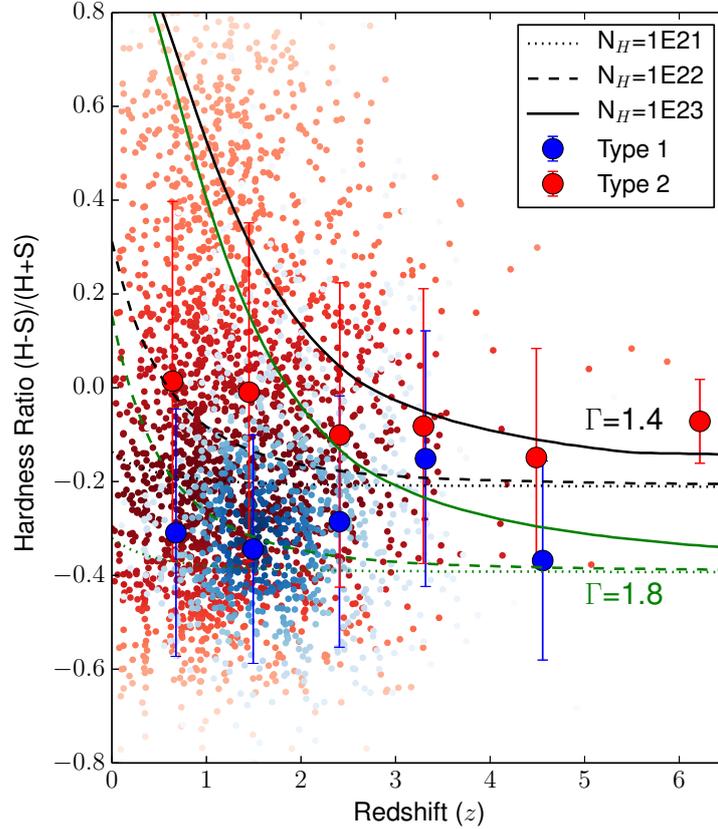


Figure 3.14: HR evolution with redshift for optically classified Type 1 (blue) and Type 2 (red) sources. The error bars represent the 68% dispersion. Three curves of different N_H (10^{21} cm^{-2} , dotted line, 10^{22} cm^{-2} , dashed line, 10^{23} cm^{-2} , solid line) are also plotted for comparison, obtained assuming a power-law spectrum with $\Gamma=1.4$ (black) and $\Gamma=1.8$ (green). Single values for each source with significant HR are plotted in the background (darker scale color indicates higher source density).

Intrinsic N_H and de-absorbed luminosity estimate

To estimate the intrinsic N_H of the sources in our sample, we used the best available redshift and the HR of each source, using a sample of redshift vs HR curves like those shown in Figure 3.14. These curves have been obtained assuming an X-ray spectral power-law with slope $\Gamma=1.8$. We did not estimate a N_H value for sources without a reliable redshift. After estimating N_H , we compute the intrinsic absorp-

tion correction $k_{abs}=f_{abs}/f_{int}$, where f_{int} and f_{abs} are the intrinsic and absorbed fluxes in a given band, respectively. Finally, we repeated the whole procedure using the HR lower and upper limits, therefore estimating upper and lower limits on the N_H .

We compared our N_H estimate with those from Lanzuisi et al. (2013a), obtained through X-ray spectral fitting on the subsample of 388 sources with more than 70 net counts in C-COSMOS, and we found a general good agreement. The sample can be divided as follows:

1. About 56% of the sources have only an upper limit on N_H in both our sample and the Lanzuisi et al. (2013a) one, and for $\simeq 95\%$ of these sources this upper limit is $<10^{22}$ cm $^{-2}$ in both samples.
2. $\simeq 18\%$ of the sources have a significant N_H value in both samples: for these sources the agreement between N_H estimates is generally good, with a mean (median) ratio $r=0.95$ (0.88) between the Lanzuisi et al. (2013a) N_H estimate and ours. We did not find a significant change in the ratio distribution at different fluxes.
3. $\simeq 26\%$ of the sources have a significant detection in one sample and only an upper limit in the other, and more than 90% of the sources in this last subsample have actually a significant detection in Lanzuisi et al. (2013a) and only an upper limit in our sample. This discrepancy can be explained with the better accuracy that the spectral analysis provides with respect to the HR-based estimate: it is also worth noticing that the majority of our upper limits are located within the 1σ uncertainty provided by Lanzuisi et al. (2013a).

A spectral analysis of the $\simeq 950$ sources with more than 70 net counts in the whole *Chandra COSMOS-Legacy* survey (included the 388 sources already analyzed) has already been planned (Lanzuisi et al. in prep.); moreover, the excellent *Chandra COSMOS-Legacy* statistics will allow us to perform stacked spectral analysis of sources with similar properties (e.g. optically classified Type 1 and Type 2 AGN), and therefore compute average N_H .

3.6.2 X-ray to Optical Flux Ratio

Since the beginning of X-ray surveys, a typical way to characterize different types of X-ray sources has been the X-ray to optical flux ratio (hereafter X/O), which is a simple first estimator of the source classification (Tananbaum et al. 1979; Maccacaro et al. 1988),

$$X/O = \log(f_X/f_{opt}) = \log(f_X) + C + m_{opt}/2.5, \quad (3.4)$$

where f_X is the X-ray flux in a given band, m_{opt} is the magnitude in the chosen optical band and C is a constant related to the filter used in the optical observations. The magnitude used in this equation is usually the i or r -band one (see Brandt & Hasinger 2005). The relation was first used in the soft X-ray band: in this band, the largest part of bright spectroscopically identified AGN, both BLAGN and non-BLAGN, lie in the region $X/O=0\pm 1$ (e.g. Schmidt et al. 1998; Stocke et al. 1991;

Lehmann et al. 2001), hereafter defined as the “soft locus”. *Chandra* and XMM-*Newton* studies extended this relation to harder bands (Hornschemeier et al. 2001, 2-8 keV; Alexander et al. 2001, 2-8 keV; Fiore et al. 2003, 2-10 keV; Della Ceca et al. 2004, 4.5-7.5 keV; Cocchia et al. 2007, 2-10 keV). The trend (i.e. the existence of a “hard locus”, a general correlation between X-ray and optical fluxes) was confirmed at bright fluxes also in these bands, but with a non negligible scatter around the median values, both in soft and hard band, at lower fluxes (Brandt & Hasinger 2005).

This scatter is linked to different types of objects: obscured AGN ($N_H > 10^{22}$ cm $^{-2}$) generally lie in the region with $X/O > 1$ (Fiore et al. 2003; Perola et al. 2004; Civano et al. 2005; B10); normal, low X-ray flux galaxies have $X/O < -2$ (Xue et al. 2011). Finally, a third class of objects is defined, formed by unobscured X-ray Bright, Optically Normal Galaxies (XBONGs, see Elvis et al. 1981, Comastri et al. 2002; Civano et al. 2007; Trump et al. 2009b). These peculiar sources were named extreme or “unconventional” (Comastri et al. 2003; Mignoli et al. 2004) or “elusive” (Maiolino et al. 2003), especially when X/O is defined in the hard X-ray band.

We studied the X-ray flux versus optical magnitude relation using the whole *Chandra COSMOS-Legacy* dataset, in order to put better constraints on it, especially at the X-ray faint end, where our sample is twice as large as the C-COSMOS one. In Figure 3.15 we show the relation between the *i*-band magnitude and the X-ray flux in both soft (left) and hard (right) bands for the whole *Chandra COSMOS-Legacy* survey: our sample comprises only sources with *i*-band magnitude and DET_ML > 10.8 in the given X-ray band, and contains 2798 sources in the soft band and 2363 sources in the hard band. The “soft locus” and the “hard locus” are also plotted, using a constant $C(i) = 5.91$ in the soft band and $C(i) = 5.44$ in the hard band. The constant has been computed on the basis of the *i*-band filters width, for all the filters in COSMOS (Subaru, CFHT and SDSS).

We studied the *i*-band-X-ray flux relation of the whole *Chandra COSMOS-Legacy* by dividing our sample in three different subsamples: (*i*) candidate AGN population (red circles), i.e., sources with $L_X > 10^{42}$ erg s $^{-1}$ in full band (2496 in the soft band and 2235 in the hard band); (*ii*) low-luminosity sources (blue squares, $\simeq 5\%$ and of $\simeq 3\%$ the soft and hard samples, respectively: 138 sources in the soft band and 81 in the hard band), i.e., objects with $L_X < 10^{42}$ erg s $^{-1}$; (*iii*) stars (cyan stars, 119 in the soft band and 33 in the hard band).

A significant fraction of sources lie outside both the soft locus ($\simeq 17\%$) and the hard locus ($\simeq 18\%$). We then computed the 90% width of the X/O distribution, i.e., tracing the 5% lower percentile and the 95% upper percentile of the *i*-band distribution of the AGN population. To do so, we divided the sources in X-ray flux bins of width 0.25 dex: the results are shown as black solid lines in Figure 3.15. We call this the *Chandra COSMOS-Legacy* locus.

The *Chandra COSMOS-Legacy* locus is shifted to fainter optical magnitudes relative to both the soft and hard locus by $\Delta(X/O) \simeq 0.3-0.5$ in both bands, and does not change significantly over 1.5 dex in flux. The *Chandra COSMOS-Legacy* locus is consistent with that of C12 at any flux and is consistent with the X/O being defined with soft X-ray selected sources, which are usually bright both in the optical band and in the X-rays.

The majority of stars and candidate low luminosity AGN or non active galaxies (i.e. sources with $L_X < 10^{42}$ erg s $^{-1}$) lie in the region of Figure 3.15 at low X-ray fluxes and bright optical magnitudes. However, there is a fraction of sources with low L_X which show X-ray to optical properties consistent with those of sources with $L_X > 10^{42}$ erg s $^{-1}$: 23 of 138 sources with $L_X < 10^{42}$ erg s $^{-1}$ (17%) lie inside the soft *Chandra COSMOS-Legacy* locus, while 25 of 81 (31%) lie inside the hard *Chandra COSMOS-Legacy* locus. The fraction is considerably higher in the hard band, where it is more likely to observe obscured AGN at low-medium redshift. A more accurate analysis of this subsample of candidate obscured AGN is beyond the purpose of this work and requires an extended analysis of several other parameters, like the HR and the morphology of the sources (see Ranalli et al. 2012; Xue et al. 2011).

We also studied the trend with X-ray soft flux of the *K* and 3.6 μ m magnitudes: the two samples contain 2824 and 2868 sources with $L_X > 10^{42}$ erg s $^{-1}$, respectively. Here the soft locus has been computed with Equation 3.4, using constants $C=6.86$ and $C=7.34$ for the *K* and 3.6 μ m bands, respectively. We computed again also the region which contains 90% of the AGN population and we found that this region is considerably smaller (1.5-2 mag) than in the *i*-band (we show the *K*- and 3.6 μ m-band relations in Figure 3.16). This narrower relation suggests that the relation of *K* and 3.6 μ m magnitudes with the X-ray flux is stronger than that of the *i*-band one, an evidence which is also reflected in the higher identification rates for *K* and 3.6 μ m counterparts. Such a result could be mainly linked to the lower contribution of the nuclear extinction at near-infrared wavelengths (Mainieri et al. 2002; Brusa et al. 2005).

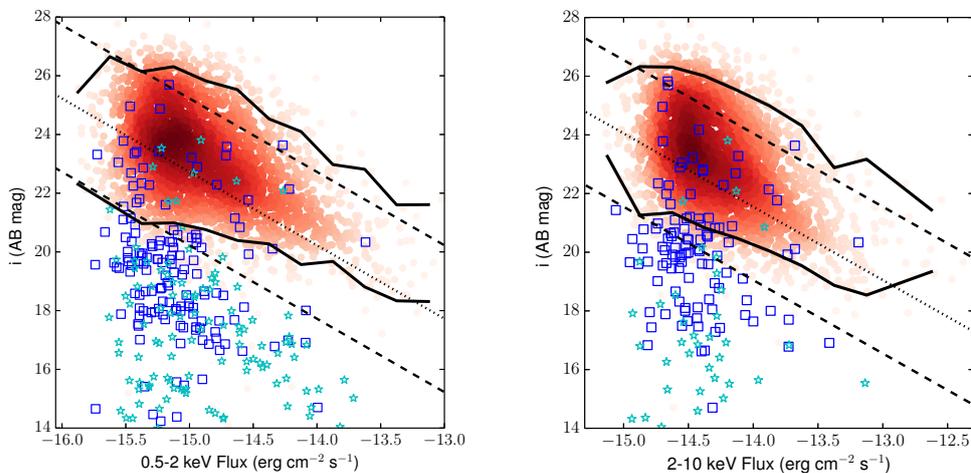


Figure 3.15: X-ray flux (soft on the left, hard on the right) versus *i*-band total (aperture corrected) magnitude, for all X-ray sources with an *i*-band counterpart. The black dashed lines define the so-called “soft locus” and “hard locus” of AGN along the correlation $X/O=0\pm 1$. Red circles are AGN ($L_X > 10^{42}$ erg s $^{-1}$, darker scale color indicates higher source density), blue squares are sources with $L_X < 10^{42}$ erg s $^{-1}$ and cyan stars are stars. Black solid lines represent the region including 90% of the *Chandra COSMOS-Legacy* AGN population.

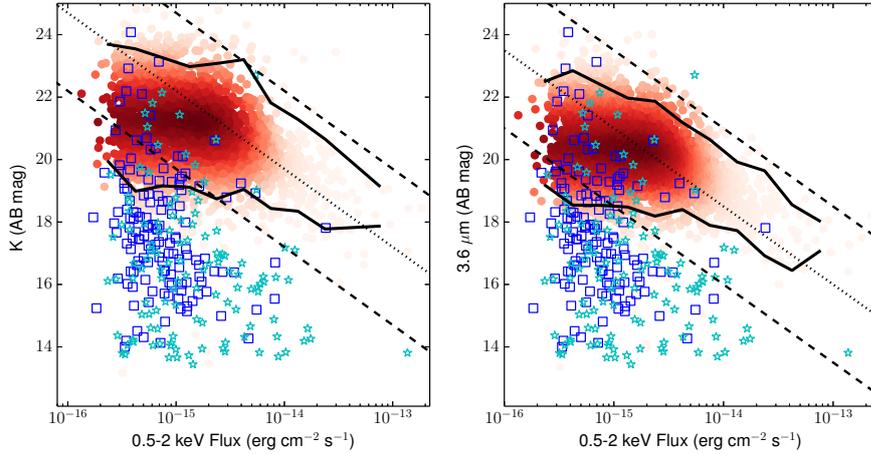


Figure 3.16: Soft X-ray flux versus K - (left) and $3.6 \mu\text{m}$ -band total (aperture corrected) magnitude, for all X-ray sources with a K -band counterpart. The black dashed lines define the so-called “soft locus” of AGN along the correlation $X/O=0\pm 1$. Red circles are AGN ($L_X > 10^{42} \text{ erg s}^{-1}$, darker scale color indicates higher source density), blue squares are sources with $L_X < 10^{42} \text{ erg s}^{-1}$ and cyan stars are stars. Black solid lines represent the region including 90% of the *Chandra COSMOS-Legacy* AGN population.

3.6.3 X/O–hard band luminosity relation

In Figure 3.17 we show X/O versus the hard band X-ray luminosity for the 2249 sources with a significant detection in the hard band, with optical counterpart and with $L_X > 10^{42} \text{ erg s}^{-1}$ in the 2–10 keV hard band. Fiore et al. (2003) showed the existence of a linear correlation between X/O and the hard X-ray luminosities for Type 2 AGN. Such a correlation is due to the fact that extinction strongly reduces the nuclear UV/optical emission (where the only remaining contribution is from the host galaxy), but it is instead not heavily attenuated in the 2–10 keV band, at least for sources with $N_H < 10^{24} \text{ cm}^{-2}$.

In the *Chandra COSMOS-Legacy* sample, Type 2 AGN (red) show a clear linear trend over more than three orders of magnitude, with slope 0.96 ± 0.02 (black solid line) and correlation coefficient $\rho = 0.79$, with p-value=0. This subsample consists of 1563 sources, out of which 611 are spectroscopic Type 2 AGN, 65 are sources with photometric redshifts and SED fitted with an obscured AGN template, and the remaining 867 sources have photo-z and SED best fitted with a galaxy template. On the other hand, unobscured AGN (blue, 695 sources, out of which 504 with spectroscopic information and the remaining 191 with only photometric information) do not show a clear trend between hard X-ray luminosity and X/O: Type 1 AGN are on average 0.5 dex more luminous than non-Type 1 AGN (95% of the Type 1 sources have $L_X > 10^{43} \text{ erg s}^{-1}$), but there are many sources with X/O < 0 even at high X-ray luminosity. This is an expected result, because BLAGN have by definition low obscuration, so the optical flux is higher than in Type 2 AGN, at any X-ray flux.

We then tested this relation only for the 809 sources with $F_X > 8 \times 10^{-15}$ erg $\text{s}^{-1} \text{cm}^{-2}$ in the hard band, i.e., the flux limit of HELLAS2XMM (Fiore et al. 2003), where the trend between X/O and L_X (2-10 keV) was first reported. For this subsample, a linear relation (black dashed line) still exists, with slope 0.92 ± 0.03 and correlation coefficient $\rho = 0.83$, with p-value=0.

At the faint end of the optical counterparts ($i > 25$, 318 sources, green) we instead found a considerably weaker trend (black dotted line), with slope 0.45 ± 0.04 and correlation coefficient $\rho = 0.54$, with p-value=0, confirming that the relation between X/O and L_X becomes flatter, if not totally disappears, moving to faint magnitudes (Barger et al. 2005; Civano et al. 2005). This trend could be partially explained with a selection effect, but when we selected other optical magnitude ranges we found that the relation still exists, even if less steep (for example, for $i = [21-23]$ the relation has slope 0.69 ± 0.04 and $\rho = 0.82$, while for $i = [22-24]$ the relation has slope 0.65 ± 0.04 and $\rho = 0.78$, with p-value=0 in both cases).

We tested several parameters to better understand the meaning of this difference.

1. There is no difference in the HR of the two samples: the mean HR value is the same for both the whole sample of candidate Type 2 (HR= 0.10 ± 0.45) and in the optically faint subsample (HR= 0.12 ± 0.44), and the hypothesis that the two distributions are actually same can not be rejected on the basis of a KS-test (p-value=0.57).
2. The mean redshift of the whole sample, $z = 1.36 \pm 0.72$, is lower than the one of the optically faint subsample (although in agreement within the errors), $z = 2.12 \pm 0.61$. The hypothesis that these two redshift distributions can be obtained by the same parent population is rejected on the basis of a KS-test (P>99.999%).
3. Suh et al. (to be submitted) performed a multi-component modeling from far-infrared ($500 \mu\text{m}$) to near-ultraviolet (2300\AA) on Type 2 sources in *Chandra COSMOS-Legacy*. They used a 3-component SED fitting with nuclear hot dust torus, galaxy, and starburst components in order to decompose the SED into a nuclear AGN and host galaxy stellar contributions. They derived an estimate of the host galaxy stellar masses using the best-fit galaxy template, then calculating the total IR luminosities, which are integrated between 8-1000 μm from the best-fit starburst template. They then combined the infrared observations with UV observations to derive the total star formation rate (SFR), $\text{SFR}_{tot} = \text{SFR}_{IR} + \text{SFR}_{UV}$, thus estimating reliable SFRs for both obscured and unobscured sources (Arnouts et al. 2013).

The specific star formation rate (sSFR= SFR/M^*) distribution spans over five orders of magnitude (sSFR= $[10^{-13}-10^{-8}] \text{yr}^{-1}$) for the whole sample of candidate Type 2 AGN, while is slightly narrower for the subsample with $i > 25$ (sSFR= $[10^{-11}-10^{-8}] \text{yr}^{-1}$). Moreover, the mean sSFR is almost two times larger in the optically faint subsample ($3.4 \times 10^{-10} \text{yr}^{-1}$) than in the whole sample of candidate Type 2 AGN ($1.9 \times 10^{-10} \text{yr}^{-1}$). Once again, the hypothesis that the two distributions have been originated by the same sSFR distribution is rejected on the basis of a KS-test (P>99.999%).

In conclusion, our data suggest that the existence of a linear trend between the hard band X-ray luminosity and X/O for Type2 AGN becomes weaker at fainter optical magnitudes, where sources have higher redshifts and the sSFR is higher. In this subsample the AGN contribution to the optical emission is less significant, while the host galaxy contamination is higher, than in the X-ray.

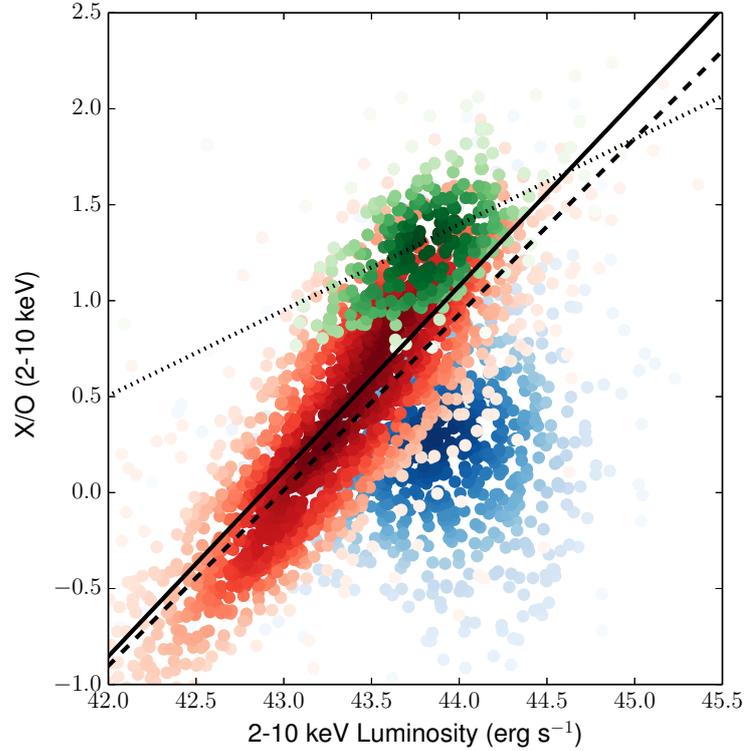


Figure 3.17: X/O versus hard band luminosity, rest frame, for *Chandra COSMOS-Legacy* sources with $L_X > 10^{42}$ erg s $^{-1}$ in the 2-10 keV band. Blue sources are Type 1 AGN; red are Type 2 AGN. Sources with $i > 25$ are plotted in green. Darker scale colours indicate higher source density. The best fit relation for all non BLAGN or obscured AGN and galaxy dominated objects with $L_X > 10^{42}$ erg s $^{-1}$ (black solid line), for those with $f_X(2-10 \text{ keV}) > 8 \times 10^{-15}$ erg s $^{-1}$ cm $^{-2}$ (black dashed line) and for those with $i > 25$ (black dotted line) are also plotted.

3.6.4 Luminosity dependence of the AGN obscured fraction

The existence of a trend between the fraction of obscured AGN and the X-ray luminosity was already shown by Lawrence & Elvis (1982). More recently, Ueda et al. (2003) confirmed the result in the 2-10 keV (rest frame) band: at low luminosities, $L_X \simeq 10^{42}$ erg s $^{-1}$, almost the whole sample is composed of obscured AGN, while unobscured sources prevail moving towards high luminosities, i.e. $L_X > 10^{44}$ erg s $^{-1}$. This trend has been confirmed over the years by other works, e.g. La Franca et al. (2005), using HELLAS2XMM; Hasinger (2008), who divided the

sample in unobscured and obscured sources on the basis of both the optical spectroscopic classification and X-ray absorption properties; Ueda et al. (2014), who also found that at higher redshifts the decline of the obscured AGN fraction starts at higher luminosities; and lastly Buchner et al. (2015). The same trend has already been confirmed by *XMM-COSMOS*, on the basis of the optical classification of the sources, for both the whole survey (B10) and in different redshift bins (Merloni et al. 2014). A different result was instead found by Lusso et al. (2013), which found no clear trend with 2-10 keV luminosity of the obscured fraction of Type 1 AGN in *XMM-COSMOS*, using SED analysis to estimate the dust covering fraction.

The whole *Chandra COSMOS-Legacy* survey has about 900 more sources with $z > 0$ and $\text{DET_ML} > 10.8$ in the hard band than *XMM-COSMOS* (2354 versus $\simeq 1450$), and twice better statistics than *XMM-COSMOS* at luminosities lower than $L_X \simeq 10^{44}$ erg s^{-1} (see Figure 3.12, left panel). We studied the relation between the obscured fraction of sources versus the 2-10 keV de-absorbed luminosity: we estimated the absorption contribution following the procedure described in Section 3.6.1. In Figure 3.18 (left panel) we plot (blue squares) the fraction of spectroscopically selected obscured AGN (i.e. the ratio between those sources which have been classified as non-BL AGN and all sources with spectroscopic type information). The whole spectroscopic type sample contains 1162 sources.

More than 90% of the sources at $L_X \leq 10^{42}$ erg s^{-1} are obscured, while the fraction of obscured sources decreases to $\simeq 80\%$ at $L_X \simeq 10^{43}$ erg s^{-1} and drops around 20% at $L_X \geq 10^{44}$ erg s^{-1} . However, there are significant uncertainties on the trend estimated using only the spectroscopic information, first of all because our spectroscopic sample is not complete (only 42% of the sources have a spectral type) and, moreover, the selection of sources for a spectral analysis in the COSMOS field has so far been biased towards the optically brightest sources (see Figure 3.6), which are more likely to be unobscured broad line AGN, which could result in an under-estimate of the obscured fraction at high luminosities ($L_X \geq 10^{44}$ erg s^{-1}).

We therefore estimated the fraction of obscured AGN using the photometric classification for all the sources without a spectral type: the total number of sources with either a spectroscopic or a photometric type is 2343. In Figure 3.18, left panel, we plot the fraction of obscured sources from the combined photometric and spectroscopic information in red circles: the agreement with the spectroscopic trend is good at low luminosities (more than 90% of sources with $L_X < 10^{42}$ erg s^{-1} are obscured). At high luminosities ($L_X \geq 10^{43.4}$ erg s^{-1}) the fraction of obscured sources is a factor $\simeq 2$ larger, i.e. $\simeq 40\%$. This trend does not change significantly while computed in complete bins of redshift and luminosity.

We also compared the optical obscuration results with those obtained using the X-ray properties of the sample, using the HR (see Section 3.6.1): we divided the sources between obscured and unobscured using the HR threshold $\text{HR}_{th} = -0.2$. This is the same threshold used in B10, and it is also the average HR value at any redshift, assuming an obscured AGN spectral slope with $\Gamma = 1.4$ (see Figure 3.14, black dotted line). The total number of sources in this third sample is 2354 (including the HR upper and lower limits), and the HR-determined obscuration fraction is plotted with cyan triangles in the right panel of Figure 3.18. The fraction of obscured sources at low luminosities is lower than in the two previous cases ($\simeq 65\%$ against $\simeq 90\%$, even at $L_X < 10^{42}$ erg s^{-1}), and is comparable with the optically based result at $L_X \geq 10^{44}$

erg s⁻¹. The discrepancy between the optical and X-ray obscured fraction at low X-ray luminosity could be due to the fact that in this luminosity range the main optical luminosity contributor is the host galaxy, the AGN being therefore hidden; conversely, in the X-rays the galaxy contribution is almost negligible at the *Chandra COSMOS-Legacy* fluxes and the AGN identification is unbiased (see also Merloni et al. 2014). As for the optical obscuration, the trend does not change significantly adopting complete samples in bins of redshift and luminosity.

Our results at $L_X > 10^{43.5}$ erg s⁻¹ are also in good agreement with the fraction of obscured sources estimated using the N_H value from Lanzuisi et al. (2013a), where the obscured fraction of AGN is between 40 and 50% in the 2-10 keV luminosity range $L_X = [10^{43.5} - 10^{45}]$ erg s⁻¹.

In Figure 3.18 we also compare our results with Merloni et al. (2014) *XMM-COSMOS* results in different bins of redshift ($z = [0.3-0.8]$, magenta diamonds; $z = [0.8-1.1]$, yellow diamonds, and $z = [2.1-3.5]$, green diamonds). There is a general agreement between these and our data, within the errors, using both the optical and the X-ray classification. Small differences are observed when comparing our results with theirs at $z = [2.1-3.5]$, where at $L_X > 10^{44}$ erg s⁻¹ their results show 10-15% more obscured sources on the basis of the optical information.

We also compare our results with the predictions of the population synthesis models of Gilli et al. (2007, black solid line) and Miyaji et al. (2015, black dotted line), both based only on the X-ray classification, with the one of and Treister & Urry (2006, black dashed line), based on both optical and X-ray classifications. For all these models, we measured the fraction of sources with $N_H > 10^{22}$ cm⁻², and folded the contribution in the two different N_H ranges through the observed flux range of our survey. We divide our results in three ranges of luminosity.

1. At $L_X < 10^{43.5}$ erg s⁻¹ the two models predictions diverge: the Treister & Urry (2006) trend is more similar to the one obtained using the optical spectroscopic and photometric classifications, while our HR-based obscured fraction is closer to the predictions of the Gilli et al. (2007) and Miyaji et al. (2015) models.
2. At $10^{43.5} < L_X < 10^{44}$ erg s⁻¹ there is a good agreement between the three models and our results obtained using both spectroscopic and photometric types or using the HR information.
3. At $L_X \geq 10^{44}$ erg s⁻¹ the Treister & Urry model overpredicts the fraction of obscured sources by 10-20% with respect to our results using the optical classification, while the Gilli model is in good agreement with both the X-ray and optical obscuration fraction.

Over the whole luminosity range, the observed behavior on the basis of the optical classification is fairly consistent with the Treister et al. (2009) model predictions, while the HR-based evidence of weak correlation between 2-10 keV luminosity and obscuration fraction is consistent with the Gilli et al (2007) and Miyaji et al. (2015) models predictions.

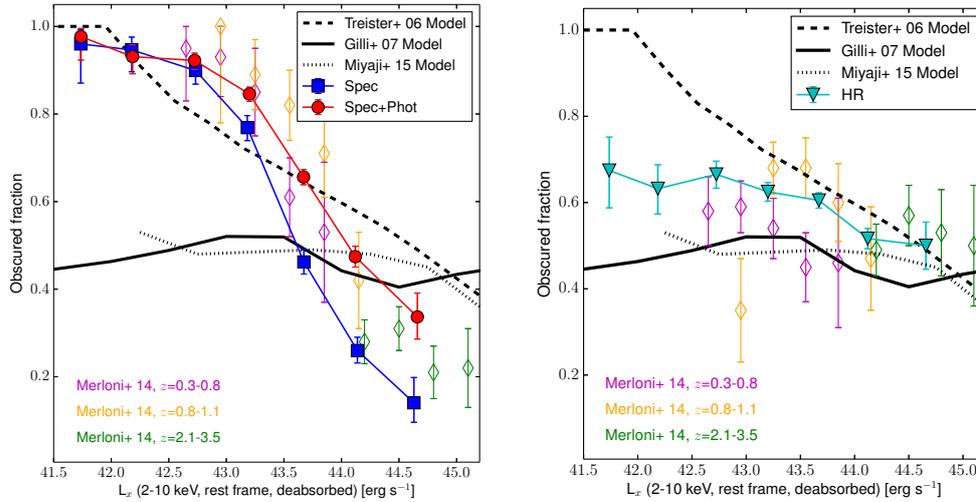


Figure 3.18: Fraction of obscured sources as a function of 2-10 keV rest frame de-absorbed luminosity, using only spectroscopic information (red circles, left), combined spectroscopic and photometric information (blue squares, left) and X-ray only HR based information, assuming as obscured all those sources with $HR > -0.2$ (cyan triangles, right). Results obtained by Merloni et al. (2014) using subsamples of XMM-COSMOS in different bins of redshift ($z=[0.3-0.8]$, magenta; $z=[0.8-1.1]$, yellow; $z=[2.1-3.5]$, green) are shown as diamonds. We also plot the fraction of AGN with $N_H > 10^{22} \text{ cm}^{-2}$ obtained using the XRB synthesis models by Gilli et al. (2007, solid black line), Miyaji et al. (2015) and Treister & Urry (2006, black dashed line). All errors are 1σ and have been calculated using Equation 26 of Gehrels (1986).

3.7 Conclusions

In this chapter we presented the identification procedure of optical/IR counterparts of the new 2273 *Chandra COSMOS-Legacy* sources. We then presented the X-ray to optical properties of the 4016 sources in the whole *Chandra COSMOS-Legacy* survey (i.e., the combination of the new survey and the 1743 C-COSMOS sources). The following are the main results of the identification process.

1. We associated the new 2273 *Chandra COSMOS-Legacy* point-like sources with optical/IR counterparts in three different bands (i , K and $3.6 \mu\text{m}$), using the likelihood ratio technique, based on both the separation between the X-ray and the optical/IR source, and the magnitude of the candidate counterpart. We found a secure counterpart in at least one of the three bands for 97% of the X-ray sources.
2. 31 of 2273 X-ray sources have no optical/IR counterpart: even if 30-50% of these sources could actually be spurious X-ray detections, or caused by bad optical/IR imaging, the remaining part of them are candidate obscured and/or high redshift sources.

Thanks to the large multiwavelength coverage in the COSMOS field and to the numerous spectroscopic campaigns, we were able to provide a redshift, either spectroscopic or photometric, for almost our whole sample (96%). We also provided a spectroscopic type and/or a photometric type from SED template best fitting.

1. 2058 sources of the 4016 in the whole *Chandra COSMOS-Legacy* survey (51.2% of the whole sample) have a reliable spectroscopic redshift. Of these sources, 36% are classified as BLAGN, while 59% do not show evidence of broad lines, but only narrow emission and absorption lines. Finally, 5% of the sources with spectroscopic information are spectroscopically identified stars.
2. We provide a photometric redshift and a related photometric classification for 3872 sources (96%). The majority (65%) of these sources are fitted with a non-active galaxy, even if only a minority of sources (26% in soft and 13% in hard band) have $L_X < 10^{42}$ erg s⁻¹. 9% of the sample is fitted with an obscured AGN template and 23% with an unobscured AGN template. Finally, 121 sources, 3% of the whole sample, have been identified as stars on the basis of the photometric template. In XMM-COSMOS (B10) there were $\simeq 50\%$ Type 1 sources and $\simeq 50\%$ Type 2 sources: the larger fraction of obscured sources in *Chandra COSMOS-Legacy* is due to its flux limit three times deeper than in XMM-COSMOS.
3. The *Chandra COSMOS-Legacy* luminosity distribution in the soft band peaks at $L_X \simeq 10^{43}$ erg s⁻¹ (Figure 3.11), and it is an excellent bridge between deep pencil beam surveys like CDF-S (Xue et al. 2011) and large area surveys like Stripe 82 (LaMassa et al. 2013a; La Massa et al. submitted). Moreover, *Chandra COSMOS-Legacy* covers with an excellent statistics (2285 sources in the soft band) the range of redshift $1 \leq z \leq 3$, i.e. at the peak of the AGN activity and the following period (Hasinger et al. 2005). Our survey also samples with solid statistics the luminosity range below the knee of the luminosity function, up to redshift $z \simeq 4$ (Figure 3.9, right panel).

Finally, we studied several X-ray-to-optical properties of our sample, especially focusing on the obscured sources.

1. We used the HR as a rough, purely X-ray based obscuration estimate. The mean (median) HR is $HR = -0.26 \pm 0.32$ (-0.30) for optically classified Type 1 sources and $HR = -0.03 \pm 0.46$ (-0.10) for optically classified Type 2 sources. We also studied the evolution with redshift of HR (Figure 3.14), and we found that, while the average HR of Type 2 sources lies above the one of Type 1 sources at any redshift, both samples show an intrinsically large dispersion. In the Type 2 sample, such a dispersion can be caused by a significant fraction of sources with a galaxy best-fit SED template being galaxy-dominated in the optical but not intrinsically obscured, therefore avoiding the possibility to correctly classify the AGN.
2. With our sample of 2798 sources in the soft band and 2363 sources in the hard band we put stronger constraints to the X-ray to optical flux ratio locus (Figure 3.15). Our results confirm, with a statistics 20% and 40% larger in

the soft and hard bands, respectively, the locus shown in C12: the new locus is shifted to faint optical magnitudes in both soft and hard X-ray band by $\Delta(X/O) \simeq 0.3-0.5$, without significantly changes at different fluxes. We also studied the trend with X-ray soft flux of the K (Figure 3.16) and $3.6 \mu\text{m}$ magnitudes and we found that the region which contains 90% of the AGN population is considerably smaller (1.5-2 mag) than the one in the i -band. This narrower relation indicates a stronger correlation of X-rays with near-infrared bands than with optical bands, a result that could be explained with a lower contribution of the nuclear extinction at near-infrared wavelengths. This last result is in general agreement with the fact that near-IR selection techniques are almost as effective as X-ray ones (Stern 2015).

3. The majority of candidate low luminosity AGN or non active galaxies (i.e. sources with $L_X < 10^{42} \text{ erg s}^{-1}$) have low X-ray fluxes and bright optical magnitudes (Figure 3.15). However, there is a fraction of sources with low L_X which shows X-ray to optical properties consistent with those of sources with $L_X > 10^{42} \text{ erg s}^{-1}$: 17% and 31% of sources with $L_X < 10^{42} \text{ erg s}^{-1}$ lie inside the *Chandra COSMOS-Legacy* X-ray to optical flux ratio locus in the soft and hard bands, respectively. The fraction is considerably higher in the hard band, where it is more likely to observe obscured AGN.
4. We confirm the existence of a correlation between X/O and the luminosity in the 2-10 keV band for Type 2 sources (Figure 3.17). We also confirm that at faint magnitudes ($i > 25$) the trend is weaker, and our data suggest that this happens at higher redshifts, where the sSFR is higher and the AGN contribution to the optical emission is less significant than the one in the X-ray.
5. We extend to low luminosities the well known, inverse correlation between the fraction of obscured AGN and the hard band luminosity: the fraction of optically classified obscured AGN is of the order of 90% at $L_X < 10^{42} \text{ erg s}^{-1}$ and drops to $\simeq 40\%$ at $L_X > 10^{43.5} \text{ erg s}^{-1}$. The observed behavior is fairly consistent with the Treister et al. (2009) AGN synthesis model predictions. On the other hand, if an X-ray classification criterion based on the HR is adopted, the lack of a strong correlation between obscured fraction and luminosity is consistent with the Gilli et al (2007) and Miyaji et al. (2015) models predictions. A higher spectroscopic completeness, coupled with a proper X-ray spectral analysis would be needed to fully capture the dependence on luminosity of the obscured AGN fraction.

Chapter 4

The $z \geq 3$ sample of the *Chandra* COSMOS Legacy survey

4.1 Introduction

In this chapter, we present the sample of 174 AGN with $z \geq 3$ from the *Chandra* COSMOS-Legacy survey (Civano et al. submitted; Marchesi et al. submitted). The work is organized as follows: in Section 4.2 we describe our sample, and the its optical and X-ray properties; in Section 4.3 we analyze the sample 0.5-2 keV LogN-LogS, while in Section 4.4 we show the 2-10 keV comoving space density of our sample in two different luminosity ranges ($\log L_X = [43.6-44.1]$ and $\log L_X > 44.1$), and dividing the sample into unobscured and obscured sources; we also compare our results with previous studies and with different AGN synthesis models. In Section 4.5 we compare our results on the number density of $z > 3$ AGN with detailed models of quasar activation via mergers, and we discuss possible alternatives in light of our newest data. In Section 4.6 we show the BH mass estimate for a subsample of sources at $z > 3$. We discuss the resultswork in Section 4.7.

4.2 Properties of the high-redshift AGN sample

We described the properties of the whole *Chandra* COSMOS-Legacy survey in chapters 2 and 3. Here we remind that the X-ray source catalog contains 4016 point-like sources, detected with a maximum likelihood threshold value $\text{DET_ML} \geq 10.8$ in at least one of three bands (0.5-2, 2-7 and 0.5-7 keV). This threshold corresponds to a probability of $\simeq 2 \times 10^{-5}$ that a source in the catalog is actually a background fluctuation (Puccetti et al. 2009). At this threshold, the flux limit of the survey is 8.9×10^{-16} in the full band (0.5-10 keV), $2.0 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the soft band (0.5-2 keV) and $1.4 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the hard band (2-10 keV). The redshifts are obtained using the spectroscopic information, when available, and the photometric otherwise. 2076 sources (i.e. $\simeq 52\%$ of the whole sample) have a secure spectroscopic redshift, while 3877 sources (i.e. $\simeq 97\%$ of the whole sample) have a photometric redshift.

4.2.1 Spectroscopic redshifts

In the spectroscopic sample, 78 sources have redshift greater than 3, 11 have $z \geq 4$ and 2 sources have $z \geq 5$. A summary of the different telescopes and surveys used to obtain the spectroscopic sample of *Chandra COSMOS-Legacy* is reported in section 3.5.1. Here we remind that the source with the highest spectroscopic redshift, $z=5.3$, is also the only X-ray source detected in a proto-cluster, and it has been analyzed in Capak et al. (2011). In Figure 4.15 we show the K -band spectra of four BL-AGN from the *Chandra COSMOS-Legacy* $z \geq 3$ sample.

Five of the 78 sources have not been detected in the soft band, while other four have $\text{DET_ML} < 10.8$: five of these nine sources with no significant detection in the soft band have been detected in both hard and full band, while four have been detected only in full band. Sources with no soft band detection are candidate obscured sources, where the 2-10 keV (rest-frame, observed in ~ 0.5 -2 keV at $z > 3$) emission is absorbed.

4.2.2 Photometric redshifts

Chandra COSMOS-Legacy photo- z have been obtained following the procedure described in section 3.5: here we remind that the COSMOS field has been observed in 31 different bands, so the analysis of the spectral energy distribution (SED) of the sources is equivalent to low-resolution spectral analysis. The accuracy of the photo- z s with respect to the whole spectroscopic redshift sample is $\sigma_{\Delta z / (1+z_{spec})} = 0.02$, with $\simeq 11\%$ of outliers ($\Delta z / (1+z_{spec}) > 0.15$). For the purposes of this work it is also worth noticing that at $z \geq 3$ there are 9 outliers, but for the remaining 69 sources the agreement between spec- z and photo- z has the same quality of the whole sample, with a normalized median absolute deviation $\sigma_{NMAD} = 1.48 \times \text{median}(\|z_{spec} - z_{phot}\| / (1+z_{spec})) = 0.015$ (Figure 4.1). As a further check, we visually inspected all the SEDs of the sources with $z_{phot} \geq 3$, together with their best fits, to verify potential inaccuracies in the fit or in the SED data points. No source has been rejected after this visual analysis.

The photo- z computation produces a nominal value of the photo- z , corresponding to the maximum of a probability distribution function (Pdz), which gives the probability of a source to be in a given redshift bin (with redshift bins $\Delta z = 0.01$ for $z \leq 6$ and $\Delta z = 0.02$ for $6 < z \leq 7$): the integrated area of the Pdz over all redshift bins is, by definition, equal to 1. The agreement between the redshift distributions computed using, for each object, either the nominal values of the photometric redshifts or the entire distribution of the Pdz is good at all redshifts (Figure 4.3). In the rest of our work we use our photo- z as a probability weighted sum of contributions, instead that just use the photo- z nominal value as the Pdz maximum.

The sample of sources with photometric redshift contains 96 sources with $z \geq 3$ ($\simeq 55\%$ of the whole sample in this redshift range), 16 sources with $z \geq 4$ ($\simeq 59\%$ of the whole sample in this redshift range), 7 sources with $z \geq 5$ ($\simeq 78\%$ of the whole sample in this redshift range) and 4 sources with $z \geq 6$ (100% of the whole sample in this redshift range). The effective weighted contribution at $z \geq 3$ is shown in Figure 4.2 and is actually equivalent to have 74.3 sources with $z \geq 3$ in the sample. 30 of the 96 sources with only a photo- z have no significant detection in the soft band: 20 have no soft band detection at all, once again a potential indication

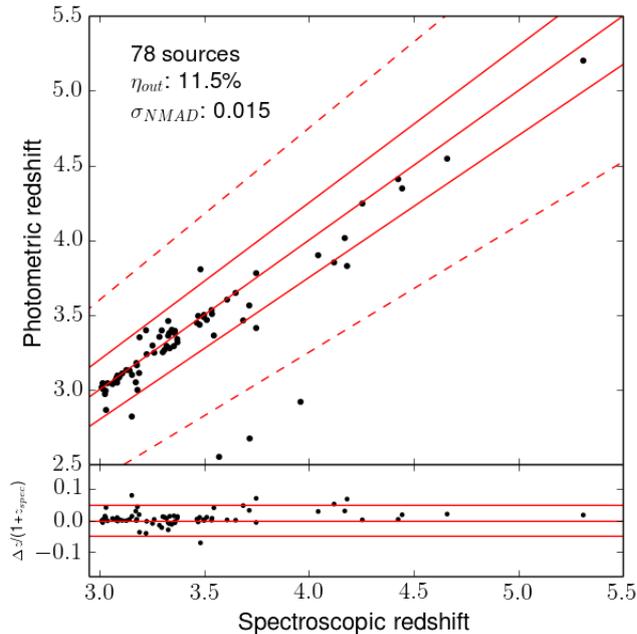


Figure 4.1: Spectroscopic versus photometric redshift for the 78 sources with a reliable spec- z in our sample. Red solid lines correspond to $z_{\text{phot}} = z_{\text{spec}}$ and $z_{\text{phot}} = z_{\text{spec}} \pm 0.05 \times (1 + z_{\text{spec}})$, respectively. The dotted lines limit the locus where $z_{\text{phot}} = z_{\text{spec}} \pm 0.15 \times (1 + z_{\text{spec}})$. Only three of the nine outliers, i.e., objects with $\Delta z / (1 + z_{\text{spec}}) > 0.15$, are shown here.

of high obscuration, while 10 have soft detection with $\text{DET_ML} < 10.8$. 17 of the sources with no significant soft detection have both hard and full band detection, the remaining 13 sources are detected only in the full band. It is interesting to note that the fraction of AGN without a significant soft detection is significantly higher among the AGN with photometric redshift (30/96, corresponding to $\simeq 31\%$) than among the AGN with spectroscopic redshift (9/78, corresponding to $\simeq 11\%$). This is an additional indication of potential different physical properties for the AGN in the two sub-samples, the sources with only a photo- z being likely more obscured, or fainter, or both.

There are also 309 sources in the *Chandra COSMOS-Legacy* catalog with only a photometric redshift < 3 , but which have a contribution to the Pd z at $z \geq 3$ (see, for example, the Pd z of source lid_1414¹, whose nominal photometric redshift value is 2.85, in Figure 4.4). All these 309 sources have been taken in account in our analysis using, for each of them, the contribution of each bin of redshift with Pd $z(z_{\text{bin}}) > 0$, weighted by the Pd z value itself: the effective contribution of these sources, i.e. the sum of all weights, is equal to add additional 39.7 sources to the $z \geq 3$ sample. In conclusion, the effective number of AGN with only photometric redshift at $z > 3$ is 114 (74.3 with nominal $z_{\text{phot}} > 3$ and 39.7 with nominal $z_{\text{phot}} < 3$). Further details

¹lid_ identifies new *COSMOS-Legacy* sources, while the cid_ prefix is used for sources already in the C-COSMOS catalog

are provided in the sections dedicated to the analysis of the number counts (Section 4.3) and of the space density (Section 4.4) of our high-redshift sample.

4.2.3 Summary

The *Chandra COSMOS-Legacy* high-redshift sample contains 174 sources with $z \geq 3$, 27 with $z \geq 4$, 9 with $z \geq 5$, and 4 with $z \geq 6$, plus other 309 sources with photometric redshift $z < 3$ and contribution to the PdZ at $z \geq 3$, whose effective contribution to the sample at $z \geq 3$ is equivalent to those of 39.7 sources. Taking in account the PdZ weighted contribution, our sample contains 192.0 sources. A summary of the distribution of these sources in the three adopted X-ray bands we is shown in Table 4.1; in the same Table, we also show how many of the sources are actually used in the computation of the number counts and of the space density, where a more conservative cut in flux limit (i.e., the flux at which at least 10% of the total area is covered) is applied to reduce the Eddington bias at faint fluxes. The fluxes at which these cuts are applied are $3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the soft band, $2 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the hard band and $1.2 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ in the full band.

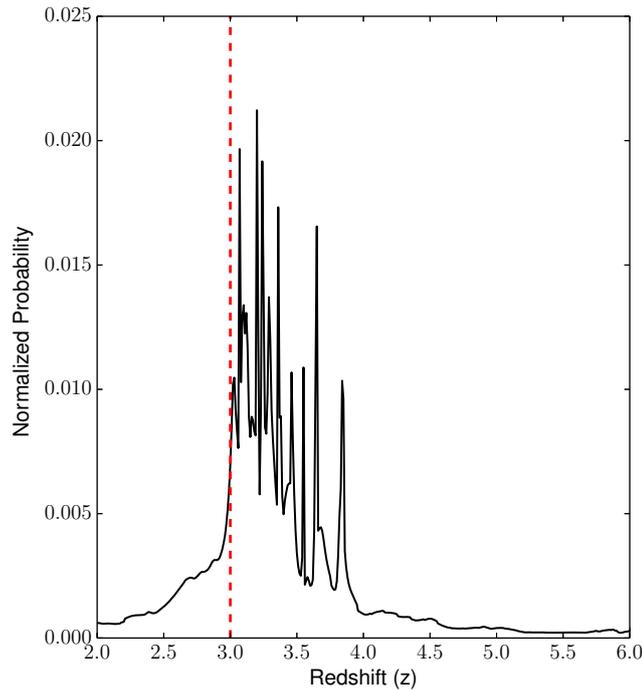


Figure 4.2: Normalized probability distribution function of redshift for all sources with $z > 3$: this distribution has also been used as probability distribution function of redshiftshigh-redshift for the 40 sources in the sample without optical counterpart.

Redshift	Total			Spec			Phot		
	S	H	F	S	H	F	S	H	F
$z>3$	135	12	27	69	3	6	66	9	21
$z>4$	20	2	5	10	0	1	10	2	4
$z>5$	6	1	2	2	0	0	4	1	2
$z>6$	3	0	1	0	0	0	3	0	1

Redshift	Total			Spec			Phot		
	S	H	F	S	H	F	S	H	F
$z>3$	132	9	19	69	2	4	63	7	15
$z>4$	19	1	3	10	0	1	9	1	2
$z>5$	5	1	1	2	0	0	3	1	1
$z>6$	2	0	1	0	0	0	2	0	1

Table 4.1: Top: number of sources in the high-redshift sample, divided by X-ray band adopted in the computation of the space density. First we use S (0.5-2 keV): if $\text{DET_ML}_S < 10.8$, we then use H (2-10 keV). If a source has $\text{DET_ML} < 10.8$ in both S and H, we then use the information from F (0.5-10 keV). Bottom: same as top, but with only the sources actually used in the computation of the space density, after the application of a cut in the flux limit.

4.2.4 Optically unidentified sources

As we said in section 3.5.3, 43 sources in the *Chandra COSMOS-Legacy* catalog are reliable X-ray sources without an optical *i*-band counterpart, but with a *K*-band or 3.6 μm IRAC counterpart, or with no counterpart at all. 19 of these sources have both a *K*-band and a 3.6 μm IRAC counterpart, 7 have only a *K*-band counterpart, 7 have only a 3.6 μm IRAC counterpart and 10 have no counterpart at all. Furthermore, 9 of these sources have no soft band detection, suggesting high obscuration.

These X-ray sources are high-redshift AGN candidates (Koekemoer et al. 2004), or highly obscured sources, or both. We take all of them into account in the estimate of the upper boundary of our 2-10 keV space density (Section 4.4), while we estimated the upper boundaries of the $z>3$ 0.5-2 keV LogN-LogS using the 34 sources detected in the soft band. To this purpose, we assume that each of these sources has a Pd z equal to the mean Pd z of all the sources in our sample with $z \geq 3$ (Figure 4.2). Then the contribution of the sources in each bin of redshift has been weighted by the value of the Pd z at that redshift, as described above. This Pd z distribution has a median (mean) redshift $z=3.20$ (3.41).

4.2.5 Optical properties

In our sample of 174 sources with nominal redshift value $z \geq 3$, 166 have *i*-band magnitude information (Capak et al. 2007, McCracken et al. 2010), 164 have *K*-band magnitude (Ilbert et al. 2009, Laigle et al. submitted) and 155 have 3.6 μm IRAC magnitude (Sanders et al. 2007, Laigle et al. submitted). We show the observed AB magnitude distribution in these three bands in Figure 4.5, dividing the sample in sources with spectroscopic (blue dashed line) and photometric redshift

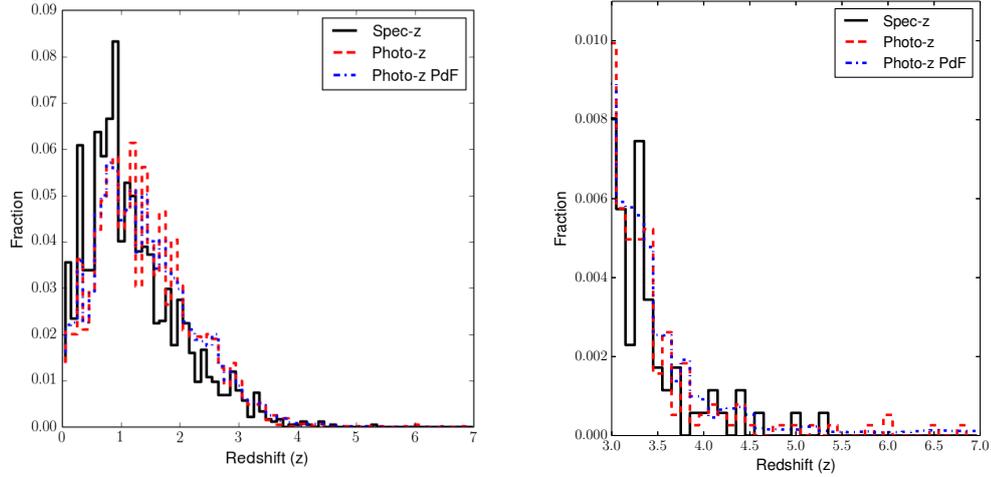


Figure 4.3: Normalized distribution of redshift for sources with a spectroscopic redshift (black solid line), a photometric redshift (red dashed line) and for the probability distribution function of all sources with a photometric redshift (blue dashed-dotted line), for all the sources in our sample (left) and for the subsample at $z \geq 3$. The agreement between the nominal value of the photometric redshifts and the distribution of the Pdz is good at all redshifts, which means that the majority of the sources have narrow and highly peaked Pdz.

only (red dashed line) sources. Sources with spectroscopic redshift have brighter optical magnitudes (mean magnitude $i_{AB}=23.3$) than sources with photometric redshift only (mean magnitude $i_{AB}=25.1$). The hypothesis that the two magnitude distributions are the same distribution is rejected by a Kolmogorov-Smirnov (KS) test, with a p-value $\simeq 1.7 \times 10^{-12}$. This difference in magnitude is less significant in K -band, where the mean magnitude for sources with spec- z is $K_{AB}=21.9$, while for sources with only photo- z is $i_{AB}=22.9$. Finally, in the $3.6 \mu\text{m}$ IRAC band the difference is of only 0.5 dex, the sources with spec- z having $IRAC_{AB}=21.7$, and those with only photo- z having $IRAC_{AB}=22.2$. Consequently, using photo- z s we cover the sub-sample of AGN with lower ultraviolet (UV) and optical (4000–6000 Å) rest-frame luminosity, which we observe in the i -band at the mean redshift of our sample. The spec- z and photo- z samples have instead similar luminosity distributions at longer wavelengths ($\simeq 6500$ – 9000 Å in the rest-frame, observed in the K -band). These objects can therefore be intrinsically fainter or more obscured than those for which we can provide a spec- z .

The mean and median magnitude values of the whole sample are 24.3, 22.4 and 22 in i -, K - and $3.6 \mu\text{m}$ IRAC band, respectively.

For most of the sources with an optical spectrum, we were able to determine the spectroscopic type of the source on the basis of the measured full width at half-maximum (FWHM) of the permitted emission lines. If one or more of these lines have $\text{FWHM} > 1000 \text{ km s}^{-1}$ (Vanden Berk et al. 2006; Stern & Laor 2012), we classify them as optical broad-line AGN (BLAGN), while sources with only narrow emission lines, or with only absorption lines, have been classified as non broad-line

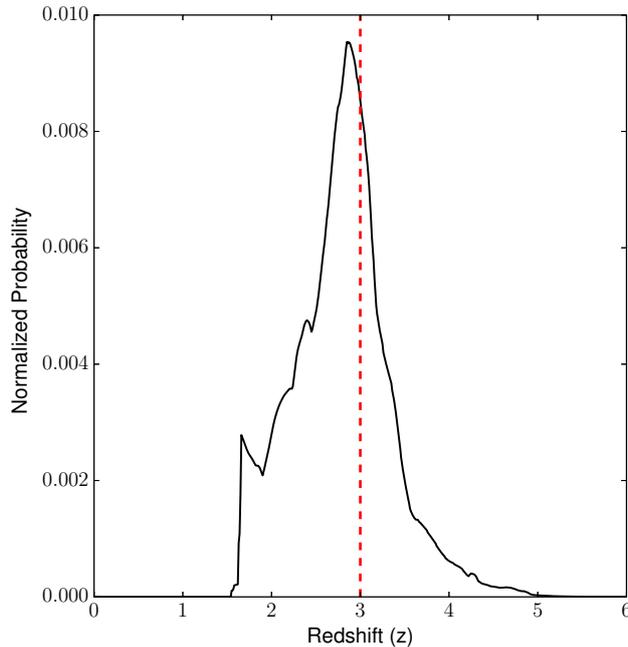


Figure 4.4: Normalized redshift probability distribution function for source lid_1414: this source has a photo- z value of 2.85, but it has $P_{dz} > 0$ at $z \geq 3$. The redshifts above the threshold, weighted by their P_{dz} , have been taken in account in the computation of the 0.5-2 keV LogN-LogS and in 2-10 keV comoving space density at $z > 3$.

AGN (non-BLAGN).

In our sample of 78 sources with spectral information, 43 are classified as BLAGN, while 32 are classified as non-BLAGN. For the remaining 3 sources, the spectral signal-to-noise ratio is not high sufficiently to draw safe conclusions on the presence or absence of broad lines. The mean i -band magnitude is brighter for BLAGN ($i_{AB}=22.6$) than for NLAGN ($i_{AB}=24.2$).

For the remaining 99 sources without spectral classification, but with a photometric redshift obtained by Salvato et al. (2011, in prep.), we used an optical classification based on the best fit of the SED, described in section 3.5.2. On the basis of this procedure, the sources in the *Chandra COSMOS-Legacy* catalog are divided in unobscured AGN, obscured AGN and galaxies. In our sample, 35 of the 99 sources are best fitted with an unobscured AGN template, 2 with an obscured AGN template and the remaining 62 with a galaxy template. The predominance of sources best fitted with a galaxy template is mainly due to the procedure of template selection before the actual fit: all extended sources with flux $f_{0.5-2} < 8 \times 10^{-15}$ erg s $^{-1}$ cm $^{-2}$ are fitted with a galaxy template, which best reproduces the SED of these usually optically faint galaxy-dominated sources (Salvato et al. 2011). Once again, the mean i -band magnitude is brighter for unobscured ($i_{AB}=24.0$) than for obscured sources ($i_{AB}=25.6$). It is worth noticing that for the 75 sources with spec-

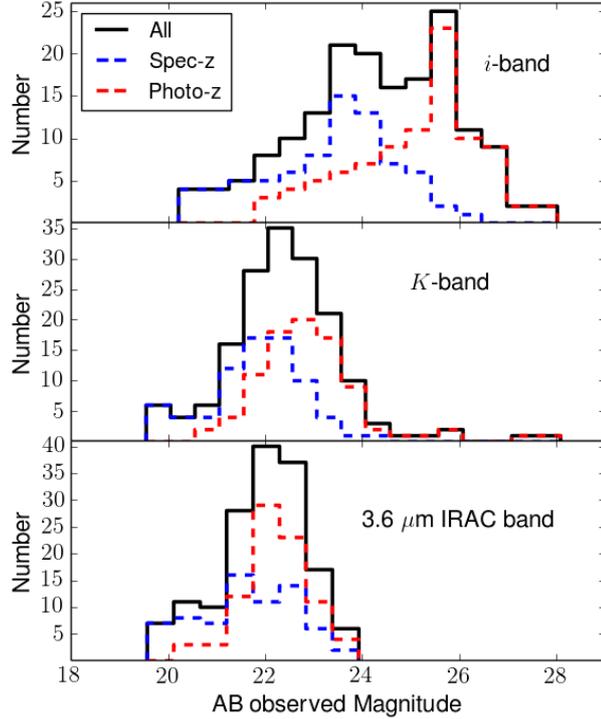


Figure 4.5: Observed AB magnitude distribution in *i*- (top), *K*- and 3.6 μm IRAC-band (bottom) for the whole sample of sources with $z \geq 3$ (black solid line), for the spectroscopic subsample (blue dashed line) and for the sources with only photometric redshift (red dashed line). Due to observational constraints, sources with spectroscopic redshift are also the optically brightest ones.

tral type the agreement between the spectral and the photometric classification is very good: 86% of spectroscopic BLAGN are best-fitted with an unobscured AGN template, while 90% of the spectroscopic NLAGN are best-fitted with an obscured AGN template or a galaxy template.

Summarizing, our sample contains 78 unobscured Type 1 sources (45% of the whole sample, 43 sources with spectral type, the remaining 35 with only photometric type) and 96 obscured Type 2 sources (55% of the whole sample, 32 sources with spectral type, the remaining 64 with photometric type). We show in Figure 4.6 the observed *i*-band AB magnitude distribution for all the sources with $z \geq 3$ (black solid line), for Type 1 AGN (blue dashed line) and for Type 2 AGN (red dashed line). The mean (median) *i*-band magnitude is 23.4 (23.2) for Type 1 AGN and 25.1 (25.2) for Type 2 AGN. The hypothesis that the two magnitude distributions are the same is rejected by a KS test, with a $p\text{-value} \simeq 1.7 \times 10^{-14}$.

Moreover, the majority (57%) of the sources with spectroscopic information are BLAGN, being brighter in *i*-band (see Figure 4.5), which, at the mean redshift of our distribution ($z \sim 3.5$), samples the so called “big blue bump”, emitting in the rest-frame UV; The majority (72%) of sources with only photometric information

are Type 2 AGN, which is consistent with the fact that these sources are intrinsically redder and fainter in the i -band (see also Brusa et al. 2010, Lanzuisi et al. 2013a).

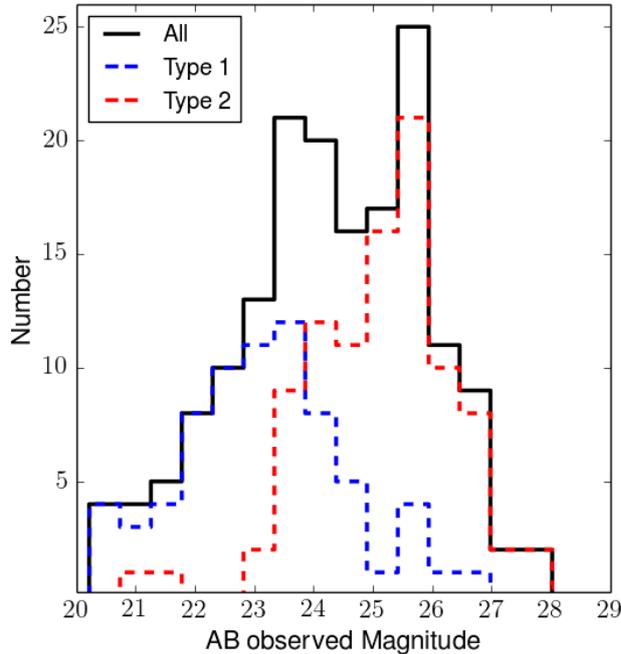


Figure 4.6: Observed AB magnitude distribution in i -band for the whole sample of sources with $z \geq 3$ (black solid line), for Type 1 or unobscured sources (blue dashed line) and for Type 2 or obscured sources (red dashed line).

4.2.6 X-ray properties

Previous studies (e.g. Kalfountzou et al. 2014) performed an analysis of their $z > 3$ sample on the basis of a pure X-ray classification, i.e. dividing the sample in obscured and unobscured sources assuming a threshold in hardness ratio (HR) at the redshift of each source. HR is derived as $HR = (C_{HB} - C_{SB}) / (C_{HB} + C_{SB})$, where C_{SB} and C_{HB} are the net counts in the soft band and hard band, respectively. Similarly, we derived HR for the sources in our sample, using the Bayesian estimate of Hardness Ratios (BEHR) method (Park et al. 2006; see Section 2.3.1). This tool is particularly effective in the estimate of uncertainties for the HR, even for sources in the low-count Poisson regime and/or for sources that have been detected only in the soft or in the hard band. For these sources, we performed aperture photometry on the *Chandra COSMOS-Legacy* dataset to obtain a measurement of the net counts in the band where the source was undetected.

As we explained in section 3.6.1, the HR is often used to derive a rough estimate of the column density of the source: in Figure 4.7 we show three curves of different column density ($N_H = 10^{21}, 10^{22}$ and 10^{23} cm^{-2} , dotted, dashed and solid line, respectively), obtained assuming a power-law spectrum with $\Gamma = 1.4$ (black) and $\Gamma = 1.8$ (green). As shown in Figure 4.7, the whole redshift range $z = [3-7]$ there is an

almost complete degeneracy between the $N_H=10^{21}$ and $N_H=10^{22}$ curves. This is due to the fact that at $z > 3$ the observed 0.5-2 keV band roughly corresponds to the rest-frame 2-10 keV band, while the rest-frame 0.5-2 keV is redshifted to energies lower than our observed energy range.

An immediate consequence of this degeneracy is shown while dividing our sample in optically classified Type 1 and Type 2 sources. The hypothesis that the two samples have the same HR distribution cannot be significantly rejected by a KS test, with a p-value $\simeq 0.1$. The mean (median) value of HR is -0.24 (-0.27) for Type 1 sources, only slightly lower than the mean and median value of HR for Type 2 sources, which is -0.19. Moreover, 26 of the 78 (33%) optically Type 1 sources have $HR > -0.19$ (i.e. lie above the threshold estimated assuming a typical obscured AGN spectrum, a power-law with slope $\Gamma=1.4$, see black dotted line in Figure 4.7), while 48 of 96 (50%) optically Type 2 sources have $HR < -0.19$ (i.e. $N_H < 10^{22} \text{ cm}^{-2}$ on the basis of the X-ray classification).

Finally, we studied the behavior with redshift of the HR: in Figure 4.7 we divide our sample in Type 1 (blue) and Type 2 (red) sources, on the basis of the optical classification. Both Type 1 and Type 2 AGN have large HR dispersions at any redshift, for the reasons described above. Nonetheless, the mean HR values of Type 2 AGN lie above the $N_H=10^{23}$ curve in the redshift bins $z=[3.4-4]$ and $z=[4.8-6.8]$, regardless the assumed slope, therefore suggesting that at least a fraction of Type 2 AGN are truly highly obscured sources.

Given the low reliability of the HR as an indicator of obscuration in the range of redshift analyzed in this work, we will not perform any analysis of the number counts and of the space density on the basis of the HR classification in the next sections.

4.3 0.5-2 keV AGN number counts

As a first step in our analysis, we produced the high- z LogN-LogS, i.e. we determined the number of sources $N(> S)$ per square degree at fluxes brighter than a given flux S ($\text{erg s}^{-1} \text{ cm}^{-2}$). We derived the LogN-LogS in the observed soft band at $z > 3$ and $z > 4$; for the first time we have a sample large enough to put constraints on the number counts also at $z > 5$ (9 objects) and $z > 6$ (4 objects). The number counts have been derived by folding our flux distribution through the sky coverage (i.e. the area of the survey covered at a given flux) of the *Chandra COSMOS-Legacy* survey (Civano et al. submitted).

The sensitivity curve is very steep in the flux regime close to the flux limit of the survey, therefore the uncertainties on the area estimate are larger here than at bright fluxes. To avoid these uncertainties, and to reduce the Eddington bias on our sample, we applied a cut in flux corresponding to 10% of the total area of the survey; we took in account in our analysis only sources with a 0.5-2 keV flux above $3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$. The sample used for the number counts therefore includes 132 sources at $z > 3$, 19 at $z > 4$, 5 at $z > 5$ and 2 at $z > 6$. We also used the contribution of sources with nominal photo- z value $z < 3$, but with $\text{Pd}z > 0$ in at least one bin of redshift $z \geq 3$: there are 188 of these sources with a $\text{Pd}z$ weighted contribution to the final LogN-LogS at $z > 3$, 39 at $z > 4$, 16 at $z > 5$ and 8 at $z > 6$.

We computed the cumulative source distribution with the equation:

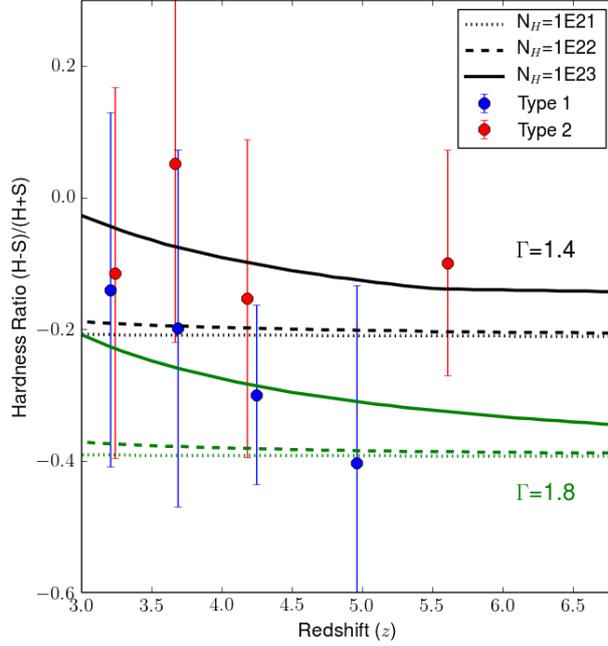


Figure 4.7: HR evolution with redshift for optically classified Type 1 (blue) and Type 2 (red) sources. The error bars represent the 68% dispersion. Three curves of different N_H (10^{21} cm^{-2} , dotted line, 10^{22} cm^{-2} , dashed line, 10^{23} cm^{-2} , solid line) are also plotted for comparison, obtained assuming a power-law spectrum with $\Gamma=1.4$ (black) and $\Gamma=1.8$ (green).

$$N(> S) = \sum_{i=1}^{N_S} \frac{w_i}{\Omega_i} [\text{deg}^{-2}], \quad (4.1)$$

where $N(> S)$ is the number of sources with flux greater than a given flux S , Ω_i is the sky coverage associated to the flux of the i th source, N_S is the number of sources above flux S and w_i is the weight associated to the Pdz contribution, $w_i = \frac{Pdz(z)}{\sum_0^7 Pdz(z)}$ ($w_i=1$ for sources with a spectroscopic redshift). Similarly, the associated error σ is

$$\sigma = \sqrt{\sum_{i=1}^{N_S} \left(\frac{w_i}{\Omega_i^2} \right)}. \quad (4.2)$$

We show our LogN-LogS relations (red circles) in Figures 4.8 ($z>3$, left, and $z>4$, right) and 4.9 ($z>5$, left, and $z>6$, right). We also estimated upper and lower boundaries of the logN-logS (plotted as black dashed lines limiting the yellow area), as follows:

1. for the upper boundary we computed Ω_i for each source adding to the observed flux the 1σ uncertainty on the flux, and we added to the sample also the 34

soft X-ray emitting sources with no optical counterpart, assuming for each of them a Pd z equal to the average Pd z of sources with $z > 3$ (see Section 4.2.4). With this second addition, we are assuming that all the non-detections in the optical bands are actually high-redshift X-ray selected sources;

2. for the lower boundary, we computed Ω_i for each source after subtracting the 1σ uncertainty on the flux to the observed flux.

In Figure 4.8 we also plot, for comparison, results from other surveys: in detail, number counts derived in Vito et al. (2013, 4-Ms *Chandra* Deep Field-South data, yellow squares), and Kalfountzou et al. (2014, C-COSMOS and Champ data, orange squares) are shown. These studies used datasets that cover the range from deep, pencil-beam area (CDF-S, 464.5 arcmin 2 , flux limit in the 0.5-2 keV band $f_X \simeq 9.1 \times 10^{-18}$ erg s $^{-1}$ cm $^{-2}$, Xue et al. 2011), to large, non contiguous areas and intermediate depth, combining C-COSMOS (0.9 deg 2 , flux limit in the 0.5-2 keV band $f_X \simeq 1.9 \times 10^{-16}$ erg s $^{-1}$ cm $^{-2}$, Elvis et al. 2009) and Champ ($\simeq 30$ deg 2 , flux limit in the 0.5-2 keV band 3.7×10^{-16} erg s $^{-1}$ cm $^{-2}$, Kim et al. 2007; Green et al. 2009).

Our results obtained using the whole sample with $z \geq 3$ are in general agreement with all the other studies, both at bright and faint fluxes. At $z > 4$, our sample of 27 sources is the largest sample on a contiguous field, about three and two times larger than those in 4 Ms CDF-S and C-COSMOS (9 and 14 sources, respectively). Moreover, it has the same size of the combined C-COSMOS and Champ sample (27 sources), and it is largely complementary to the latter sample. In fact, only $\simeq 40\%$ of the sources in our survey come from C-COSMOS, and Champ contribution is significant only at fluxes brighter than 3×10^{-15} erg s $^{-1}$ cm $^{-2}$, where the contribution of *Chandra COSMOS-Legacy* is instead negligible. Our number counts are slightly lower, but consistent within the uncertainties, than those in Civano et al. (2011) and Kalfountzou et al. (2014) at $f_X < 5 \times 10^{-16}$ erg s $^{-1}$ cm $^{-2}$. Our data also show a trend consistent with the results from Vito et al. (2013) at $f_X < 3 \times 10^{-16}$ erg s $^{-1}$ cm $^{-2}$.

Due to our good statistics, we are able to improve the constraints on predictions of different phenomenological models. We show these predictions in Figures 4.8 and 4.9, as black curves.

1. The thick solid lines show the predictions of the XRB synthesis model of Gilli et al. (2007). This model is based on the extrapolation of the X-ray luminosity function observed in a low-redshift regime (Hasinger et al. 2005), parametrized with an LDDE model and with a high-redshift exponential decline, as in Schmidt, Schneider & Gunn (1995): in detail, $\Phi(z) = \Phi(z_0) \times 10^{-0.43(z-z_0)}$ (with $z_0 = 2.7$). This model was developed in order to fit the optical luminosity function in the redshift range $z = [2.5-6]$ (Fan et al. 2001). It is worth noticing that the Gilli et al. (2007) model *without* the high- z exponential decline has already been ruled out in several works (e.g. Kalfountzou et al. 2014), so we excluded it from our analysis.
2. The dashed lines show the predictions of the LADE model (Aird et al. 2010), developed fitting the hard X-ray luminosity function derived in the same work using the 2 Ms *Chandra* Deep Fields and the AEGIS-X survey (200 ks).

3. The dotted lines are derived from the model of X-ray background population synthesis by Treister, Urry & Virani (2009).

At $z > 3$ (Figure 4.8, left) our results indicate that in this range of redshifts a decline in the number of sources is needed also in the X-ray, as assumed by the Gilli et al. (2007) model with the exponential decline, and not only in the optical band, showing therefore a discrepancy with the model of Treister, Urry & Virani (2009) model (dotted line). In fact, their predictions are too high by a factor $\simeq 2$ in comparison to our data, at any flux. In this redshift range the LDDE model with an exponential decline (solid line) and the LADE model (thick dashed line) predictions are too close to discard one of the two models.

A different behaviour is observed at $z > 4$ (Figure 4.8, right), where our results are in excellent agreement with the LDDE model with an exponential decline. The LADE model overpredicts the number of sources in the whole 0.5-2 keV flux range [3×10^{-16} - 1.4×10^{-15} erg s $^{-1}$ cm $^{-2}$], especially at fluxes fainter than 5×10^{-16} erg s $^{-1}$ cm $^{-2}$, even with respect to our upper boundary. Our result confirms and improves the one reported in Kalfountzou et al. (2014), which also showed a good agreement between the data and LDDE model with an exponential decline, and ruled out the LADE model at $z > 4$ at bright fluxes.

Finally, we show in Figure 4.9 the first analysis ever of X-ray selected AGN number counts at $z > 5$ (left) and $z > 6$ (right). At $z > 5$, our data (red circle) are in agreement with the LDDE model with an exponential decline (solid line), while the LADE model overpredicts the number of 0.5-2 keV sources, even on the basis of our upper boundary. Finally, at $z > 6$ our data are slightly above the predictions of the LDDE model with an exponential decline (solid line), although in agreement within 1σ .

4.4 2-10 keV comoving space density

For the computation of the space density in the 2-10 keV band, we applied the flux cut described in Section 4.2.3: it is worth noticing that with this criterion the source at $z = 6.84$ is excluded from the sample. We report a summary of the final number of sources included in the space density sample in Table 4.1 (bottom). More than 80% of the sources in the sample are detected in the 0.5-2 keV observed band, that at $z > 3$ roughly corresponds to the 2-10 keV rest-frame band. However, to complete our analysis, we used the extrapolated 2-10 keV rest-frame luminosity also for the sources with no soft band detection, using first the 2-10 keV observed flux and, for those sources with no 2-10 keV detection, the 0.5-10 keV observed flux.

We computed the comoving space density using the $1/V_{Max}$ method (Schmidt 1968). This technique has been developed to be used in cases where the survey area is flux dependent, as in our survey. We also worked with the assumptions described in Avni & Bahcall (1980), which take into account the fact that each source could in principle have been found in any region of the survey, and therefore at any X-ray depth.

For every redshift associated to a source in our sample, spectroscopic or photometric with an associated $Pdz(z_{bin}) > 0$ in at least one bin of redshift $z_{bin} \geq 3$, we

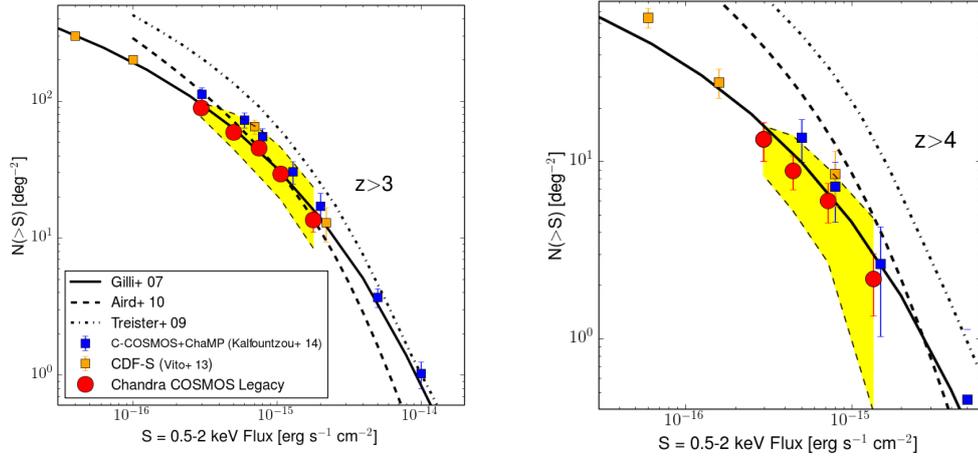


Figure 4.8: LogN-LogS relation in the 0.5-2 keV band for *Chandra COSMOS-Legacy* (red circles), for $z>3$ (left) and $z>4$ (right). Results from Vito et al. (2013, 4 Ms CDF-S, orange squares), and Kalfountzou et al. (2014, C-COSMOS and Champ data, blue squares) are also shown for comparison, together with models from Gilli et al. with an exponential decline (2007, solid line), Aird et al. (2010, dashed line) and Treister et al. (2009, dotted line). The yellow area in the left figure is obtained computing the space density adding and subtracting to the flux value its 1σ error. In the computation of the upper boundary the weighted contribution of sources with no optical counterpart is also taken in account.

computed the maximum available volume over which the source can be detected, using the equation

$$V_{max} = \int_{z_{min}}^{z_{max}} w(z) \Omega(f(L_X, z)) \frac{dV}{dz} dz, \quad (4.3)$$

where w is the weight linked to the Pdz contribution, $w = \frac{Pdz(z)}{\sum_0^7 Pdz(z)}$ ($w=1$ for sources with a spectroscopic redshift), $\Omega(f(L_X, z))$ is the sky coverage at the flux $f(L_X, z)$ observed from a source with redshift z and intrinsic luminosity L_X . z_{min} is the lower value of the redshift bin and z_{max} is the minimum value between the maximum observable redshift of the source at the flux limit of the survey and $z_{up,bin}$, the upper value of the redshift bin. The 2-10 keV rest-frame luminosities are estimated assuming $\Gamma=1.4$; no absorption correction is applied. We took the flux in the first available band where $DET_ML > 10.8$, starting from 0.5-2 keV, then 2-10 keV and finally 0.5-10 keV. The flux was then converted to the 2-10 keV luminosity, using the equation

$$L_{2-10keV,rest} = \frac{4\pi d_l(z)^2 \times (10^{2-\Gamma} - 2^{2-\Gamma})}{(E_{max}(1+z))^{2-\Gamma} - (E_{min}(1+z))^{2-\Gamma}}, \quad (4.4)$$

where E_{min} and E_{max} are the minimum and maximum energies in the range where the flux is measured, and $d_l(z)$ is the luminosity distance at the given redshift.

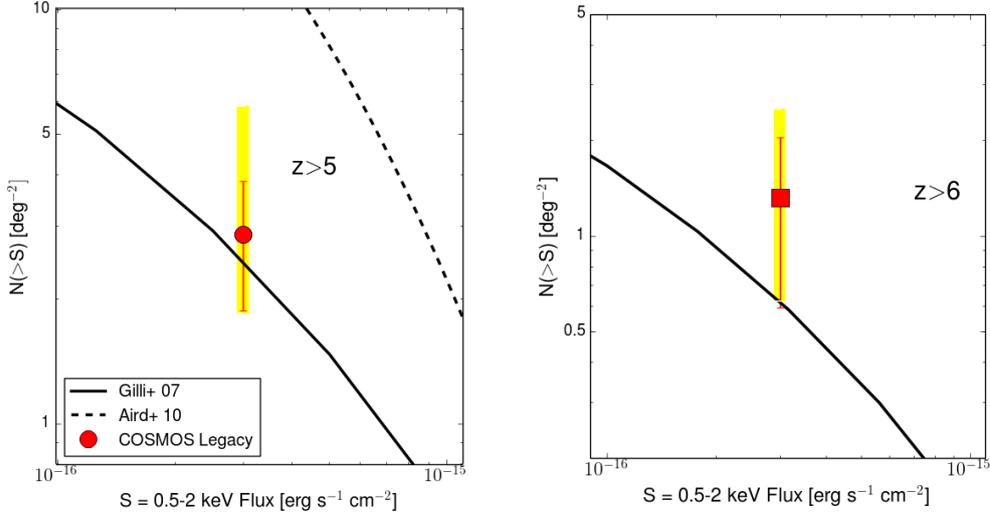


Figure 4.9: LogN-LogS relation in the 0.5-2 keV band for *Chandra COSMOS-Legacy* (red circles), for $z > 5$ (left) and $z > 6$ (right). Models from Gilli et al. (2007, black solid line) and Aird et al. (2010, black dashed line) are also shown for comparison. The yellow area in the left figure is obtained computing the space density adding and subtracting to the flux value its 1σ error. In the computation of the upper boundary the weighted contribution of sources with no optical counterpart is also taken in account.

Finally, we summed the reciprocal of all V_{max} values in each redshift bin $[z_{min}-z_{max}]$ in order to compute the comoving space density value Φ :

$$\Phi = \sum_{i=1}^{z_{min} < z < z_{max}} \left(\frac{1}{V_{max,i}} \right). \quad (4.5)$$

The 1σ error on the space density is

$$\sigma_{\Phi} = \sqrt{\sum_{i=1}^{z_{min} < z < z_{max}} \left(\frac{1}{V_{max,i}} \right)^2}. \quad (4.6)$$

where i is the index of each of the sources in the given bin of redshift. This means that for a source with photometric redshift $V_{max,i}$ is actually the sum of each of the V_{max} in that bin of redshift where $Pdz > 0$.

We divided our sample in two different luminosity ranges for completeness reasons (see Figure 4.10). The high-luminosity space density has therefore been computed in six redshift bins in the range $z=3-6.85$, with $\text{Log}(L_X) > 44.1$. The low-luminosity space density, instead, has been computed in three redshift bins in the range $z=3-3.5$, with $43.6 \leq \text{Log}(L_X) < 44.1$. These luminosity ranges are slightly different from those adopted for C-COSMOS (Civano et al. 2011), where the low-luminosity range was $\text{Log}_{10}(L_X) = [43.55-44.15]$ and the high-luminosity one was

$\text{Log}_{10}(L_X) > 44.15$. This difference is due to the fact that Civano et al. (2011) a power-law with $\Gamma=2$ was adopted in the flux computation.

To improve our analysis, we estimated the upper and lower boundaries of the space densities, using as input parameters in the X-ray fluxes plus or minus their 1σ uncertainties.

In the computation of the upper boundary we also take into account the 40 sources with no optical counterpart. As explained in Section 4.2.4, these sources are candidate high-redshift AGN. For each source, we assumed as Pd z the mean Pd z for all the sources in our sample with $z \geq 3$ (Figure 4.2). We then computed the space density for this subsample with the same technique described above. The values of Φ that we obtained have then been summed to the upper boundary obtained using $f_X + \sigma(f_X)$ as input parameter.

Our results are shown in Figure 4.11, with the high-luminosity space density on the left and the low-luminosity one on the right. We compare our results (red dots) with the results of Vito et al. (2014, orange squares). We found that their results are in good agreement with our data, with the difference in normalization caused by the different slope assumed in the flux estimate ($\Gamma=1.4$ in our work, $\Gamma=1.8$ in Vito et al. 2014).

We also compared our results with the predictions from the LDDE model with an exponential decline from Gilli et al. (2007, black solid line), Ueda et al. (cyan dashed line), Miyaji et al. (2015, green solid line), and with those from the FDPL model of Aird et al. (2015, black dashed line). We described the Gilli et al. (2007) and the Aird et al. (2015) models in Section 4.3. The Ueda et al. (2014) and the Miyaji et al. (2015) models are both derivations of the LDDE model.

As for the number counts results (Section 4.3), the FDPL model is higher than our data by a factor 2-4 at $3 < z < 5$, even if the upper boundary is considered, at high luminosities (Figure 4.11, left). Moreover, our data lie below the predictions of the various LDDE with an exponential decline models by a factor of 2-2.5 in the redshift range $z=[3-4]$. The predictions of the model are in agreement with our data in the high-redshift regime, $z \geq 5$. The best linear fit to our data, $\Phi = a + bz$ (see Table 4.2), has a slope $b = -0.45 \pm 0.04$, but there is a significant difference between the steeper slope in the redshift range $z=[3-4]$, $b = -0.61 \pm 0.09$, consistent with the slope of the Gilli et al. (2007) LDDE model with an exponential decline at $z > 3$, $b = -0.53$, and the flatter slope in the redshift range $z=[4-6]$, $b = -0.37 \pm 0.05$. This result is in agreement with the one obtained with the number counts analysis, both suggesting a less strong decline in the AGN number at $z > 4$.

We also show results from optical surveys such as Masters et al. (2012, black diamonds, left), McGreer et al. (2013, blue diamonds, left), Ikeda et al. (2011, black diamonds, right) and Glikman et al. (2011, purple diamonds, right). To compare the optical results to ours in *Chandra COSMOS Legacy*, we assumed the relation between the X-ray luminosity at 2 keV, L_{2keV} , and the luminosity at 1500 Å, L_{1500} , from Young et al. (2010),

$$\alpha_{OX} = 1.929 - 0.119 \log L_{1500}, \quad (4.7)$$

with

$$\alpha_{OX} = (\log L_{2keV} - \log L_{1500}) \left(\log \frac{\nu_{2keV}}{\nu_{1500}} \right)^{-1}. \quad (4.8)$$

We then integrated the luminosity functions of Masters et al. (2012) and McGreer et al. (2013) down to $M_{1450} = -24.5$, corresponding approximately to $\text{Log}(L_X) \sim 44.1$, and we compared them with our high-luminosity space density. The slope derived from the optical surveys ($b = -0.68 \pm 0.02$) is in good agreement with our data in the redshift range $z = [3-4]$ and with the Gilli et al. (2007) model; the normalization is instead significantly different, due to the large fraction of obscured sources that are missed in the optical band.

Unlike in the high-luminosity regime, we are not able to put any constraint on the fainter luminosities regime (Figure 4.11, right). This is due to both the small difference between the LDDE and FDPL models in the redshift range where our survey is complete in this range of luminosities, i.e. $\text{Log} L_X = [43.6-44.1]$, and because of the size of the errors associated to our data points, especially in the first redshift bin. It is also worth noticing that in this luminosity range we are only able to compute a lower limit on the space density, using as input parameter the flux value minus the 1σ error on the flux (red solid line): in fact, space density values computed using as input parameter the flux value plus the 1σ error are lower than those computed with the nominal value of the fluxes. This result is again due to the fact that some of the sources close to the flux limit of the survey have considerably smaller values of Φ once the 1σ error is added to the nominal value of the flux. However, we find that our results are in rough agreement with those of Vito et al. (2014).

In Figure 4.11, right panel, we also show the optical luminosity functions of Ikeda et al. (2011, black diamonds) and Glikman et al. (2011, purple diamonds): we integrated their luminosity functions in the absolute magnitude range $M_{1450} = [-23.5; -21.8]$. We are not able to distinguish between the two different optical results: in fact, the slope of our space density is steeper than the models predictions ($b = -1.31 \pm 0.84$) and fit well the results of Ikeda et al. (2011) at $z = 4$. If we instead remove the point at $z = 3.1$ (the one most affected by uncertainties) the trend ($b = -0.46$) seems more in agreement with the result of Glikman et al. (2011), and with the predictions of both the different LDDE models with an exponential decline and the FDPL model. However, in this luminosity range our statistics is not large enough to make strong assumptions and safely rule one of the two optical results.

4.4.1 Obscured versus unobscured AGN

The high-redshift decline of space densities has been measured in both optical and X-ray selected AGN samples. Therefore, given that X-ray selected samples suffer considerably less obscuration bias compared to those optical selected, a similar trend should imply that the fraction of obscured AGN does not change significantly above $z = 3$. In fact, previous works showed an increase in the fraction of obscured objects in the redshift range $z = [1-2]$ (e.g., Ballantyne et al. 2006; Iwasawa et al. 2012), followed by a decline of this fraction at higher redshifts (Hasinger 2008; see also Gilli 2010 for a general review and an analysis of possible selection biases). We test this result with our sample, which we divide in two subsamples on the basis either of the

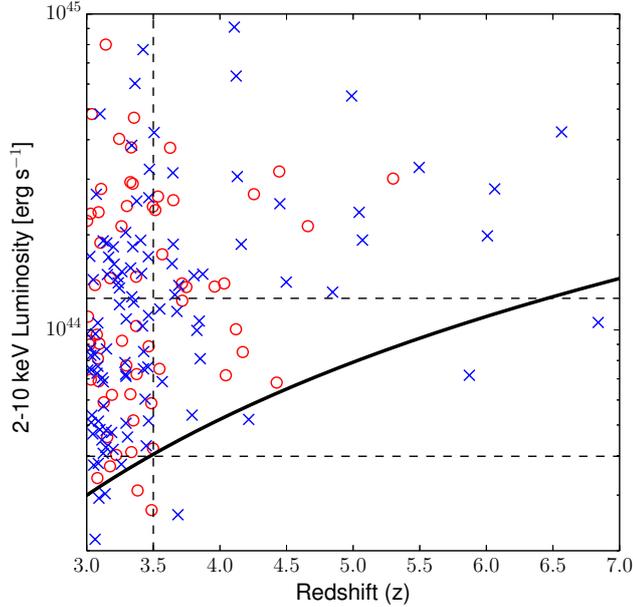


Figure 4.10: Evolution with redshift of the 2-10 keV K-corrected luminosity for all sources in our sample with a spectroscopic (red circles) or a photometric (blue crosses) redshift. The black solid line shows the 10%-area limit of the survey, while the black dashed lines highlight the two subsamples used in the computation of the space density.

spectroscopic classification (where available, i.e. for 75 sources) or the best fitting SED template. More details are reported in Section 4.2.5; here we summarize that 78 sources with nominal redshift value $z \geq 3$ are classified as Type 1, or unobscured, while the remaining 96 are classified as Type 2, or obscured. For the analysis of the space density, however, we also take into account (as for the general case) the weighted contribution of those sources with photometric redshift $z < 3$ and $Pdz > 0$ in at least one bin of redshift with $z \geq 3$.

The space densities for sources with $L_X > 10^{44.1}$ (left) and $10^{43.6} \leq L_X < 10^{44.1}$ (right), in the 2-10 keV band, are shown in Figure 4.12. The sample of type 1 AGN is plotted with blue circles, while the sample of type 2 AGN is plotted with red squares. Our results are also compared with the predictions of the LDDE model with an exponential decline of Gilli et al. (2007), where the contribution of sources with $N_H \leq 10^{22} \text{ cm}^{-2}$ (i.e. the unobscured ones) is plotted as a black solid line, while the contribution of sources with $N_H > 10^{22} \text{ cm}^{-2}$ (i.e. the obscured ones) is plotted as a black dashed line. At high luminosities (left in Figure 4.12), the unobscured sources ($b = -0.65 \pm 0.07$) are in good agreement with the predictions of the model, although a potential stronger decline is present in the data at $z > 5.5$, where the number of unobscured sources is about a factor of two smaller than the predictions of the model. The trend of decline in obscured sources is instead flatter ($b = -0.34 \pm 0.04$) than the predictions of the model, with the number of obscured sources being smaller than the predictions of the model by a factor $\simeq 2$ at $z < 4$,

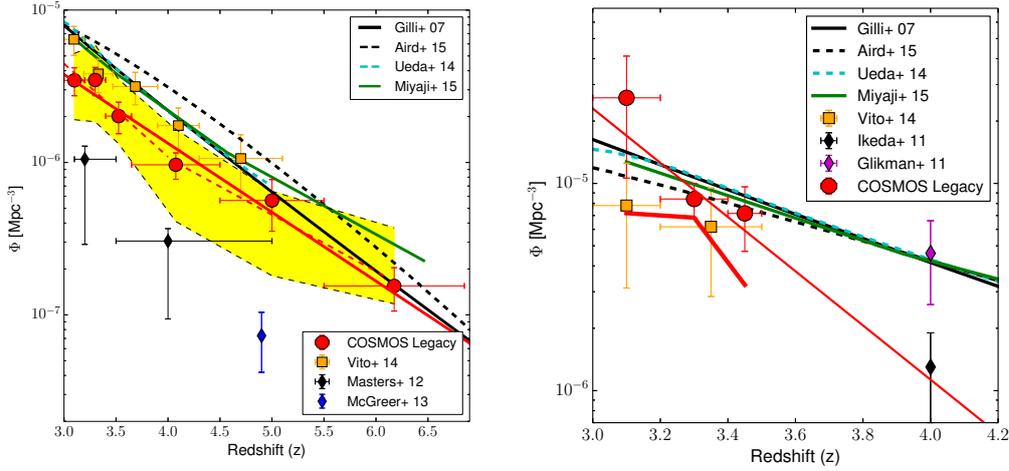


Figure 4.11: Space density for sources with $L_X > 10^{44.1}$ (left) and $10^{43.6} \leq L_X < 10^{44.1}$ (right), in the 2-10 keV band. The *Chandra COSMOS-Legacy* survey results are plotted with red dots, while results from Vito et al. (2014, orange squares) are also shown for comparison, together with optical space density from Masters et al. (2012, black diamonds) and McGreer et al. (2013, blue diamonds). Four different models of X-ray population synthesis are also shown, those of Gilli et al. (2007, black solid line), Aird et al. (2015, black dashed line), Ueda et al. (2014, cyan dashed line) and Miyaji et al. (2015, green solid line). The red solid line is the best fit to the *Chandra COSMOS-Legacy* data, assuming an equation $\text{Log}(\Phi) = a + b \times z$, while the red dashed lines (left) show the two different slopes of our high-luminosity space density in different redshift range, $b = -0.61 \pm 0.09$ for $z = [3-4]$ and $b = -0.37 \pm 0.05$ for $z = [4-6]$. The yellow area in the left figure is obtained computing the space density adding and subtracting to the flux value its 1σ error. In the computation of the upper boundary the weighted contribution of sources with no optical counterpart is also taken in account. The red bold solid line in the right figure is a lower limit to our space density and is obtained computing the space density subtracting to the flux value its 1σ error.

while at $z > 4$ the data and the model agree. The ratio between obscured and unobscured sources is $\simeq 1$ in the redshift range $z = [3-4]$, while it grows to $\simeq 2$ at $z = 5$ and further increases at $z > 5.5$.

We also compare our results with those from the optical surveys of Masters et al (2012, black diamonds, left) and McGreer et al. (2013, blue diamond, left): there is a good agreement (within 1σ) between the optical space densities and our unobscured space density, which also have consistent slopes ($b = -0.68 \pm 0.02$ and $b = -0.65 \pm 0.07$ for the unobscured X-ray sources). This result is consistent with our expectations, due to the fact that the optical surveys are limited to Type 1, unobscured sources.

At low luminosities (right in Figure 4.12) there are larger uncertainties, but our data are in rough agreement with the predictions of the LDDE model with decline for both unobscured and obscured sources.

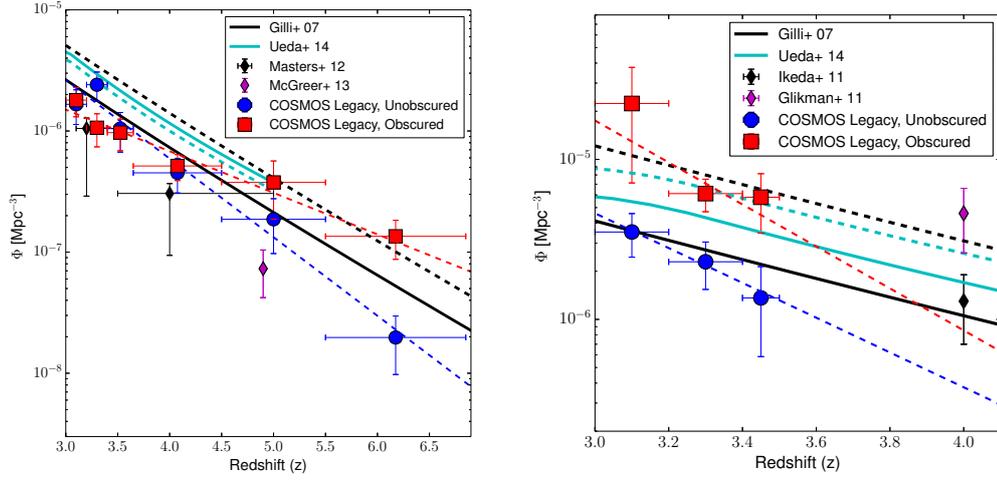


Figure 4.12: Space density for sources with $L_X > 10^{44.1}$ (left) and $10^{43.6} \leq L_X < 10^{44.1}$ (right), in the 2-10 keV band. The sample of type 1 AGN is plotted with blue dots, while the sample of type 2 AGN is plotted with red squares. The red and blue dashed lines are the best fit to the unobscured and obscured samples, respectively, assuming the equation $\text{Log}(\Phi) = a + b \times z$. The model of X-ray population synthesis from Gilli et al. (2007) is also shown, one with $N_H = [20-22]$ (black solid line), the other with $N_H = [22-26]$ (black dashed line); the model from Ueda et al. (2014) is shown in cyan, for $N_H = [20-22]$ (solid line) and $N_H = [22-24]$ (dashed line). Optical space densities from Masters et al. (2012, black diamonds, left), McGreer et al. (2013, magenta diamonds, left), Ikeda et al. (2011, black diamonds, right) and Glikman et al. (2011, magenta diamonds, right) are also shown for comparison.

L_X range	a	b
$\text{Log} L_X > 44.1$	-4.07 ± 0.17	-0.45 ± 0.04
$43.6 \leq \text{Log} L_X < 44.1$	-0.71 ± 2.79	-1.31 ± 0.84

Table 4.2: Parameters of the best fit of the space density in two intervals of luminosity, where the fit model is described by the equation $\text{Log}(\Phi) = a + b \times z$. In this fit we also take into account the uncertainty on Φ .

	Type 1		Type 2	
	a	b	a	b
$\text{Log} L_X > 44.1$	-3.63 ± 0.27	-0.65 ± 0.07	-4.79 ± 0.16	-0.34 ± 0.04
$43.6 \leq \text{Log} L_X < 44.1$	-2.07 ± 0.54	-1.09 ± 0.16	-0.82 ± 3.73	-1.31 ± 1.12

Table 4.3: Parameters of the best fit of the space density for both obscured and unobscured sources, in two intervals of luminosity, where the fit model is described by the equation $\text{Log}(\Phi) = a + b \times z$. In this fit we also take into account the uncertainty on Φ .

4.5 Comparison with merger models

In this section we compare our results with those predicted by the basic quasar activation merger model by Shen (2009). Basic quasar activation merger models have been developed to constrain the accretion mechanism of BH growth and to disentangle between models of BH and galaxy co-evolution: mergers have been proposed as efficient triggering mechanisms for especially luminous quasars (e.g., Barnes & Hernquist 1991, Menci et al. 2014). The model outlined in Shen (2009) was built upon the dark matter halo (major) merger rate extracted from numerical simulations, which provides the number of triggering events per unit time, convolved with an assumed AGN light curve, which characterizes the evolution of individual quasars. The light curve is described by an exponentially ascending phase, and a power-law descending phase. The end of the exponential growth is controlled by an AGN feedback self-regulation condition between the peak luminosity and the host dark matter haloes of the type (e.g., Wyithe & Loeb 2003) $L_{\text{peak}} \propto M_{\text{halo}}^{5/3}$, valid in the whole range of host halo masses above $M_{\text{halo}} > 2 \times 10^{11} M_{\odot}/h$. The parameters of the model were tuned by Shen (2009) to broadly reproduce the full bolometric, obscuration-corrected, AGN luminosity function at $0.1 < z < 6$, as well as the available large-scale clustering measurements of optical quasars available at the time.

In our work, the aim is to use the space density at high-redshift to possibly constrain the accretion mechanisms of BH growth and to disentangle between models of BH and galaxy co-evolution. Following Civano et al. (2011) and Allevato et al. (2014), we compare the Shen (2009) merger model with the newest available AGN data at $z > 3$, including the ones presented in this work.

Figure 4.13 shows that the predictions of the reference merger model (black solid line) match well with high-luminosity part of the optical quasar luminosity function (LF) in the redshift range $z = [3.08-3.27]$ by BOSS (Ross et al. 2013). For this comparison, we corrected the model LF by a luminosity-dependent fraction from Ueda et al. (2014) to account only for Type 1 unobscured sources with $N_H < 21$. The model predictions, however, tend to gradually overestimate the observed number counts when moving to fainter luminosities ($L_{\text{bol}} < 10^{47} \text{ erg s}^{-1}$). This is even more evident when comparing the Shen (2009) model with the number densities of fainter AGN derived in this work and others in the literature (Figure 4.14). The reference model (black solid line) is higher than the data by a factor of 3 to 10, depending on the redshift. This behaviour is not fully unexpected. The Shen (2009) model was calibrated mostly on bright AGN at $z > 3$, while the faint AGN data available at the time were poor.

At fixed redshift, the parameters defining the model seem to be well suited to reproduce the bright end of the AGN luminosity, but tend to fall short in matching the most up-to-date number counts from X-ray surveys. There are two main ways to improve the match between merger models and data: to modify either the light curve or the host halo mass distribution, or a combination of both. The black, dotted lines in Figures 4.13 and 4.14 mark the predictions from a modified Shen (2009) model in which we cut out the post-peak descending phase, with all other parameters held fixed. This alteration naturally represents an improvement with respect to the faint-end luminosity function, because a smaller number of low-luminosity AGN are

now predicted by the model, though it also tends to cause an under-prediction of the bright-end of the AGN LF. Alternatively, we have explored a second variant to the Shen (2009) model, characterized by a steepening in the $L_{\text{peak}}-M_{\text{halo}}$ relation below $M_{\text{halo}} \simeq 10^{12} M_{\odot}/h$, with $L_{\text{peak}} \propto M_{\text{halo}}^5$, implying that preferentially lower-luminosity quasars are now associated to more massive, less numerous host dark matter haloes. The outcome of this third model is shown with dashed lines in Figures 4.13 and 4.14. With this model the number densities of very luminous quasars are preserved, while those of lower luminosities AGN gradually decrease, in better agreement with the data. Evidence for a break in the black hole-galaxy scaling relations is also now claimed in the local universe (Scott & Graham 2013) and by independent theoretical models (Cirasuolo et al. 2005, Fontanot et al. 2006, Fontanot et al. submitted). An independent test of the Shen (2009) model will be presented in Allevalo et al. (in prep.) making use of the clustering analysis.

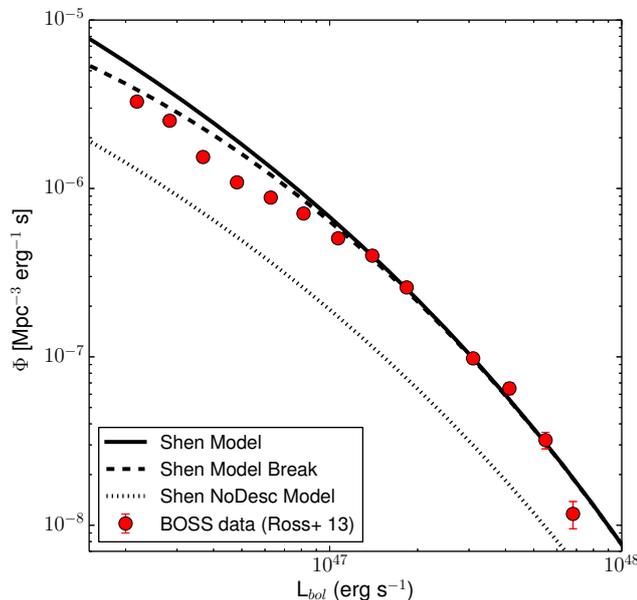


Figure 4.13: SDSS-III BOSS bolometric luminosity function computed in the redshift range $z=[3.08-3.27]$ (Ross et al. 2013, red dots). Different models from Shen (2009) are also plotted for comparison.

4.5.1 Alternatives to mergers

At face value, theoretical merger models predict enough, if not even too many, major mergers to account for all high-redshift AGN of moderate-to-high-luminosity. However, mergers may not be the unique driver for the evolution of AGN, especially at lower luminosities. Other “in-situ” processes such as disk instabilities and/or clumpy accretion may be effective in channelling flows of gas down towards the very center of the host galaxy, eventually fuelling the black hole (e.g., Bower et al. 2006; Bournaud et al. 2011; Di Matteo et al. 2012). Dedicated studies based on advanced

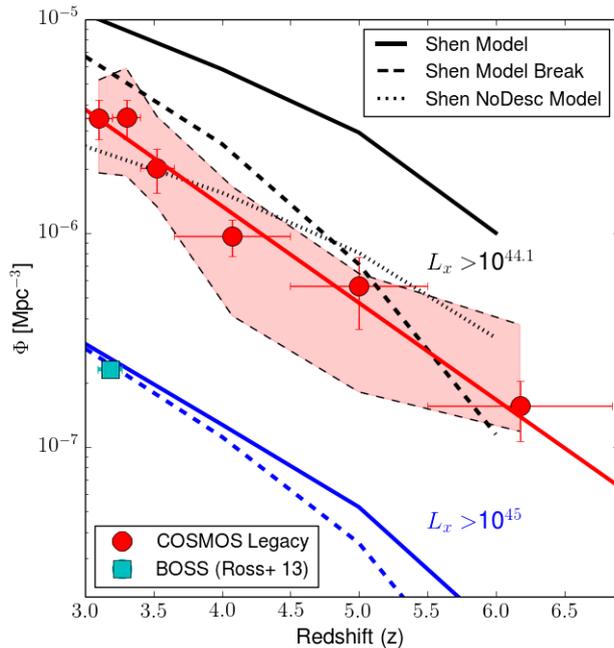


Figure 4.14: *Chandra* COSMOS-Legacy space density for sources with $L_X > 10^{44.1}$ (red), compared with different models from Shen (2009, black lines). Space density from BOSS data at $L_X > 10^{45}$ erg s $^{-1}$ (Ross et al. 2013, cyan square), together with the models of Shen (2009, blue lines) in the same luminosity range.

semi-analytic models have shown however that disk instabilities alone may not be enough to account for the full distribution of AGN (e.g., Menci et al. 2014, Gatti et al. 2015, Gatti et al. submitted), and in fact direct observations suggest that mergers may be at work in the most luminous sources (e.g., Treister et al. 2012).

To provide some additional clues towards these issues, we have carefully examined the F814W *HST/ACS* images for our X-ray sources (Scoville et al. 2007; Koekemoer et al. 2007). At $z > 3$, the F814W filter corresponds to a rest-frame wavelength of ≈ 2000 Å, where the galaxy emission is dominant and is therefore possible to identify SF and merger evidence.

We found that only a fraction of $< 20\%$ carry evident signs of dynamical perturbations, or have a close-by companion within $2''$. This does not imply that moderate mergers may not have happened in these systems, given that disk regrowth in gas-rich systems may be a viable possibility at these masses (e.g., Hopkins et al. 2009, Puech et al. 2014, Huertas-Company et al. 2015). Nevertheless our data challenge a pure merger-driven scenario, in agreement with the results of Cisternas et al. (2013), based on galaxy morphology in the local universe (41 galaxies within 35 Mpc).

4.6 Black hole mass estimate at $z > 3$

In this section we briefly summarize the results from the work of Trakhtenbrot et al. (in preparation) on a subsample of 10 *Chandra COSMOS-Legacy* sources at $z \sim 3.3$.

As we discussed in section 1.7.2, while phenomenological models of population synthesis have well constrained the AGN evolution at $z < 3$, so far the SMBH statistics at $z > 3$ is poor and the model predictions show significant disagreement. Moreover, all the phenomenological models have been so far able only to describe the AGN space density evolution with redshift, but without extensively discuss the physical causes of this evolution. Therefore, to improve our knowledge on the SMBH growth history it is necessary to properly characterize the basic physical properties of high- z accreting SMBHs, properties like black hole masses (M_{BH}), accretion rates (in terms of L/L_{Edd} or M_{BH}) and radiative efficiency (η).

So far, BH mass estimates have been focused on the most luminous AGN at a given redshift (e.g., Shemmer et al. 2004; Kurk et al. 2007; Netzer et al. 2007; Dietrich et al. 2009; Marziani et al. 2009; Willott et al. 2010; De Rosa et al. 2011; Trakhtenbrot et al. 2011). These works proved that black holes with $M_{BH} > 10^9 M_{\odot}$ exist up to $z \sim 7$, and their accretion rates are close to the Eddington limit only at $z \geq 5$. However, if we want to sample the most common SMBH with lower masses ($\sim 10^8 M_{\odot}$) and lower L_{bol} , X-ray datasets are a good place where to find these sources.

The sample used in this analysis includes 10 spectroscopically identified broad-line AGN *Chandra COSMOS-Legacy* in the redshift range $z=[3-3.7]$. The redshift range is selected in order to observe in the K -band the $H\beta$ emission line and the continuum adjacent to it, from which is possible to derive an estimate of the continuum luminosity.

We obtained K -band Keck/MOSFIRE² spectra for all the 10 sources: each source was observed for a total time between ~ 2500 and ~ 14400 seconds. The observation were conducted in six different nights between January 2014 and February 2015 and the observational conditions during the two nights were generally good, with typical seeing of $\sim 1''$ (or $\sim 0.8''$ in the NIR). The spectra were modeled using the procedure from Trakhtenbrot & Netzer (2012): the model components are a linear (pseudo) continuum, a broadened Fe II template and a combination of different Gaussians, to take in account different broad and narrow emission lines. These lines are He II, $H\beta$, [O III] $\lambda 4959$ and [O III] $\lambda 5007$. The continuum flux at 5100 \AA is estimated from the best-fit linear continuum and then used to measure the monochromatic continuum luminosity at 5100 \AA (rest-frame).

The black hole mass was then estimated from the $H\beta$ equivalent width, using the equation

$$M_{BH}(H\beta) = 1.05 \times 10^8 \left(\frac{L_{5100}}{10^{46} \text{ erg s}^{-1}} \right)^{0.65} \left[\frac{FWHM(H\beta)}{10^3 \text{ km s}^{-1}} \right]^2 M_{\odot}. \quad (4.9)$$

(see section 1.2 for further details). In Figure 4.15 we show four of the spectra, together with the model fitted to estimate M_{BH} . Two of the ten sources did not

²MOSFIRE is a NIR multi-object spectrograph in operation at the Cassegrain focus of the Keck I telescope. It has a $6.1' \times 6.1'$ field of view, and it can have up to 46 slits, which can be reconfigured in less than 5 minutes, using a unique cryogenic robotic slit mask system.

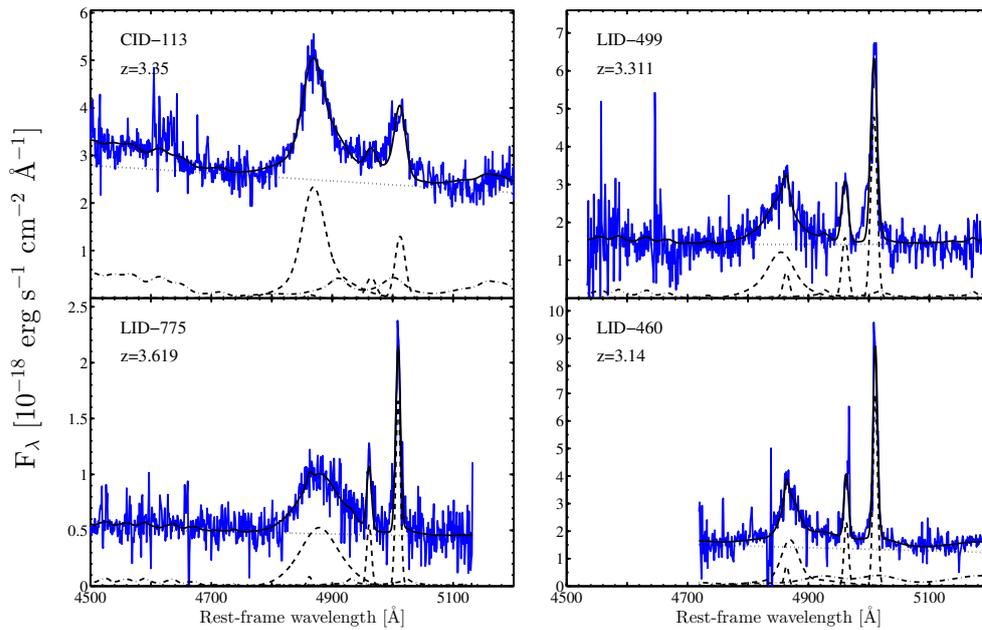


Figure 4.15: Spectra for four X-ray selected COSMOS AGN at $z \sim 3.3$ (blue), along with the best-fitting spectral model (solid black lines). The different model components, i.e., a linear continuum (dotted), a broadened Iron template and a combination of broad and narrow Gaussians (dot-dashed), are also shown.

show evidence of broad $H\beta$ emission: the estimated upper limit on the line equivalent width is $EW(H\beta) < 10\text{--}15 \text{ \AA}$, lower by more than a factor of 4 than the median value of $EW(H\beta)$ of the remaining 8 sources in the sample. The average black hole mass for the eight sources is $M_{BH} \sim 3 \times 10^8 M_{\odot}$, and the black hole mass range is $\text{Log}(M_{BH}) = [8.32\text{--}9.88] M_{\odot}$.

The bolometric luminosity was estimated from the optical continuum, using the luminosity-dependent prescription from Trakhtenbrot & Netzer (2012). At L_{5100} , the correction term can be expressed with the equation

$$f_{bol}(5100\text{\AA}) = 6.57 - 0.88L_{5100,45} + 0.26L_{5100,45}^2, \quad (4.10)$$

where $L_{5100,45} \equiv \log(L_{5100}/10^{45} \text{ erg s}^{-1})$. The bolometric luminosities obtained with this method are in the range $L_{bol} = [5.8 \times 10^{45} \text{--} 2.6 \times 10^{46}] \text{ erg s}^{-1}$. A consistency check was also performed, deriving the bolometric luminosities from the 2–10 keV *Chandra* luminosities, using the prescriptions of Marconi et al. (2004). The agreement between the two techniques is generally good, with a median offset of about 0.07 dex.

Finally, the accretion rates in our sample range around $0.1 \leq L/L_{Edd} \leq 0.5$.

In Figure 4.16 we show the distribution of apparent K magnitude, L_{bol} , M_{BH} and L/L_{Edd} for the ten sources in the *Chandra COSMOS-Legacy* sample (red circles), for a parent sample of *Chandra COSMOS-Legacy* sources at $z \sim 2.5$ (red squares) and for a sample of optically selected unobscured AGN from different surveys (open markers). These optical results have been presented in Shemmer et al. (2004) and

Netzer et al. (2007, triangles at $z \sim 2.4$ and ~ 3.3), in Trakhtenbrot et al. (2011, squares at $z \sim 4.8$) and in Kurk et al. (2007) and Willott et al. (2010, diamonds at $z \sim 6.2$).

As can be seen, the X-ray analysis allows us to observe objects with lower bolometric luminosities and with BH masses smaller than those known so far; the accretion rates are instead consistent with those observed in optically selected AGN. The masses and accretion rates observed in our sample imply that most of these sources had to grow from massive BH seeds ($M_{seed} > 10^4 M_{\odot}$). Lower BH seeds, and particularly stellar seeds, can only be possible if the accretion rate L/L_{Edd} was significantly higher at higher redshifts. A similar requirement, however, would imply duty cycles of 100%, i.e. continuous accretion at $L/L_{Edd} \sim 1$ along the whole BH life ($t \leq 10^9$ yrs), which is extremely unlikely.

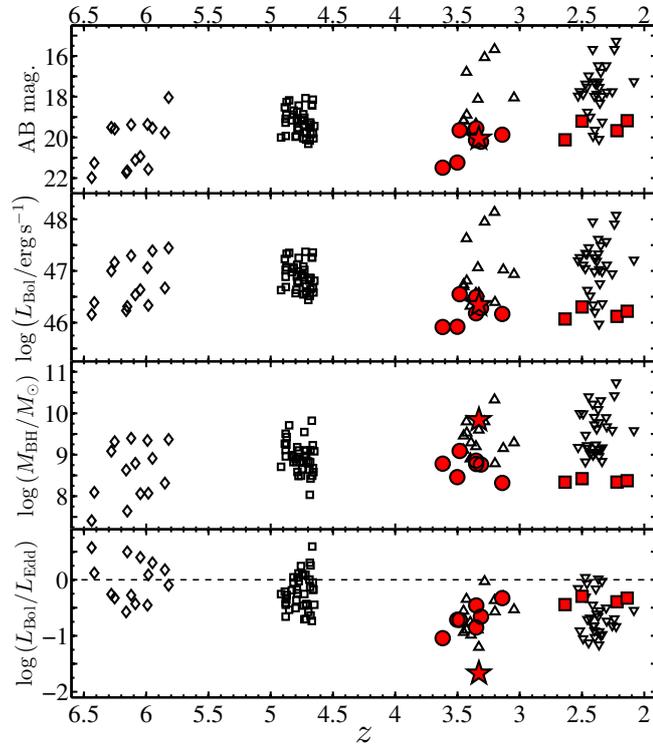


Figure 4.16: From top to bottom, trends of observed (NIR) brightness, L_{bol} , M_{BH} and L/L_{Edd} for the available samples of unobscured AGN at $z > 2$, with reliable determinations of M_{BH} . The red symbols represent the measurements reported in this work, at $z \sim 3.3$ and 2.5. CID_947, which was analyzed in detail in Trakhtenbrot et al. (2015), is highlighted as a pentagram. Estimates for optically selected sources are plotted in black and come from different surveys: the combined sample of Shemmer et al. (2004) and Netzer et al. (2007, triangles at $z \sim 2.4$ and ~ 3.3); Trakhtenbrot et al. (2011, squares at $z \sim 4.8$); and the combined samples of Kurk et al. (2007) and Willott et al. (2010, diamonds at $z \sim 6.2$). The Eddington limit, i.e., $L/L_{Edd} = 1$, is marked with a dashed line in the bottom panel.

Finally, it is worth noticing that while on average X-ray selected AGN allow us to study the properties of less extremely accreting objects than the optically selected ones, it is still possible to find rare, extreme X-ray selected sources. This is the case of CID_947 ($z=3.328$, red star in Figure 4.15), one of the ten sources in the subsample described above. CID_947, as discussed in Trakhtenbrot et al. (2015) and shown in Figure 4.17, has $M_{BH} \sim 7.6 \times 10^9 M_{\odot}$ and a BH-to-stellar mass ratio M_{BH}/M_* , significantly higher than the typical value in local, inactive galaxies (at most, $M_{BH}/M_* \sim 1/500$; dotted line) and a factor of ~ 2.5 higher than expected at these redshifts. This mass and the accretion rate of CID_947 suggest that this source has to be in the final phase of its BH growth, having already a mass comparable to the most massive sources in the local Universe (Figure 1.1). On the contrary, the galaxy is just starting to form stars and increase its mass. Therefore, we believe the BH and galaxy growth in this source are un-synchronous, with the BH dominating the process and the galaxy adjusting to it later on (see Volonteri 2012).

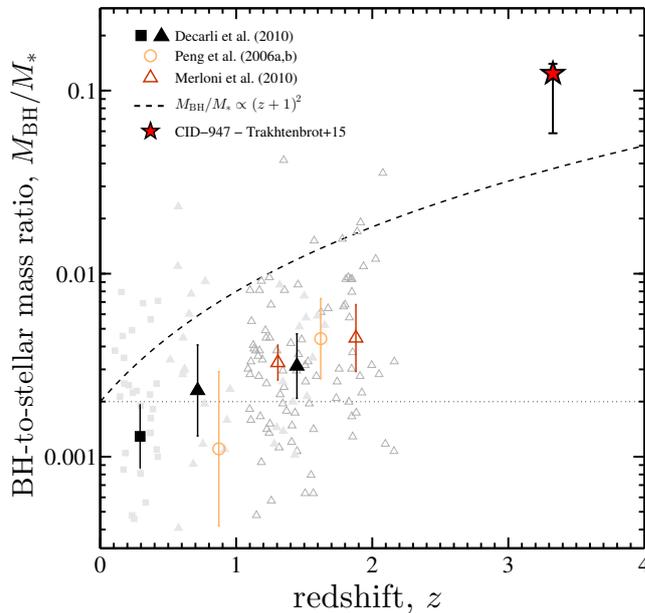


Figure 4.17: Evolution with redshift of the BH-to-stellar mass ratio, M_{BH}/M_* : at $z > 2$ a extrapolation on the basis of the actual data is shown. CID_947 (red star) has $M_{BH}/M_* = 1/8$ at $z \sim 3.3$, ~ 50 higher than is observed in typical local SMBH-galaxy inactive systems. Taken from Trakhtenbrot et al. (2015).

4.7 Discussion and conclusions

In this work we have selected a sample of 174 $z \geq 3$ sources from the *Chandra COSMOS-Legacy* survey, the largest sample of $z \geq 3$ X-ray selected sources on a contiguous field. Here we summarize the main results we obtained.

- 78 of the 174 have a spectroscopic redshift, while the remaining 96 have a photometric redshift. We treated our photo- z as a probability weighted sum

of contributions, using only the contribution of the Pd z at $z \geq 3$. The sum of all these contributions is equivalent to have 74.3 sources with $z \geq 3$. In a similar way, we also selected an additional sample of other 309 sources with $z_{phot} < 3$ but with significant Pd z contribution at $z \geq 3$ (with a total contribution equal to 39.7 sources with $z \geq 3$), therefore obtaining a sample equivalent to 192.0 sources.

- 78 sources are unobscured Type 1 (45% of the whole sample, 43 sources with spectral type, the remaining 35 with only photometric type), while the remaining 96 are obscured Type 2 (55% of the whole sample, 32 sources with spectral type, the remaining 64 with photometric type). The mean (median) i -band magnitude is 23.3 (23.2) for Type 1 AGN and 25.1 (25.2) for Type 2 AGN.
- In our sample and in this range of redshifts a pure X-ray classification of obscured and unobscured source, using the HR to have a rough estimate of the column density of the source, is not reliable. HR vs redshift curves for $N_H=10^{21}$ and $N_H=10^{22}$ are in fact almost completely degenerate. This is due to the fact that at $z > 3$ the observed 0.5-2 keV band roughly corresponds to the rest-frame 2-10 keV band, while the rest-frame 0.5-2 keV is redshifted to energies lower than our energy range.
- Our analysis of the number counts in the observed 0.5-2 keV band shows a decline in the number of sources at $z > 3$ (Figure 4.8, left), in agreement with the predictions of both a LDDE model with an exponential decline and a LADE model. A different behaviour is observed at $z > 4$ (Figure 4.8, right) where our results remain in agreement with the LDDE model, while they are a factor of two below the predictions of the LADE model, especially at fluxes fainter than $5 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$.
- For the first time, we were able to put constraints on the number counts at $z > 5$ (Figure 4.9, left) and $z > 6$ (right).
- We compared our rest frame 2-10 keV comoving space density in the high-luminosity range of our survey ($L_X > 10^{44.1} \text{ erg s}^{-1}$, Figure 4.11, left) with different phenomenological models: the FDPL model overpredicts our data by a factor 2-4 at $3 < z < 5$, while our data are a factor of 2-2.5 below the predictions of the LDDE with an exponential decline model in the redshift range $z=[3-4]$, and they are in agreement at higher redshifts, $z \geq 5$. Our data suggest a less strong decline in the AGN number at $z > 4$ than in the models predictions.
- Thanks to our good statistics, we investigated the 2-10 keV space density for Type 1 (or unobscured) and Type 2 (or obscured) AGN (Figure 4.12). We found that obscured sources have a slope significantly flatter ($b=-0.34 \pm 0.04$) than unobscured sources ($b=-0.65 \pm 0.07$). The ratio between obscured and unobscured sources is $\simeq 1$ in the redshift range $z=[3-4]$, while it grows to $\simeq 2$ at $z=5$.

- We compared our data with the quasar activation merger models of Shen (2009), calibrated mostly on bright AGN at $z > 3$. We found that the Shen (2009) reference model in fact well matches the data at high luminosities, but it tends to overpredict our newest number counts at faint luminosities by a factor of 3-10. To decrease the number of faint AGN, we attempted to modify the basic model in two distinct ways, by either shortening the input light curve characterizing each quasar, or by imposing that most of $z > 3$ AGN are preferentially hosted in more massive haloes. The former solution is disfavoured as it provides a drop also at the bright end. The latter solution, better in agreement with the data, predicts instead a specific clustering pattern that we plan to test in future work within our COSMOS data. We have also briefly discussed and tested alternatives to merger models, showing that only a minority (<20%) of our faint AGN show some sign of an ongoing interaction.
- A subsample of $\simeq 15$ sources discussed in this work has already been observed with *Keck* MOSFIRE, allowing to estimate the BH mass and put better constraints on the accretion properties of SMBH in early universe. One of these sources, cid.947 ($z=3.328$) showed an extremely massive accreting BH, with $M_{BH} \simeq 0.1 M_{galaxy}$, therefore suggesting that the BH mass accretion process took place on a significantly shorter time-scale than the host galaxy growth process (Trakhtenbrot et al. 2015).

All the results presented in this work have to be verified with future works and larger samples, and possibly with a larger number of spectroscopic redshifts (our sample contains only two sources with spec- z at $z > 5$, and none at $z > 6$). A spectroscopic follow-up of two candidate $z > 6$ low-luminosity AGN in the *Chandra COSMOS Legacy Survey* (CID.2550 at $z=6.84$ and LID.2595 at $z=6.46$) will be performed in early 2016 (P.I.: F. Civano) using *Keck* LRIS. Given that there no known $z > 6$ spectroscopically confirmed X-ray selected AGN known, this would be a real breakthrough. The confirmation of these redshifts would allow to constrain for the first time the AGN number counts and the faint end of the AGN luminosity function at $z > 6$. We will provide even tighter constraints on available models of galaxy and BH formation during the epoch of reionization. In Figure 4.18 we show the $5'' \times 5''$ cutouts of the two sources in ten different bands: B Subaru - V Subaru - i Subaru - HST/ACS F814W - z Subaru - J UltraVISTA - H UltraVISTA - K UltraVISTA - $3.6 \mu m$ IRAC. As can be seen, both sources show no clear signal up to the H UltraVISTA band, where a clear detection is visible for both objects. In Figure 4.19 we show the SED of cid.2550: once again, a dropout in flux is clearly visible at $\lambda < 1 \mu m$, the Lyman break observed wavelength.

In the next 10-20 years, missions like *Athena* (Nandra et al. 2013) and *X-ray Surveyor* (Vikhlinin et al. 2012) will be able to detect SMBH activity in a significant fraction of high-redshift galaxies, providing samples of ~ 100 sources at $z > 6$ and possibly even reaching luminosities of the orders of $10^{42.5}$ erg s $^{-1}$ (Aird et al. 2013).

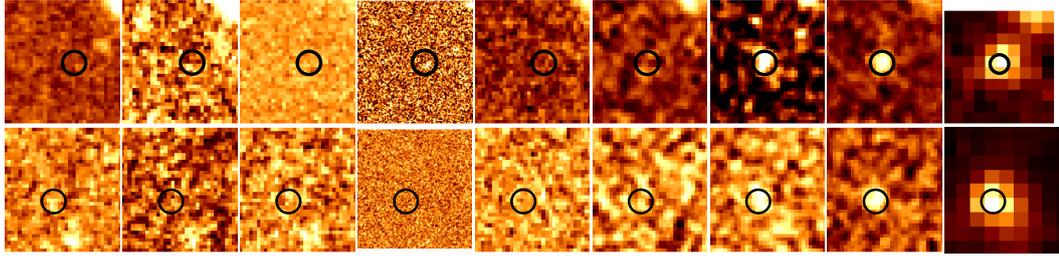


Figure 4.18: The cutouts ($5'' \times 5''$) of the two (CID-2550 at $z=6.84$ and LID-2595 at $z=6.46$) candidate $z > 6$ sources (B Subaru - V Subaru - i Subaru - HST/ACS F814W - z Subaru - J UltraVISTA - H UltraVISTA - K UltraVISTA - $3.6 \mu\text{m}$ IRAC).

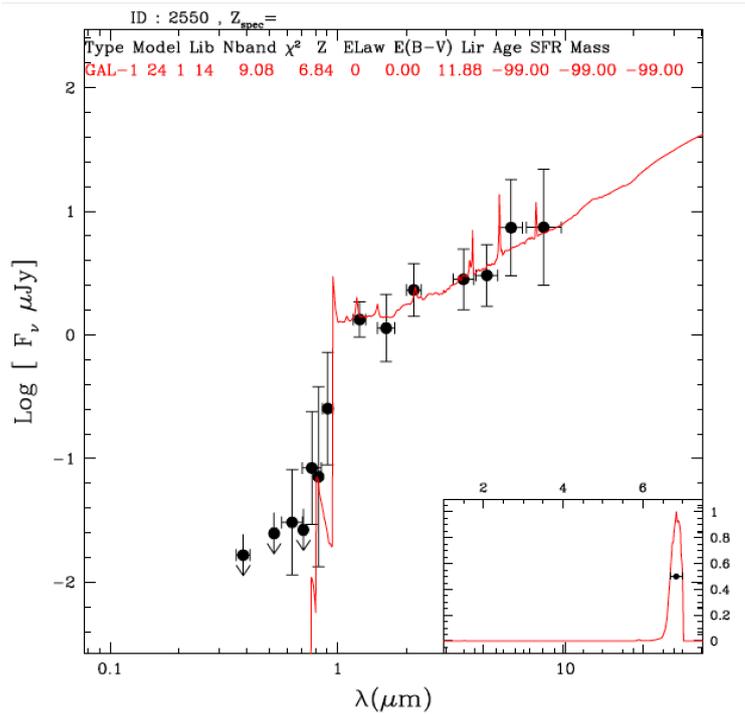


Figure 4.19: Spectral energy distribution of source CID-2550. The width of the probability distribution function (shown in the inset) is $\Delta z=0.35$ (LID-2595 has a width of 0.3) which is easily covered by the LRIS-R wavelength range.

Chapter 5

Discussion and conclusions

In this thesis, I have presented the *Chandra COSMOS-Legacy* survey (Civano et al. submitted; Marchesi et al. submitted), a 2.2 deg² *Chandra* survey of the COSMOS field. A total of 4.6 Ms of exposure time was employed: 1.8 Ms were already been granted over the central 0.9 deg² (C-COSMOS, Elvis et al. 2009; Puccetti et al. 2009; Civano et al. 2012), while other 2.8 Ms have been obtained as an X-ray Visionary Project during *Chandra* Cycle 14. X-ray surveys are strategic to study AGN with intrinsic lower luminosity and with higher obscuration than optical surveys.

In chapter 2 I described the procedure of data detection and analysis on these new 2.8 Ms observations, and I presented a catalog of point-like X-ray sources, obtained combined the new detected sources with those in C-COSMOS. The catalog contains 4016 sources, detected down to a flux limit $f \sim 2 \times 10^{-16}$ erg s⁻¹ cm⁻² in the 0.5-2 keV band, three times deeper than XMM-COSMOS (Cappelluti et al. 2009; Brusa et al. 2010).

Thanks to the extended multiwavelength dataset available in the COSMOS field, we were able to associate an optical/IR counterpart to $\sim 97\%$ of the X-ray sources (see chapter 3). We provided a spectroscopic redshift for $\sim 51\%$ of the sources in our sample, while $\sim 96\%$ of the sources have a photometric redshift; only 138 out of 4016 sources have no redshift information.

Several models predict that the main BH growth phase takes place in optically obscured, gas-rich environments. X-ray emission is significantly less effected by obscuration than optical emission, therefore it allows us to study the active BH properties during the peak of AGN activity. Simultaneously, the host galaxy properties can be studied in the optical with no/limited contamination from the active BH. In *Chandra COSMOS-Legacy* $\sim 65\%$ of the sources are obscured, on the basis either of their optical spectrum or of their best-fitting SED template, and therefore allow us to study the BH growth over a large range of redshifts and luminosities, with unprecedented good statistics.

The *Chandra COSMOS-Legacy* 0.5–2 keV luminosity distribution peaks at $L_X \simeq 10^{43}$ erg s⁻¹ (Figure 3.11); as such, it represents an excellent bridge between different classes of surveys. At lower luminosities there are deep pencil-beam surveys like CDF-S (Xue et al. 2011), which are very effective in detecting both AGN and non-AGN, but cover small areas of the sky and can be plagued by cosmic variance. Besides, they detect a limited number of sources. At higher luminosities, instead,

large area surveys like Stripe 82 (LaMassa et al. 2013; La Massa et al. submitted) are particularly effective in finding rare objects, i.e., high-luminosity and/or high-redshift AGN. These surveys, however, are not effective in detecting low-luminosity AGN, because of their lack of depth.

Chandra COSMOS-Legacy covers with an excellent statistics (2285 sources in the soft band) the range of redshift $1 \leq z \leq 3$, i.e. at the peak of the AGN activity (e.g., Hasinger et al. 2005). *Chandra COSMOS-Legacy* also samples with solid statistics the luminosity range below the knee of the luminosity function, up to redshift $z \simeq 4$ (Figure 4.10, right panel), which allows studies of the low-luminosity, potentially high-obscured sources that dominate the AGN population and that are largely missed by optical surveys.

By combining the X-ray and optical information, we were able to confirm the existence of a correlation between X/O and the luminosity in the 2-10 keV band for Type 2 sources, and we also confirmed that this correlation becomes weaker at faint optical magnitudes ($i > 25$), i.e, for sources at higher redshift (average redshift $z = 2.12 \pm 0.61$, compared to $z = 1.36 \pm 0.72$ of the whole sample) and with higher specific star formation rate. Overall, our analysis indicates that X-ray data are capable of detecting AGN emission in optically faint galaxies. These AGN would not been recognized as such in optical surveys. We also extended to low luminosities the known inverse correlation between the fraction of obscured AGN and the hard band luminosity. The fraction of optically classified obscured AGN is of the order of 90% at $L_X < 10^{42}$ erg s $^{-1}$ and drops to $\simeq 40\%$ at $L_X > 10^{43.5}$ erg s $^{-1}$. This behavior is fairly consistent with the Treister et al. (2009) AGN synthesis model predictions. A different trend is instead observed while assuming an X-ray obscuration classification criterion, based on the hardness ratio (HR). In this case there is no clear evidence of correlation between obscured fraction and luminosity: this result is in agreement with the predictions of the models by Gilli et al (2007) and Miyaji et al. (2015). While this result suggests that a fraction of sources optically classified as obscured are actually objects where the galaxy contribution dominates at optical wavelengths, a complete analysis of this discrepancy will require a higher spectroscopic completeness, and the X-ray spectral analysis for a significant fraction of the sample.

The study of the high-redshift Universe is strategic to understand how the first BHs and their host galaxy grow, and which kind of BH seeds were originally formed. X-ray selected AGN sample a population of BHs with on average smaller mass and lower luminosity than optically selected AGN, and can therefore explore a more standard BH-host galaxy population. We selected the sample of 174 $z \geq 3$ sources from the *Chandra COSMOS-Legacy* survey (chapter 4): 78 of the 174 have a spectroscopic redshift, while the remaining 96 have a photometric redshift. In our analysis we treated the photo- z as a probability weighted sum of contributions, using only the contribution of the redshift probability distribution function (Pdz) at $z \geq 3$. Summing these contributions, we obtain a final sample equivalent to 192.0 sources. This sample is one of the largest samples of $z \geq 3$ X-ray selected sources, and the largest sample on a contiguous field. This is also the sample with the largest statistics at $z > 5$ (9 sources) and at $z > 6$ (4 sources, although none of which have a spec- z). Our analysis of the number counts in the observed 0.5-2 keV band shows a decline in the number of sources at $z > 3$ (Figure 4.8, left): we compared our

results with different AGN phenomenological models and found an agreement with the predictions of both a Luminosity-Dependent Density Evolution (LDDE; Gilli et al. 2007) model with an exponential decline and with those of a Luminosity And Density Evolution (LADE; Aird et al. 2010) model. At $z > 4$ (Figure 4.8, right), instead, only the predictions of the LDDE model remain in agreement with our data, while those of the LADE model are a factor of two above our data.

We compared our rest-frame 2-10 keV comoving space density in the high-luminosity range of our survey ($L_X > 10^{44.1}$ erg s $^{-1}$, Figure 4.11, left) with different phenomenological models: the Flexible Double Power Law (FDPL) model, a derivation of the LADE model, overpredicts with respect to our data by a factor 2-4 at $3 < z < 5$, while our data are a factor of 2-2.5 below the predictions of the LDDE with an exponential decline model in the redshift range $z = [3-4]$, and they are in agreement at higher redshifts ($z \geq 5$). Our data suggest a less strong decline in the AGN number at $z > 4$ than in the models predictions.

In order to investigate which kind of processes are mainly responsible in driving the black hole accretion, if galaxy major merger or secular accretion, we compared our data also with a physical model of quasar activation through major merger (Shen 2009); this model was calibrated mostly on luminous optically selected AGN at $z > 3$. In the *Chandra COSMOS-Legacy* luminosity range, i.e., ~ 2 orders of magnitude below the interval of luminosity used for the Shen (2009) model calibration, we found that the model predictions are above our data by a factor 3–10. A possible way to overcome this discrepancy is to assume that most of $z > 3$ AGN are preferentially hosted in more massive haloes than suggested by the Shen (2009) model. This modified model also predicts a specific clustering pattern that we plan to test in a future work using COSMOS data. The analysis of the AGN space density at high redshift leads us to the evidence that in our luminosity range the secular accretion scenario is likely preferred to the major-merger one.

To probe BH accretion at high redshift in a different BH mass–luminosity than the one sampled by optically selected AGN, a subsample of $\simeq 10$ sources at $z = [3.2-3.7]$ has already been followed-up with *Keck* MOSFIRE. In this range of redshift, X-ray selected black holes have, on average, masses and bolometric luminosities higher by $\gtrsim 0.5$ dex than those of optically selected AGN (Figure 4.16), therefore it is possible that these objects are also the byproduct of a different class of black hole seeds than those optically selected, an hypothesis that should be tested with theoretical models and cosmological simulations. It is also worth noticing that one of these sources, cid_947, has $M_{BH} \simeq 0.1 M_{galaxy}$, significantly higher than observed in the local Universe, and a factor of ~ 2.5 higher than expected at $z \sim 3.5$, suggesting that in this system the BH mass accretion process took place well before the host-galaxy growth (Trakhtenbrot et al. 2015).

The results at high redshift have to be validated with future works and larger samples, including a larger fraction of sources with spectroscopic identification, given that photometric redshifts at these redshifts can often have large probability distribution functions and therefore give an intrinsically less reliable estimate of the redshift than a spec- z . So far, our sample contains only two sources with spec- z at $z > 5$, and none at $z > 6$. A first step in this direction will be the spectroscopic follow-up of two candidate $z > 6$ low-luminosity AGN in the *Chandra COSMOS Legacy Survey* that will be performed in early 2016 (P.I.: F. Civano) using *Keck* LRIS.

These objects would be the first $z > 6$ spectroscopically confirmed X-ray selected AGN, and would allow us to constrain for the first time the AGN number counts and the faint end of the AGN luminosity function at $z > 6$.

The *Chandra COSMOS-Legacy* dataset, combined with the other multiwavelength COSMOS catalogs, can be used to answer questions related to a large number of astrophysical topics: for example, (i) the study of the host galaxy-SMBH system and its physical properties (e.g. sSFR, M_{BH}) can be performed through X-ray spectral analysis (e.g., Lanzuisi et al. 2013, 2014). To do so, X-ray spectra of optically obscured and unobscured sources can be studied in different range of redshift to estimate the average spectral parameters (N_H and spectral slope) and intrinsic X-ray luminosity. While the largest majority of *Chandra COSMOS-Legacy* sources have less than 50 counts and a proper spectral analysis can be performed only on the $\simeq 800$ brightest *Chandra COSMOS-Legacy* sources, the combination of the *Chandra COSMOS-Legacy* and *XMM-COSMOS* datasets (Cappelluti et al. 2009; Brusa et al. 2010) allows us to study average X-ray properties of sources in different bins of redshift and luminosity, stacking the signal from multiple objects. Moreover, more than 80% of the COSMOS field has also been covered with *NuSTAR* (Civano et al. 2015), thus providing the possibility to study the spectral shape at energies up to 30 keV. (ii) Combining the deep *Chandra COSMOS-Legacy* X-ray coverage with the large multiwavelength dataset on the COSMOS field will allow us to measure with excellent statistics the average properties of X-ray undetected sources using stacking analysis. A preliminary analysis on the 1.8 Ms C-COSMOS dataset, using optically selected galaxies with reliable photometric redshift and no X-ray detection from Ilbert et al. (2009; $\simeq 21500$ sources at $z > 2$, $\simeq 4500$ at $z > 3$), shows evidence of a 3σ signal up to redshift $z \simeq 3$, with average luminosity $L_X > 2 \times 10^{41}$ erg s^{-1} . With the whole *Chandra COSMOS-Legacy* dataset, and the possibility to use the Ultravista *K*-selected photo- z (Laigle et al. in preparation), it will be possible to repeat this analysis using a sample $\simeq 3$ times larger, thus possibly improving the significance of our detections by a factor $\simeq 1.5-2$, and finding significant signal even at $z > 3$. This analysis will allow us to put better constraints on the average properties of the low mass and low luminosity AGN in *COSMOS-Legacy*.

We showed that most X-ray surveys lie on a locus (yellow shaded area in Figure 1.8) in a area-flux diagram determined by our current X-ray telescope capabilities. *COSMOS-Legacy* is the first survey to significantly move away from this locus, getting an additional factor 2-3 deeper at the areas it covers, by using a total exposure time which is unusually large (4.6 Ms total) for that given area flux combination and laying the groundwork for future facilities. In future decades, facilities like *Athena* (Nandra et al. 2013) and *X-ray Surveyor* (Vikhlinin et al. 2012) will be able to explore a new region of area-flux parameter space, moving towards the bottom right corner of Figure 1.8, i.e., simultaneously at larger areas and deeper flux limits than the current surveys. For example, *Athena* large effective area and field of view will enable X-ray surveys to be carried out two orders of magnitude faster than *XMM-Newton* and *Chandra*. With a *Chandra*-like resolution over $10'$ and with an effective area 30 times better than *Chandra* at 1 keV, *X-ray Surveyor* will be able to cover the same *COSMOS-Legacy* area at the same flux in only 55 ks, 80 times faster than *Chandra*.

Appendix A

List of new *Chandra* *COSMOS-Legacy* observations

In this Appendix we report the list of new *Chandra COSMOS-Legacy* observations. The observations took place in four blocks: November, 2012 to January, 2013; March to July, 2013; October, 2013 to January, 2014; and March, 2014. Due to observational constraints, 11 of the 56 ACIS-I pointings were scheduled as two or more separate observations, for a total of 68 pointings.

The mean net effective exposure time per field was 48.8 ks: the maximum exposure was 53 ks (observation 15227) while the minimum exposure was 45.2 ks (combined observations 15208 and 15998).

The sequence of the observations was designed to start from the N-E top corner tile of C-COSMOS moving towards W and proceeding clockwise around the central C-COSMOS area, in such a way that the outer frame of the C-COSMOS survey overlaps with the inner frame of the new *Chandra* observations. To maximize the overlap, the observing roll angle was constrained to be within 70 ± 20 or 250 ± 20 degrees.

Field	Obs. ID	RA	Dec.	Date	Exp. time second	Roll (deg)
CLS_1	15207	150.544451	2.499045	2012-11-25	14883	70.2
	15590	150.544402	2.499094	2012-11-23	14893	70.2
	15591	150.544454	2.499065	2012-11-25	19828	70.2
CLS_2	15208	150.415643	2.543225	2012-12-07	22985	70.2
	15598	150.415625	2.543213	2012-12-08	22193	70.2
CLS_3	15209	150.295749	2.588083	2012-12-03	23775	70.2
	15600	150.295747	2.588106	2012-12-05	21795	70.2
CLS_4	15604	150.164741	2.639752	2012-12-10	20988	70.2
	15210	150.164738	2.639709	2012-12-16	24365	70.2
CLS_5	15211	150.045569	2.682903	2012-12-13	23572	70.2
	15605	150.045586	2.682879	2012-12-15	21801	70.2
CLS_6	15212	149.913425	2.732850	2012-12-21	25249	70.2
	15606	149.913418	2.732845	2012-12-23	25219	70.2
CLS_7	15213	149.796052	2.772968	2013-01-01	49435	62.2
CLS_8	15214	149.751287	2.655331	2013-01-03	45983	61.75
CLS_9	15215	149.704144	2.525446	2013-01-07	49437	63.2
CLS_10	15216	149.654208	2.399733	2013-01-16	46459	56.7
CLS_11	15217	149.627509	2.272922	2013-03-23	46057	265.2
CLS_12	15218	149.584767	2.145874	2013-03-22	46475	265.2
CLS_13	15219	149.538688	2.017596	2013-03-30	49432	261.6
CLS_14	15220	149.614659	1.846399	2013-04-04	49924	60.1
CLS_15	15221	149.753949	1.801935	2013-04-10	49431	58.2
CLS_16	15222	149.870306	1.757718	2013-04-04	49407	60.0
CLS_17	15223	149.999623	1.706079	2013-04-17	50905	55.2
CLS_18	15224	150.115609	1.664373	2013-04-19	49426	55.2
CLS_19	15225	150.245495	1.621716	2013-04-05	49631	59.8
CLS_20	15226	150.411336	1.697830	2013-06-21	49428	250.2
CLS_21	15227	150.463753	1.829216	2013-05-02	53051	50.2
CLS_22	15228	150.504029	1.950647	2013-04-30	49432	50.2
CLS_23	15229	150.551660	2.080265	2013-05-10	49012	52.2
CLS_24	15230	150.592692	2.199969	2013-05-08	49429	52.2
CLS_25	15231	150.642972	2.325853	2013-05-13	48446	51.0
CLS_26	15232	150.690403	2.449284	2013-05-16	35085	50.6
	15649	150.690409	2.449315	2013-06-03	15251	50.65
CLS_27	15233	150.734924	2.575875	2013-05-21	46476	50.20
CLS_28	15234	150.616710	2.623373	2013-05-22	5439	50.20
	15653	150.594589	2.629150	2014-01-16	44895	58.21
CLS_29	15235	150.480833	2.671101	2013-06-01	49440	48.31
CLS_30	15236	150.364822	2.714260	2013-06-01	49435	48.23
CLS_31	15237	150.228563	2.765929	2013-06-08	25246	50.65
	15655	150.228550	2.765907	2013-06-10	24466	50.65
CLS_32	15238	150.114727	2.808579	2013-06-09	49429	50.65
CLS_33	15239	149.981160	2.858739	2013-06-11	49430	50.20
CLS_34	15240	149.617459	2.695912	2013-10-15	48450	77.09
CLS_35	15241	149.593992	2.566795	2014-03-28	48600	260.21
CLS_36	15242	149.547344	2.443553	2013-06-22	49432	50.20
CLS_37	15243	149.499113	2.312060	2013-07-05	47985	50.20
CLS_38	15244	149.547442	1.723759	2014-01-21	47461	53.21
CLS_39	15245	149.680897	1.673516	2014-01-23	49437	53.21
CLS_40	15246	149.796705	1.629086	2013-10-22	48850	75.21
CLS_41	15247	149.953115	1.578784	2014-03-18	49545	267.21
CLS_42	15248	150.047419	1.537353	2013-11-13	49438	71.61
CLS_43	15249	150.516513	1.655025	2013-11-29	45635	70.21
CLS_44	15250	150.566083	1.783479	2013-12-12	49315	70.21
CLS_45	15251	150.612991	1.904134	2013-12-03	29702	67.91
	16544	150.613008	1.904121	2013-12-04	19830	67.91
CLS_46	15252	150.660018	2.034094	2013-12-14	49434	70.21
CLS_47	15253	150.707972	2.162869	2014-01-28	49132	53.21
CLS_48	15254	150.753963	2.289683	2014-01-29	49139	53.21
CLS_49	15255	150.661405	2.741395	2014-03-24	49435	260.21
CLS_50	15256	150.504801	2.795740	2014-01-13	49943	59.21
CLS_51	15257	150.384246	2.838987	2014-01-04	49435	61.85
CLS_52	15258	149.497504	2.746858	2014-01-01	49432	62.27
CLS_53	15259	149.451159	2.620733	2014-01-27	49435	53.21
CLS_54	15260	150.690908	1.740589	2014-01-05	22793	60.21
	16562	150.690921	1.740576	2014-01-25	26736	60.21
CLS_55	15261	150.736977	1.863191	2014-01-18	46474	59.21
CLS_56	15262	150.782957	1.992193	2014-01-12	50236	59.21

Table A.1: Observation Summary. CLS stands for COSMOS Legacy Survey.

Appendix B

The *Chandra* COSMOS X-ray catalog

We report in Table B.1 the columns of the *Chandra COSMOS-Legacy* X-ray catalog. The catalog will be available with the published Paper (Civano et al. submitted), in the COSMOS repository and online.

No.	Field	Note
1	Name	<i>Chandra</i> source name
2	R.A.	<i>Chandra</i> Right Ascension (J2000, hms)
3	DEC	<i>Chandra</i> Declination (J2000, dms)
4	pos.err	Positional error [arcsec]
5	DET_ML_F	maximum likelihood detection value in 0.5-7 keV band
6	rate_F	0.5-7 keV count rate [counts s ⁻¹]
7	rate_F_err	0.5-7 keV count rate error [counts s ⁻¹]
8	flux_F	0.5-10 keV flux [erg cm ⁻² s ⁻¹]
9	flux_F_err	0.5-10 keV flux error [erg cm ⁻² s ⁻¹]
10	snr_F	0.5-7 keV S/N Ratio
11	exptime_F	0.5-7 keV exposure time [ks]
12	cts_ap_F	0.5-7 aperture photometry counts [counts]
13	cts_ap_F_err	0.5-7 aperture photometry counts error [counts]
14	DET_ML_S	maximum likelihood detection value in 0.5-2 keV band
15	rate_S	0.5-2 keV count rate [counts s ⁻¹]
16	rate_S_err	0.5-2 keV count rate error [counts s ⁻¹]
17	flux_S	0.5-2 keV flux [erg cm ⁻² s ⁻¹]
18	flux_S_err	0.5-2 keV flux error [erg cm ⁻² s ⁻¹]
19	snr_S	0.5-2 keV S/N Ratio
20	exptime_S	0.5-2 keV exposure time [ks]
21	cts_ap_S	0.5-2 aperture photometry counts [counts]
22	cts_ap_S_err	0.5-2 aperture photometry counts error [counts]
23	DET_ML_H	maximum likelihood detection value in 2-7 keV band
24	rate_H	2-7 keV count rate [counts s ⁻¹]
25	rate_H_err	2-7 keV count rate error [counts s ⁻¹]
26	flux_H	2-10 keV flux [erg cm ⁻² s ⁻¹]
27	flux_H_err	2-10 keV flux error [erg cm ⁻² s ⁻¹]
28	snr_H	2-7 keV S/N Ratio
29	exptime_H	2-7 keV exposure time [ks]
30	cts_ap_H	2-7 aperture photometry counts [counts]
31	cts_ap_H_err	2-7 aperture photometry counts error [counts]
32	hr	Hardness ratio
33	hr_lo_lim	Hardness ratio 90% lower limit
34	hr_up_lim	Hardness ratio 90% upper limit

Table B.1: Data fields in the catalog.

Appendix C

The multiwavelength catalog of *Chandra* COSMOS Legacy sources

In this Appendix we describe the multiwavelength properties reported in the catalog of optical counterparts of the *Chandra* COSMOS-Legacy sources. The catalog will be available with the published Paper (Marchesi et al. submitted), in the COSMOS repository and online.

1. *Column 1.* Source ID. Sources are listed in the same order used in the catalog of X-ray sources reported in Appendix B: first all sources detected in full band, then those detected in soft band only, then those detected in hard band only.
2. *Columns 2-3.* X-ray coordinates of the source, from the catalog of X-ray sources reported in Appendix B.
3. *Columns 4-6.* Maximum likelihood detection (DET_ML) value in 0.5-7 keV, 0.5-2 and 2-7 keV band, from the catalog of X-ray sources reported in Appendix B.
4. *Columns 7-9.* X-ray fluxes in full, soft and hard bands, from the catalog of X-ray sources reported in Appendix B. Negative fluxes represent upper limits.
5. *Columns 10-12* Hardness ratio and hardness ratio 90% lower and upper limit, from the catalog of X-ray sources reported in Appendix B.
6. *Column 13.* Identifier number of the optical counterpart from the Ilbert et al. (2009) catalog.
7. *Columns 14-15.* Optical coordinates of the source, from the Ilbert et al. (2009) catalog.
8. *Columns 16-17.* *i*-band magnitude and magnitude error in 3'' aperture, from the Ilbert et al. (2009) catalog.
9. *Column 18.* *i*-band magnitude origin: 1 Subaru, 2 CFHT, 3 SDSS, 5 manual photometry

10. *Column 19.* Identifier number of the *K*-band counterpart from the UltraVISTA catalog from Laigle et al. (submitted).
11. *Columns 20-21.* UltraVISTA *K*-band counterpart coordinates, from the Laigle et al. (submitted) catalog.
12. *Columns 22-23.* UltraVISTA *K*-band magnitude and magnitude error in 3'' aperture, from the Laigle et al. (submitted) catalog.
13. *Column 24.* Identifier number of the *K*-band counterpart from the CFHT catalog from Ilbert et al. (2009) catalog.
14. *Columns 25-26.* CFHT *K*-band counterpart coordinates, from the Ilbert et al. (2009) catalog.
15. *Columns 27-28.* CFHT *K*-band magnitude and magnitude error in 3'' aperture, from the Ilbert et al. (2009) catalog.
16. *Column 29-30.* Coordinates of the 3.6 μm counterpart from the Sanders catalog.
17. *Column 31-32.* 3.6 μm flux (μJy) and flux error in 1.9'' aperture, from the Sanders catalog. To convert to total flux, the standard factor suggested in the IRAC user guide has to be applied (dividing by 0.765).
18. *Column 33-34.* Coordinates of the 3.6 μm counterpart from the SPLASH catalog.
19. *Column 35-36.* 3.6 μm flux (μJy) and flux error in 1.9'' aperture, from the SPLASH catalog. To convert to total flux, the standard factor suggested in the IRAC user guide has to be applied (dividing by 0.765).
20. *Column 37.* Final identification flag: 1= secure, 10= ambiguous, 100= sub-threshold, -99= unidentified
21. *Column 38.* Star flag: 1= spectroscopically confirmed star, 10= photometric star, 100= visually identified star.
22. *Column 39.* Best redshift available. This is the spectroscopic redshift if the spectroscopic redshift quality flag is $Q_g \geq 1.5$ (see below) and the photometric redshift otherwise.
23. *Column 40.* Spectroscopic redshift.
24. *Column 41.* Spectroscopic redshift origin.
25. *Column 42.* Spectroscopic redshift quality. 2= "secure" redshift, spectroscopic reliability $>99.5\%$, 1.5= "reliable" redshift, spectroscopic reliability $<99.5\%$ but there is a photometric redshift such that $\frac{\Delta z}{1+z_{spec}} < 0.1$, 1= "not reliable" redshift, spectroscopic reliability $<99.5\%$ and there is a photo-z such that $\frac{\Delta z}{1+z_{spec}} > 0.1$.
26. *Column 43.* Spectroscopic identification. 1=BLAGN, 2=non-BLAGN, 0=star.

27. *Column 44.* Photometric redshift from Salvato et al. (in preparation).
28. *Column 45.* Photometric identification from SED fitting (1=unobscured, 2=obscured, 3=galaxy, 5=star).
29. *Column 46.* Identifier number of the *XMM-COSMOS* counterpart, from the Cappelluti et al. (2009) catalog.
30. *Column 47.* Luminosity distance (in Mpc).
31. *Columns 48-50.* Rest-frame luminosity, in 0.5-10 keV, 0.5-2 keV and 2-10 keV bands, obtained assuming an X-ray spectral index $\Gamma=1.4$.
32. *Column 51.* Intrinsic neutral hydrogen (N_H) column density, estimated using the best redshift available and the hardness ratio from the catalog of X-ray sources reported in Appendix B, assuming an X-ray spectral index $\Gamma=1.8$.
33. *Columns 52-54.* Luminosity absorption correction, in 0.5-10 keV, 0.5-2 keV and 2-10 keV bands, obtained assuming the intrinsic N_H reported in *Column 47* and a power-law with spectral index $\Gamma=1.8$.
34. *Column 55.* Lower limit on intrinsic N_H column density, estimated using the best redshift available and the hardness ratio lower limit from the catalog of X-ray sources reported in Appendix B, assuming an X-ray spectral index $\Gamma=1.8$.
35. *Columns 56-58.* Luminosity absorption correction, in 0.5-10 keV, 0.5-2 keV and 2-10 keV bands, obtained assuming the intrinsic N_H reported in *Column 51* and a power-law with spectral index $\Gamma=1.8$.
36. *Column 59.* Upper limit on intrinsic N_H column density, estimated using the best redshift available and the hardness ratio upper limit from the catalog of X-ray sources reported in Appendix B, assuming an X-ray spectral index $\Gamma=1.8$.
37. *Columns 60-62.* Luminosity absorption correction, in 0.5-10 keV, 0.5-2 keV and 2-10 keV bands, obtained assuming the intrinsic N_H reported in *Column 55* and a power-law with spectral index $\Gamma=1.8$.

Bibliography

- Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, *MNRAS*, 451, 1892
- Aird, J., Nandra, K., Laird, E. S., et al. 2010, *MNRAS*, 401, 2531
- Aird, J., Comastri, A., Brusa, M., et al. 2013, arXiv:1306.2325
- Akylas, A., Georgantopoulos, I., Georgakakis, A., Kitsionas, S., & Hatziminaoglou, E. 2006, *A&A*, 459, 693
- Alexander, D. M., Brandt, W. N., Hornschemeier, A. E., et al. 2001, *AJ*, 122, 2156
- Alexander, D. M., Chary, R.-R., Pope, A., et al. 2008, *ApJ*, 687, 835
- Alexander, D. M., & Hickox, R. C. 2012, *New Astronomy Review*, 56, 93
- Allevato, V., Finoguenov, A., Cappelluti, N., et al. 2011, *ApJ*, 736, 99
- Allevato, V., Finoguenov, A., Civano, F., et al. 2014, *ApJ*, 796, 4
- Antonucci, R. 1993, *Annual Review of Astron and Astrophysics*, 31, 473
- Arnaud, K., Dorman, B., & Gordon, C. 1999, *Astrophysics Source Code Library*, 9910.005
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540
- Arnouts, S., Le Floch, E., Chevillard, J., et al. 2013, *A&A*, 558, A67
- Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, *ApJ*, 772, 26
- Avni, Y., & Bahcall, J. N. 1980, *ApJ*, 235, 694
- Ballantyne, D. R., Shi, Y., Rieke, G. H., et al. 2006, *ApJ*, 653, 1070
- Barger, A. J., & Cowie, L. L. 2005, *ApJ*, 635, 115
- Barger, A. J., Cowie, L. L., Mushotzky, R. F., et al. 2005, *AJ*, 129, 578
- Barnes, J. E., & Hernquist, L. E. 1991, *ApJL*, 370, L65
- Basu-Zych, A. R., Lehmer, B. D., Hornschemeier, A. E., et al. 2013, *ApJ*, 762, 45
- Batcheldor, D., Robinson, A., Axon, D. J., Perlman, E. S., & Merritt, D. 2010, *ApJL*, 717, L6

- Baum, S. A., Gallimore, J. F., O’Dea, C. P., et al. 2010, *ApJ*, 710, 289
- Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, *MNRAS*, 370, 289
- Bekenstein, J. D. 1973, *ApJ*, 183, 657
- Bellovary, J., Brooks, A., Volonteri, M., et al. 2013, *ApJ*, 779, 136
- Bennert, V. N., Auger, M. W., Treu, T., Woo, J.-H., & Malkan, M. A. 2011, *ApJ*, 742, 107
- Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, *MNRAS*, 362, 25
- Bevington, P. R., & Robinson, D. K. 1992, New York: McGraw-Hill, —c1992, 2nd ed.,
- Bianchi, S., Maiolino, R., & Risaliti, G. 2012, *Advances in Astronomy*, 2012, 782030
- Blecha, L., Civano, F., Elvis, M., & Loeb, A. 2013, *MNRAS*, 428, 1341
- Blecha, L., Cox, T. J., Loeb, A., & Hernquist, L. 2011, *MNRAS*, 412, 2154
- Bongiorno, A., Merloni, A., Brusa, M., et al. 2012, *MNRAS*, 427, 3103
- Bongiorno, A., Mignoli, M., Zamorani, G., et al. 2010, *A&A*, 510, A56
- Bongiorno, A., Zamorani, G., Gavignaud, I., et al. 2007, *A&A*, 472, 443
- Bonning, E. W., Shields, G. A., & Salviander, S. 2007, *ApJL*, 666, L13
- Bournaud, F., Jog, C. J., & Combes, F. 2005, *A&A*, 437, 69
- Bournaud, F., Dekel, A., Teyssier, R., et al. 2011, *ApJL*, 741, L33
- Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, *MNRAS*, 370, 645
- Brandt, W. N., & Alexander, D. M. 2015, *A&A Reviews*, 23, 1
- Brandt, W. N., & Hasinger, G. 2005, *Annual Review of Astron and Astrophysics*, 43, 827
- Brightman, M., Nandra, K., Salvato, M., et al. 2014, *MNRAS*, 443, 1999
- Bromm, V., & Loeb, A. 2003, *Nature*, 425, 812
- Brusa, M., Civano, F., Comastri, A., et al. 2010, *ApJ*, 716, 348
- Brusa, M., Comastri, A., Daddi, E., et al. 2005, *A&A*, 432, 69
- Brusa, M., Comastri, A., Gilli, R., et al. 2009, *ApJ*, 693, 8
- Brusa, M., Zamorani, G., Comastri, A., et al. 2007, *ApJS*, 172, 353
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, *ApJ*, 802, 89
- Burlon, D., Ajello, M., Greiner, J., et al. 2011, *ApJ*, 728, 58

- Capak, P., Aussel, H., Ajiki, M., et al. 2007, *ApJS*, 172, 99
- Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, *Nature*, 470, 233
- Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, *A&A*, 497, 635
- Cappelluti, N., Kashlinsky, A., Arendt, R. G., et al. 2013, *ApJ*, 769, 68
- Cash, W. 1979, *ApJ*, 228, 939
- Chen, C.-T. J., Hickox, R. C., Alberts, S., et al. 2015, *ApJ*, 802, 50
- Chiaberge, M., Capetti, A., & Celotti, A. 2000, *A&A*, 355, 873
- Cicone, C., Maiolino, R., Sturm, E., et al. 2014, *A&A*, 562, A21
- Cimatti, A., Daddi, E., & Renzini, A. 2006, *A&A*, 453, L29
- Cirasuolo, M., Shankar, F., Granato, G. L., De Zotti, G., & Danese, L. 2005, *ApJ*, 629, 816
- Cisternas, M., Gadotti, D. A., Knapen, J. H., et al. 2013, *ApJ*, 776, 50
- Civano, F., Brusa, M., Comastri, A., et al. 2011, *ApJ*, 741, 91
- Civano, F., Elvis, M., Brusa, M., et al. 2012, *ApJS*, 201, 30
- Civano, F., Elvis, M., Lanzuisi, G., et al. 2012, *ApJ*, 752, 49
- Civano, F., Elvis, M., Lanzuisi, G., et al. 2010, *ApJ*, 717, 209
- Civano, F., Fabbiano, G., Pellegrini, S., et al. 2014, *ApJ*, 790, 16
- Civano, F., Hickox, R. C., Puccetti, S., et al. 2015, *ApJ*, 808, 185
- Civano, F., Mignoli, M., Comastri, A., et al. 2007, *A&A*, 476, 1223
- Civano, F., Comastri, A., & Brusa, M. 2005, *MNRAS*, 358, 693
- Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2008, *ApJ*, 672, 757
- Cocchia, F., Fiore, F., Vignali, C., et al. 2007, *A&A*, 466, 31
- Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, *ApJ*, 741, 8
- Colpi, M., & Dotti, M. 2009, arXiv:0906.4339
- Comastri, A., Brusa, M., Mignoli, M., & HELLAS2XMM Team 2003, *Astronomische Nachrichten*, 324, 28
- Comastri, A., Mignoli, M., Ciliegi, P., et al. 2002, *ApJ*, 571, 771
- Comastri, A., Ranalli, P., Iwasawa, K., et al. 2011, *A&A*, 526, L9
- Comastri, A., Setti, G., Zamorani, G., & Hasinger, G. 1995, *A&A*, 296, 1
- Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, 575, A40

- Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, *AJ*, 112, 839
- Cruddace, R. G., Hasinger, G. R., & Schmitt, J. H. 1988, *European Southern Observatory Conference and Workshop Proceedings*, 28, 177
- Dadina, M., Cappi, M., Malaguti, G., Ponti, G., & de Rosa, A. 2005, *A&A*, 442, 461
- Damiani, F., Maggio, A., Micela, G., & Sciortino, S. 1997, *ApJ*, 483, 350
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010, *MNRAS*, 401, 1552
- De Rosa, G., Decarli, R., Walter, F., et al. 2011, *ApJ*, 739, 56
- Della Ceca, R., Caccianiga, A., Severgnini, P., et al. 2008, *A&A*, 487, 119
- Della Ceca, R., Maccacaro, T., Caccianiga, A., et al. 2004, *A&A*, 428, 383
- Delvecchio, I., Gruppioni, C., Pozzi, F., et al. 2014, *MNRAS*, 439, 2736
- Devecchi, B., & Volonteri, M. 2009, *ApJ*, 694, 302
- Di Matteo, T., Khandai, N., DeGraf, C., et al. 2012, *ApJL*, 745, L29
- Dietrich, M., Mathur, S., Grupe, D., & Komossa, S. 2009, *ApJ*, 696, 1998
- Done, C., Davis, S. W., Jin, C., Blaes, O., & Ward, M. 2012, *MNRAS*, 420, 1848
- Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142
- Donley, J. L., Rieke, G. H., Pérez-González, P. G., & Barro, G. 2008, *ApJ*, 687, 111
- Dubois, Y., Pichon, C., Haehnelt, M., et al. 2012, *MNRAS*, 423, 3616
- Dwelly, T., & Page, M. J. 2006, *MNRAS*, 372, 1755
- Ebrero, J., Carrera, F. J., Page, M. J., et al. 2009, *A&A*, 493, 55
- Eisenstein, D. J., & Loeb, A. 1995, *ApJ*, 443, 11
- Elitzur, M. 2008, *New Astronomy Review*, 52, 274
- Elvis, M., Hao, H., Civano, F., et al. 2012, *ApJ*, 759, 6
- Elvis, M., Schreier, E. J., Tonry, J., Davis, M., & Huchra, J. P. 1981, *ApJ*, 246, 20
- Elvis, M. 2000, *ApJ*, 545, 63
- 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
- Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, *ApJS*, 201, 23
- Fabian, A. C. 2006, *Astronomische Nachrichten*, 327, 943
- Fabian, A. C., Sanders, J. S., Etori, S., et al. 2000, *MNRAS*, 318, L65

- Fabian, A. C., & Miniutti, G. 2005, arXiv:astro-ph/0507409
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, 132, 117
- Fan, X., Strauss, M. A., Richards, G. T., et al. 2006, *AJ*, 131, 1203
- Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, *AJ*, 121, 54
- Ferrarese, L., & Merritt, D. 2000, *ApJL*, 539, L9
- Feruglio, C., Maiolino, R., Piconcelli, E., et al. 2010, *A&A*, 518, L155
- Finoguenov, A., Guzzo, L., Hasinger, G., et al. 2007, *ApJS*, 172, 182
- Fiore, F., Brusa, M., Cocchia, F., et al. 2003, *A&A*, 409, 79
- Fiore, F., Puccetti, S., Brusa, M., et al. 2009, *ApJ*, 693, 447
- Fischer, J., Sturm, E., González-Alfonso, E., et al. 2010, *A&A*, 518, L41
- Fontanot, F., Monaco, P., Cristiani, S., & Tozzi, P. 2006, *MNRAS*, 373, 1173
- Fragos, T., Lehmer, B., Tremmel, M., et al. 2013, *ApJ*, 764, 41
- Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, *Proceedings of the SPIE*, 6270, 62701V
- Gatti, M., Lamastra, A., Menci, N., Bongiorno, A., & Fiore, F. 2015, *A&A*, 576, A32
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, *ApJL*, 539, L13
- Gehrels, N. 1986, *ApJ*, 303, 336
- Georgakakis, A., Aird, J., Buchner, J., et al. 2015, *MNRAS*, 453, 1946
- Georgakakis, A., Mountrichas, G., Salvato, M., et al. 2014, *MNRAS*, 443, 3327
- Georgantopoulos, I., Comastri, A., Vignali, C., et al. 2013, *A&A*, 555, A43
- Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
- Gilli, R., Zamorani, G., Miyaji, T., et al. 2009, *A&A*, 494, 33
- Giustini, M., Cappi, M., Chartas, G., et al. 2011, *A&A*, 536, A49
- Glikman, E., Djorgovski, S. G., Stern, D., et al. 2011, *ApJL*, 728, L26
- Goulding, A. D., & Alexander, D. M. 2009, *MNRAS*, 398, 1165
- Goulding, A. D., Alexander, D. M., Lehmer, B. D., & Mullaney, J. R. 2010, *MNRAS*, 406, 597
- Goulding, A. D., Alexander, D. M., Mullaney, J. R., et al. 2011, *MNRAS*, 411, 1231
- Green, P. J., Aldcroft, T. L., Richards, G. T., et al. 2009, *ApJ*, 690, 644

- Greene, J. E., Ho, L. C., & Barth, A. J. 2008, *ApJ*, 688, 159
- Grier, C. J., Martini, P., Watson, L. C., et al. 2013, *ApJ*, 773, 90
- Guainazzi, M., Bianchi, S., & Dovčiak, M. 2006, *Astronomische Nachrichten*, 327, 1032
- Guedes, J., Madau, P., Mayer, L., & Callegari, S. 2011, *ApJ*, 729, 125
- Haardt, F., & Maraschi, L. 1991, *ApJL*, 380, L51
- Haardt, F., & Maraschi, L. 1993, *ApJ*, 413, 507
- Haehnelt, M. G., & Rees, M. J. 1993, *MNRAS*, 263, 168
- Hao, H., Elvis, M., Civano, F., et al. 2014, *MNRAS*, 438, 1288
- Harrison, F. A., Craig, W. W., Christensen, F. E., et al. 2013, *ApJ*, 770, 103
- Hasinger, G. 2008, *A&A*, 490, 905
- Hasinger, G., Burg, R., Giacconi, R., et al. 1993, *A&A*, 275, 1
- Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, *ApJS*, 172, 29
- Hasinger, G., Miyaji, T., & Schmidt, M. 2005, *A&A*, 441, 417
- Hickox, R. C., & Markevitch, M. 2006, *ApJ*, 645, 95
- Ho, L. C. 2008, *Annual Review of Astron and Astrophysics*, 46, 475
- 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., Cox, T. J., & Kereš, D. 2008, *ApJS*, 175, 356
- Hornschemeier, A. E., Brandt, W. N., Garmire, G. P., et al. 2001, *ApJ*, 554, 742
- Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, *ApJ*, 796, 60
- Huertas-Company, M., Pérez-González, P. G., Mei, S., et al. 2015, *ApJ*, 809, 95
- Ikeda, H., Nagao, T., Matsuoka, K., et al. 2011, *ApJL*, 728, L25
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, 690, 1236
- Iwasawa, K., Gilli, R., Vignali, C., et al. 2012, *A&A*, 546, A84
- Iwasawa, K., & Taniguchi, Y. 1993, *ApJL*, 413, L15
- Jiang, Y.-F., Greene, J. E., Ho, L. C., Xiao, T., & Barth, A. J. 2011, *ApJ*, 742, 68
- Jonker, P. G., Torres, M. A. P., Fabian, A. C., et al. 2010, *MNRAS*, 407, 645

- Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, *A&A*, 440, 775
- Kalfountzou, E., Civano, F., Elvis, M., Trichas, M., & Green, P. 2014, *MNRAS*, 445, 1430
- Kartaltepe, J. S., Sanders, D. B., Le Floch, E., et al. 2010, *ApJ*, 709, 572
- Kaspi, S., Maoz, D., Netzer, H., et al. 2005, *ApJ*, 629, 61
- Kettula, K., Finoguenov, A., Massey, R., et al. 2013, *ApJ*, 778, 74
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, *ApJ*, 556, 121
- Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, *MNRAS*, 372, 961
- Kim, D.-W., & Fabbiano, G. 2013, *ApJ*, 776, 116
- Kim, M., Kim, D.-W., Wilkes, B. J., et al. 2007, *ApJS*, 169, 401
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- Koekemoer, A. M., Alexander, D. M., Bauer, F. E., et al. 2004, *ApJL*, 600, L123
- Komossa, S. 2012, *Advances in Astronomy*, 2012, 364973
- Komossa, S., Zhou, H., & Lu, H. 2008, *ApJL*, 678, L81
- Kormendy, J., Bender, R., & Cornell, M. E. 2011, *Nature*, 469, 374
- Kormendy, J., & Kennicutt, R. C., Jr. 2004, *Annual Review of Astron and Astrophysics*, 42, 603
- Kormendy, J., & Richstone, D. 1995, *Annual Review of Astron and Astrophysics*, 33, 581
- Koss, M., Mushotzky, R., Veilleux, S., & Winter, L. 2010, *ApJL*, 716, L125
- Koushiappas, S. M., Bullock, J. S., & Dekel, A. 2004, *MNRAS*, 354, 292
- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, 218, 15
- Krumpe, M., Miyaji, T., Brunner, H., et al. 2015, *MNRAS*, 446, 911
- Kurk, J. D., Walter, F., Fan, X., et al. 2007, *ApJ*, 669, 32
- La Franca, F., Fiore, F., Comastri, A., et al. 2005, *ApJ*, 635, 864
- Lackner, C. N., Silverman, J. D., Salvato, M., et al. 2014, *AJ*, 148, 137
- LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2013, *MNRAS*, 436, 3581
- LaMassa, S. M., Urry, C. M., Glikman, E., et al. 2013, *MNRAS*, 432, 1351
- Lamastra, A., Perola, G. C., & Matt, G. 2006, *A&A*, 449, 551

- Lansbury, G. B., Alexander, D. M., Del Moro, A., et al. 2014, *ApJ*, 785, 17
- Lansbury, G. B., Gandhi, P., Alexander, D. M., et al. 2015, *ApJ*, 809, 115
- Lanzuisi, G., Civano, F., Elvis, M., et al. 2013, *MNRAS*, 431, 978
- Lanzuisi, G., Civano, F., Marchesi, S., et al. 2013, *ApJ*, 778, 62
- Lanzuisi, G., Ponti, G., Salvato, M., et al. 2014, *ApJ*, 781, 105
- Lanzuisi, G., Ranalli, P., Georgantopoulos, I., et al. 2015, *A&A*, 573, A137
- Lawrence, A. 1991, *MNRAS*, 252, 586
- Lawrence, A., & Elvis, M. 1982, *ApJ*, 256, 410
- Lawrence, A., & Elvis, M. 2010, *ApJ*, 714, 561
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79
- Lehmann, I., Hasinger, G., Schmidt, M., et al. 2001, *A&A*, 371, 833
- Lehmer, B. D., Alexander, D. M., Bauer, F. E., et al. 2010, *ApJ*, 724, 559
- Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, *ApJ*, 752, 46
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, *ApJS*, 184, 218
- Liu, X., Shen, Y., Strauss, M. A., & Hao, L. 2011, *ApJ*, 737, 101
- Lodato, G., & Natarajan, P. 2006, *MNRAS*, 371, 1813
- Loeb, A. 2007, *Physical Review Letters*, 99, 041103
- Loeb, A., & Rasio, F. A. 1994, *ApJ*, 432, 52
- Longinotti, A. L., Sim, S. A., Nandra, K., & Cappi, M. 2007, *MNRAS*, 374, 237
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2011, *ApJ*, 740, 37
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, *ApJS*, 187, 560
- Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, *MNRAS*, 425, 623
- Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013, *ApJ*, 777, 86
- Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, *ApJ*, 326, 680
- Madau, P., & Dickinson, M. 2014, *Annual Review of Astron and Astrophysics*, 52, 415
- Madau, P., & Rees, M. J. 2001, *ApJL*, 551, L27
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285

- Mainieri, V., Bergeron, J., Hasinger, G., et al. 2002, *A&A*, 393, 425
- Mainieri, V., Bongiorno, A., Merloni, A., et al. 2011, *A&A*, 535, A80
- Maiolino, R., Comastri, A., Gilli, R., et al. 2003, *MNRAS*, 344, L59
- Maiolino, R., Gallerani, S., Neri, R., et al. 2012, *MNRAS*, 425, L66
- Maiolino, R., & Rieke, G. H. 1995, *ApJ*, 454, 95
- Maiolino, R., Shemmer, O., Imanishi, M., et al. 2007, *A&A*, 468, 979
- Marconi, A., Risaliti, G., Gilli, R., et al. 2004, *MNRAS*, 351, 169
- Markevitch, M., Bautz, M. W., Biller, B., et al. 2003, *ApJ*, 583, 70
- Marziani, P., Sulentic, J. W., Stirpe, G. M., Zamfir, S., & Calvani, M. 2009, *A&A*, 495, 83
- Masters, D., Capak, P., Salvato, M., et al. 2012, *ApJ*, 755, 169
- Mateos, S., Warwick, R. S., Carrera, F. J., et al. 2008, *A&A*, 492, 51
- Mathur, S., Fields, D., Peterson, B. M., & Grupe, D. 2012, *ApJ*, 754, 146
- Matt, G., Porquet, D., Bianchi, S., et al. 2005, *A&A*, 435, 857
- McConnell, N. J., & Ma, C.-P. 2013, *ApJ*, 764, 184
- McCracken, H. J., Capak, P., Salvato, M., et al. 2010, *ApJ*, 708, 202
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- McGreer, I. D., Jiang, L., Fan, X., et al. 2013, *ApJ*, 768, 105
- Menci, N., Gatti, M., Fiore, F., & Lamastra, A. 2014, *A&A*, 569, A37
- Merloni, A., Bongiorno, A., Bolzonella, M., et al. 2010, *ApJ*, 708, 137
- Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, *MNRAS*, 437, 3550
- Merloni, A., Predehl, P., Becker, W., et al. 2012, *arXiv:1209.3114*
- Merritt, D., Storchi-Bergmann, T., Robinson, A., et al. 2006, *MNRAS*, 367, 1746
- Mezcua, M., Farrell, S. A., Gladstone, J. C., & Lobanov, A. P. 2013, *MNRAS*, 436, 1546
- Mignoli, M., Pozzetti, L., Comastri, A., et al. 2004, *A&A*, 418, 827
- Mignoli, M., Vignali, C., Gilli, R., et al. 2013, *A&A*, 556, A29
- Mineo, S., Gilfanov, M., & Sunyaev, R. 2012, *MNRAS*, 419, 2095
- Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, *ApJ*, 804, 104
- Mor, R., & Trakhtenbrot, B. 2011, *ApJL*, 737, L36

- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Nature*, 474, 616
- Murray, S. S., Kenter, A., Forman, W. R., et al. 2005, *ApJS*, 161, 1
- Mushotzky, R. F., Done, C., & Pounds, K. A. 1993, *Annual Review of Astron and Astrophysics*, 31, 717
- Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1999, *ApJL*, 523, L17
- Nandra, K., Laird, E. S., Aird, J. A., et al. 2015, *ApJS*, 220, 10
- Nandra, K., O'Neill, P. M., George, I. M., & Reeves, J. N. 2007, *MNRAS*, 382, 194
- Nandra, K., Barret, D., Barcons, X., et al. 2013, arXiv:1306.2307
- Nardini, E., & Risaliti, G. 2011, *MNRAS*, 415, 619
- Narsky, I. 2000, arXiv:hep-ex/0005019
- Nenkova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž., & Elitzur, M. 2008, *ApJ*, 685, 160
- Netzer, H., Lira, P., Trakhtenbrot, B., Shemmer, O., & Cury, I. 2007, *ApJ*, 671, 1256
- Netzer, H., & Trakhtenbrot, B. 2007, *ApJ*, 654, 754
- Novak, M., Smolčić, V., Civano, F., et al. 2015, *MNRAS*, 447, 1282
- Onken, C. A., Ferrarese, L., Merritt, D., et al. 2004, *ApJ*, 615, 645
- Orban de Xivry, G., Davies, R., Schartmann, M., et al. 2011, *MNRAS*, 417, 2721
- Page, M. J., Stevens, J. A., Ivison, R. J., & Carrera, F. J. 2004, *ApJL*, 611, L85
- Paggi, A., Fabbiano, G., Civano, F., et al. 2015, arXiv:1507.03170
- Park, T., Kashyap, V. L., Siemiginowska, A., et al. 2006, *ApJ*, 652, 610
- Peres, A. 1962, *Physical Review*, 128, 2471
- Perola, G. C., Puccetti, S., Fiore, F., et al. 2004, *A&A*, 421, 491
- Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, *ApJ*, 613, 682
- Peterson, B. M., & Horne, K. 2004, *Astronomische Nachrichten*, 325, 248
- Peterson, B. M. 1993, *Publications of the ASP*, 105, 247
- Pierre, M., Valtchanov, I., Altieri, B., et al. 2004, *Journal of Cosmology and Astroparticle Physics*, 9, 011
- Puccetti, S., Vignali, C., Cappelluti, N., et al. 2009, *ApJS*, 185, 586

- Pudritz, R. E., Hardcastle, M. J., & Gabuzda, D. C. 2012, *Space Science Reviews*, 169, 27
- Puech, M., Hammer, F., Rodrigues, M., et al. 2014, *MNRAS*, 443, L49
- Ranalli, P., Comastri, A., & Setti, G. 2003, *A&A*, 399, 39
- Ranalli, P., Georgantopoulos, I., Corral, A., et al. 2015, *A&A*, 577, A121
- Reeves, J. N., Pounds, K., Uttley, P., et al. 2005, *ApJL*, 633, L81
- Reynolds, C. S., & Nowak, M. A. 2003, *Physics Reports*, 377, 389
- Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, *ApJS*, 166, 470
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, *AJ*, 131, 2766
- Risaliti, G., & Elvis, M. 2004, *Supermassive Black Holes in the Distant Universe*, 308, 187
- Robinson, A., Young, S., Axon, D. J., Kharb, P., & Smith, J. E. 2010, *ApJL*, 717, L122
- Rosario, D. J., McGurk, R. C., Max, C. E., et al. 2011, *ApJ*, 739, 44
- Ross, N. P., McGreer, I. D., White, M., et al. 2013, *ApJ*, 773, 14
- Rupke, D. S. N., & Veilleux, S. 2011, *ApJL*, 729, L27
- Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61
- Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, 172, 86
- Satyapal, S., Vega, D., Dudik, R. P., Abel, N. P., & Heckman, T. 2008, *ApJ*, 677, 926
- Sazonov, S., Willner, S. P., Goulding, A. D., et al. 2012, *ApJ*, 757, 181
- Schawinski, K., Dowlin, N., Thomas, D., Urry, C. M., & Edmondson, E. 2010, *ApJL*, 714, L108
- Schinnerer, E., Smolčić, V., Carilli, C. L., et al. 2007, *ApJS*, 172, 46
- Schmidt, M., Hasinger, G., Gunn, J., et al. 1998, *A&A*, 329, 495
- Schmidt, M. 1968, *ApJ*, 151, 393
- Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, *AJ*, 110, 68
- Schneider, R., Omukai, K., Inoue, A. K., & Ferrara, A. 2006, *MNRAS*, 369, 1437
- Schnittman, J. D., & Krolik, J. H. 2010, *ApJ*, 712, 908
- Scott, N., & Graham, A. W. 2013, *ApJ*, 763, 76
- Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, *ApJS*, 172, 38

- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Setti, G., & Woltjer, L. 1989, *A&A*, 224, L21
- Shankar, F. 2009, *New Astronomy Review*, 53, 57
- Shemmer, O., Netzer, H., Maiolino, R., et al. 2004, *ApJ*, 614, 547
- Shen, Y. 2009, *ApJ*, 704, 89
- Shields, G. A., Bonning, E. W., & Salvander, S. 2009, *ApJ*, 696, 1367
- Sijacki, D., Springel, V., & Haehnelt, M. G. 2011, *MNRAS*, 414, 3656
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, *ApJ*, 509, 103
- Silverman, J. D., Daddi, E., Rodighiero, G., et al. 2015, *ApJL*, 812, L23
- Silverman, J. D., Green, P. J., Barkhouse, W. A., et al. 2008, *ApJ*, 679, 118
- Simpson, C. 2005, *MNRAS*, 360, 565
- Soltan, A. 1982, *MNRAS*, 200, 115
- Steffen, A. T., Barger, A. J., Cowie, L. L., Mushotzky, R. F., & Yang, Y. 2003, *ApJL*, 596, L23
- Steinhardt, C. L., Schramm, M., Silverman, J. D., et al. 2012, *ApJ*, 759, 24
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, *ApJ*, 753, 30
- Stern, J., & Laor, A. 2012, *MNRAS*, 426, 2703
- Stoche, J. T., Morris, S. L., Gioia, I. M., et al. 1991, *ApJS*, 76, 813
- Sturm, E., González-Alfonso, E., Veilleux, S., et al. 2011, *ApJL*, 733, L16
- Sturm, E., Hasinger, G., Lehmann, I., et al. 2006, *ApJ*, 642, 81
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, 259, 413
- Tananbaum, H., Avni, Y., Branduardi, G., et al. 1979, *ApJL*, 234, L9
- Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, *ApJS*, 172, 9
- Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, *ApJ*, 565, 773
- Tombesi, F., Cappi, M., Reeves, J. N., et al. 2010, *A&A*, 521, A57
- Tombesi, F., Meléndez, M., Veilleux, S., et al. 2015, *Nature*, 519, 436
- Trakhtenbrot, B., & Netzer, H. 2012, *MNRAS*, 427, 3081
- Trakhtenbrot, B., Netzer, H., Lira, P., & Shemmer, O. 2011, *ApJ*, 730, 7
- Trakhtenbrot, B., Urry, C. M., Civano, F., et al. 2015, *Science*, 349, 168

- Treister, E., Schawinski, K., Urry, C. M., & Simmons, B. D. 2012, *ApJL*, 758, L39
- Treister, E., Urry, C. M., & Virani, S. 2009, *ApJ*, 696, 110
- Treister, E., Krolik, J. H., & Dullemond, C. 2008, *ApJ*, 679, 140
- Treister, E., & Urry, C. M. 2006, *ApJL*, 652, L79
- Trichas, M., Green, P. J., Silverman, J. D., et al. 2012, *ApJS*, 200, 17
- Trump, J. R., Impey, C. D., Elvis, M., et al. 2009, *ApJ*, 696, 1195
- Trump, J. R., Impey, C. D., McCarthy, P. J., et al. 2007, *ApJS*, 172, 383
- Trump, J. R., Impey, C. D., Taniguchi, Y., et al. 2009, *ApJ*, 706, 797
- Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, *ApJ*, 786, 104
- Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, *ApJ*, 598, 886
- Urry, C. M., & Padovani, P. 1995, *Publications of the ASP*, 107, 803
- Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, *AJ*, 122, 549
- Vanden Berk, D. E., Shen, J., Yip, C.-W., et al. 2006, *AJ*, 131, 84
- Vaughan, S., & Uttley, P. 2008, *MNRAS*, 390, 421
- Veilleux, S., & Osterbrock, D. E. 1987, *ApJS*, 63, 295
- Vikhlinin, A., Reid, P., Tananbaum, H., et al. 2012, *Proceedings of the SPIE*, 8443, 844316
- Vito, F., Gilli, R., Vignali, C., et al. 2014, *MNRAS*, 445, 3557
- Vito, F., Vignali, C., Gilli, R., et al. 2013, *MNRAS*, 428, 354
- Volonteri, M., & Begelman, M. C. 2010, *MNRAS*, 409, 1022
- Volonteri, M., Haardt, F., & Madau, P. 2003, *ApJ*, 582, 559
- Volonteri, M., & Natarajan, P. 2009, *MNRAS*, 400, 1911
- Volonteri, M. 2012, *Science*, 337, 544
- Weigel, A. K., Schawinski, K., Treister, E., et al. 2015, *MNRAS*, 448, 3167
- Wild, V., Kauffmann, G., Heckman, T., et al. 2007, *MNRAS*, 381, 543
- Wilkes, B. J., Kilgard, R., Kim, D.-W., et al. 2009, *ApJS*, 185, 433
- Wilkes, B. J., Kuraszekiewicz, J., Haas, M., et al. 2013, *ApJ*, 773, 15
- Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, *AJ*, 140, 546
- Willott, C. J., Delorme, P., Reyl e, C., et al. 2010, *AJ*, 139, 906

- Willott, C. J., Delorme, P., Reylé, C., et al. 2009, *AJ*, 137, 3541
- Willott, C. J., McLure, R. J., & Jarvis, M. J. 2003, *ApJL*, 587, L15
- Wright, N. J., Drake, J. J., & Civano, F. 2010, *ApJ*, 725, 480
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, 195, 10
- Yaqoob, T., & Serlemitsos, P. 2005, *ApJ*, 623, 112
- Young, M., Elvis, M., & Risaliti, G. 2010, *ApJ*, 708, 1388
- Zakamska, N. L., Gómez, L., Strauss, M. A., & Krolik, J. H. 2008, *AJ*, 136, 1607
- Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2006, *AJ*, 132, 1496
- Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, *ApJS*, 172, 468