Technische Universität München
Max-Planck-Institut für extraterrestrische Physik
Garching bei München

Large scale structures in X-ray surveys

Nico Cappelluti

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:		Univ.–Prof. Dr. M. Ratz
Prüfer:	1.	Hon.–Prof. Dr. G. Hasinger
	2.	Univ.–Prof. Dr. F. von Feilitzsch

Die Dissertation wurde am 27.09.2007 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 17.10.2007 angenommen.

Zusammenfassung

Die Beziehung zwischen Aktiven Galaxien und den sie umgebenden großräumigen Strukturen wurde untersucht. Die Untersuchungen wurden unter Benutzung der Daten der ROSAT-NEP und XMM-COSMOS Durchmusterungen durchgeführt. Ein spezielles Datenanalyseverfahren für großflächige Röntgendurchmusterungen wurde entwickelt. Das wesentliches Ergebnis ist, dass sich Aktive Galaxien bevorzugt in Halos aus dunkler Materie der Größenordnung log(M)=13 Sonnenmassen befinden. Außerdem ergab sich, dass Aktive Galaxien stark mit Galaxienhaufen korreliert sind.

Abstract

The relation between AGN and the large scale structures environment in which they reside has been investigate. The work has been performed making use of the ROSAT-NEP and XMM-COSMOS survey data. A sophisticated data analysis technique has been developed for wide field X-ray surveys. The main result is that AGN preferentially reside in Dark Matter halos of the order of Log(M)=13 solar masses. It has also been determined that AGN are strongly correlated with Galaxy clusters.

Contents

Chapter 1

Large scale structures in X-ray surveys: an overview

1.1 The history of the X-ray background

In 1962 during a rocket experiment aimed to detect possible fluorescent X-ray emission from the moon, Riccardo Giacconi and his collaborators serendipitously discovered the first extrasolar X-ray source, Sco-X1. Together with this emission, the Geiger counters on board the rocket detected diffuse emission X-ray coming from all the sky: the Cosmic X-ray Background (hereinafter XRB). It is worth noticing that the XRB was the first discovered cosmic background.

In the 70's the first X-ray surveys with *Uhuru* and *Ariel V* showed that because of its high level of isotropy the XRB should have an extragalactic origin. Setti & Woltjer (1973) predicted that if the XRB was produced by unresolved extragalactic X-ray sources, their source density would have to be relatively high ($> 10^6$ sr^{-1}). HEAO-1 showed also that the spectrum of the XRB could be well fit by thermal Bremsstrahlung model with a temperature of $\sim$40 keV. This originally suggested that the XRB could arise from a hot intergalactic medium; this hypothesis was however discarded in 1990 by not observing the Compton distortion on the CMB spectrum with COBE (Mather et al. 1990).

With the utilization of Wolter telescopes, the discrete nature of the XRB became rapidly clear. The *Einstein* observatory was in fact able to resolve $\sim$25% of the soft-XRB into discrete sources which were mainly identified with AGN (Giacconi et al. 1979). A milestone in the study of the XRB was the all-sky-survey conducted by the German X-ray satellite ROSAT (Trümper 1982)

and its deep surveys, which resolved $\sim$75% of the XRB into discrete sources (see e.g Hasinger et al. 1993, 1998). The source density of AGN measured with ROSAT (i.e. $\sim$850 deg^{-2}) is larger than in any other wavelength and was able, for the first time in the X-ray band, to constrain the cosmological evolution of SuperMassive Black Holes (SMBHs) (Miyaji et al. 2000) on a broad range of redshifts and luminosities. In the 2-10 keV band, despite the source confusion introduced by the broad PSF of the telescope, ASCA performed several deep survey reaching a limit of $\sim$ 100 deg^{-2} resolving $\sim$35% of the XRB (see e.g Ueda et al. 1999). At higher energies the Italian satellite *Beppo Sax* performed a survey in the 5-10 keV band resolving $\sim$30% of the XRB (Fiore et al. 2001).

A revolution in the study of the XRB happened with the launch of high-throughput X-ray telescopes XMM-*Newton* and *Chandra*. These telescopes with their high angular resolution (FWHM$\sim$0.5" for *Chandra* and FWHM$\sim$6" for XMM-*Newton*) and high throughput (up to 3000 cm^2 @1.5 keV with XMM-*Newton*), gave a final push for the solution of the XRB enigma. It is worth citing the deep *Chandra* and XMM-*Newton* survey of the *Chandra* deep fields and of the Lockman hole (Rosati et al. 2002; Bauer et al. 2004; Hasinger et al. 2001) which were able to resolve $\sim$95% of the 0.5–2 keV XRB into discrete sources, mainly AGN (Moretti et al. 2003). These surveys reach a flux limit of the order of $\sim 1\times10^{-16}$ cgs in the 0.5–2 keV band and allow to observe up to 3000 AGN deg^{-2}. With such a high photon statistics modern X-ray surveys are able to detect and constrain the spectral properties of AGN up to z $\sim$4-5. Together with the observations, important studies were also conducted to understand the spectral shape of the XRB. There is in fact a contradiction between the experimental evidence that the XRB is made by AGN and the observed quasi-thermal spectrum. In the 1-10 keV band, for example, this thermal spectrum can be well approximated with a power-law with spectral photon index Γ=1.4, while the average X-ray spectrum of AGN observed till that period, showed a spectral index $\Gamma \sim 1.7 \div 1.9$. This contradiction is known as the spectral paradox of the XRB. The shape of the XRB spectrum (see Fig. 1.1) was interpreted by population synthesis models based on the unified model of AGN including effects introduced by dust absorption (Madau et al. 1994; Comastri et al. 1995; Gilli et al. 2001; Gilli, Comastri & Hasinger 2006).

The models predict that the XRB is mainly formed by AGN and clusters of galaxies. In particular to explain the cut off of the XRB spectrum, below 20-30 keV, a conspicuous fraction of AGN should be surrounded by obscuring

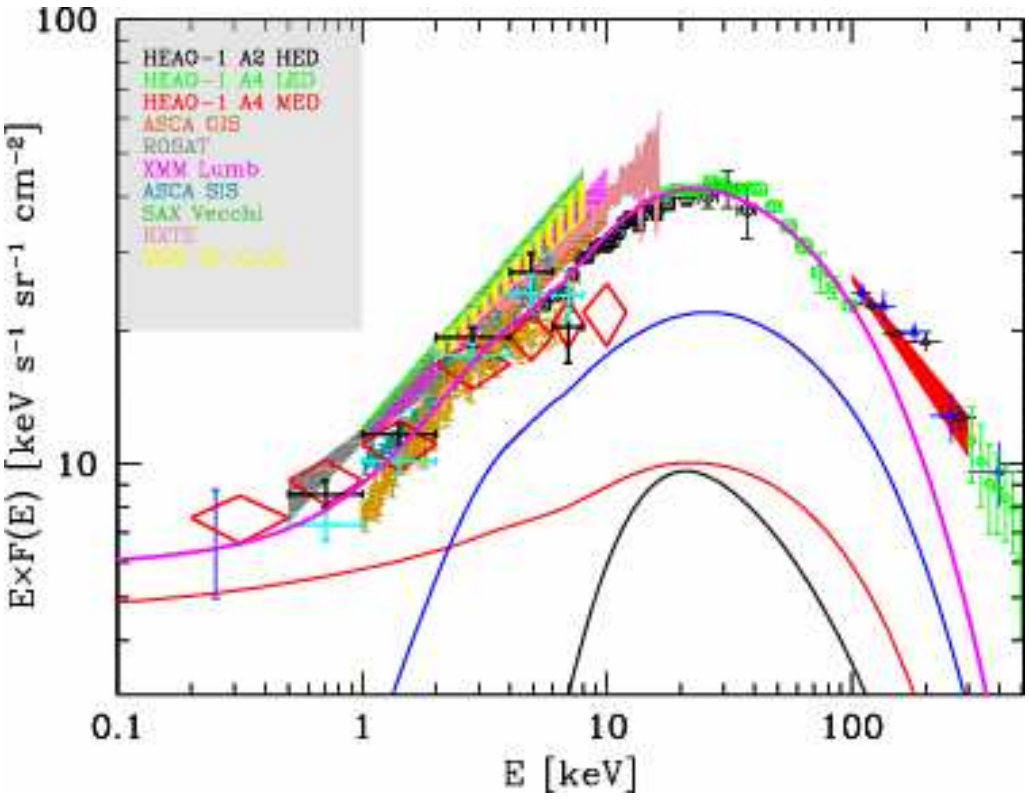

Figure 1.1: The spectrum of the XRB as measured by different instruments. The magenta line represents the spectrum obtained with the XRB population synthesis by Gilli, Comastri & Hasinger (2006) including all the classes of AGN and galaxy clusters. The contribution of unobscured AGN, Compton Thin AGN and Compton thick AGN are plotted with red, blue and black lines, respectively.

gas. Among these absorbed sources, models predict a remarkable fraction of Compton thick AGN to match the observed intensity at 30 keV. Compton thick AGN are supposed to be X-ray emitting SMBH surrounded by obscuring dust with a column density nH$> \sigma_T^{-1}$; yielding an optical depth for Compton scattering $\tau_C = 1$. This causes most of the light below 5 keV to be completely absorbed, making the detection of these objects very difficult in the energy range of focusing X-ray telescopes. Being very faint in the 0.5-10 keV energy band, at the flux limit of the modern surveys the fraction of Compton thick AGN observed up top now is of the order of 5-6%.

The X-ray observatory HEAO-1 mapped the all sky distribution of the X-ray background. Below 2 keV the XRB the large scale anisotropy is dominated by a galactic contribution, also a dipole contribution has been detected,

aligned with the motion of the earth against the CMB reference frame (Scharf et al. 2000). The amplitude of the dipole depends mostly on two components: the kinematical effect of our motion and the excess emission due to the structures in the great attractor region. Treyer et al. (1998), analyzed the fluctuations of the HEAO-1 A2 XRB at higher multipoles. They discovered that the discrete nature of the XRB contributes with a constant term to all the multipoles which scales as $S_{cut}^{0.5}$, where S_{cut} is the limiting flux at which the sources were excised. The signal showed a growth toward low order multipoles according to a gravitational collapse scenario. Their analysis lead to an estimate of the bias factor for the pointlike sources in the XRB $b_X \sim 1-2$. On scales of few arcminutes (typical scale of galaxy clustering), data from imaging telescopes have been used. Carrera et al. 1992 estimated that on those scales the autocorrelation of the XRB should reflect the autocorrelation function of X-ray sources at redshift of ∼1-2. The advent of modern X-ray telescopes made possible the study of the anisotropy of the XRB in terms of source clustering, opening a completely new research branch.

1.2 Cosmology and large scale structures with AGNs

The clustering of galaxies, which are supposed to be tracers of the underlying dark matter distribution, gives a powerful test of hierarchical structure formation theory. The galaxy autocorrelation function can be represented by a power-law $\xi(r) = (r/r_0)^{-\gamma}$, with $\gamma \sim 1.7$ and $r_0 \sim 5\ h^{-1}$ Mpc (see e.g. Hawkins et al. 2003,and references therein). Interestingly, measurement of the z>3 galaxy clustering showed no evidence of an evolution of the comoving clustering length (Giavalisco et al. 1998). According to the theory of structure formation the clustering at high redshift should be weaker than now. This discrepancy has been explained with the theory of biasing. The linear theory of biasing (Kaiser 1987) was first introduced to explain the different amplitude of the galaxy and galaxy clusters correlation functions. In this framework biasing is assumed to be statistical in nature: galaxies and clusters are identified as high peaks of an underlying initially random density field. Let us consider the correlation function of a certain kind of tracers $\xi_{tr}(r)$ (such as galaxies or galaxy clusters), the linear bias parameter is given by:

$$\xi_{tr}(r) = b_{tr}^2 \xi_m(r), \tag{1.1}$$

where b is the linear bias factor. Using the Press & Schechter formalism Mo & White (1996), linked the clustering of dark matter halos and their mass. In this context the biasing of mass tracers is related to the mass of the dark haloes in which they preferentially reside. Another important implication of such a formalism is that the bias turns out to be a function of redshift. A good determination of the bias factor at different epochs is therefore crucial to understand the structure formation in the universe. Though galaxies represent the most abundant observable tracers of the matter in the universe, optical surveys are able to provide constraints on the clustering of these objects only at low redshift. Because of their high luminosity (i.e. up to 10^{47} erg s^{-1}), bright QSO are the most distant observable objects in the universe. In a recent paper Shen et al. (2007) measured the clustering of 4426 luminous QSO at z>3 in the Sloan Digital sky survey. They measured a correlation length of the order of $\sim$15 h^{-1} Mpc and a slope of $\gamma \sim 2$. This indicates that QSO form in massive, and therefore highly biased, dark matter haloes. This conclusion is also in close agreement with recent observation of the relation between the mass of the central black hole in nearby galaxy and the velocity dispertion. Under the assumption of a linear dependence of the QSO luminosity and the mass the central black hole, bright QSO are powered by very massive black holes. Extrapolating this consideration means that high-z luminous QSO reside in massive dark matter haloes. They in fact estimate a minimum mass of the halos that harbor bright QSO of the order of $\sim 5 \times 10^{12} M_{\odot}$. With the same formalism they were also able to connect the number density of QSO and their clustering with the QSO lifetime. They estimated a QSO lifetime of the order of $\sim$100 Myr.

Modern population synthesis models of the XRB estimate that the majority of the accretion power in the universe, is obscured (see e.g. Gilli, Comastri & Hasinger 2006). According to this paradigm optical surveys are not sensitive to medium-low luminosity obscured AGN, since most of the optical photons from the nucleus are absorbed by the dust. Therefore, the resulting spectrum is dominated by the galaxy component.

AGN strongly emit X-rays, which on the contrary are capable the penetrate the absorbing gas and are not sensitive to the dilution by the galaxy starlight. X-ray surveys are therefore capable to obtain the highest surface density of AGN since they are able to detect also faint obscured sources; as an example deep XMM-*Newton* and Chandra surveys detected >1500 AGN deg^{-2} against the <500 AGN deg^{-2} of optical surveys. In comparison with

the optical, X-ray surveys have the disadvantage of a small field of view and the long observing times. The studies of clustering of AGN in the X-ray band are therefore potentially fruitful since they are less affected by the selection effects mentioned above.

Vikhlinin & Forman (1995) measured for the first time a statistically significant signal of clustering using data of the ROSAT all-sky survey. By making use of the angular correlation function of X-ray sources they measured an angular correlation length of $\sim$4" which, using the Limber's equation for the deprojection yielded a clustering length of $\sim$6 Mpc (technical details in Chap. 3). On the other hand Carrera et al. (1998), measured a small significance low clustering signal using spectroscopic counterparts of ROSAT sources. By using the counts in cell technique, Cowie et al. (2002) measured a 30% field to field variation of the source number counts in Chandra deep surveys. Such an amplitude of the fluctuations is larger than that expected from Poisson statistics. A similar result was obtained by Manners et al. (2003) and both works suggested that this effect is due to the large scale structure. Yang et al. (2003) measured a significant clustering signal both in the soft and hard X-ray bands. Interesting enough, the hard X-ray clustering showed a factor 10 higher normalization than in the soft band. This result was not confirmed by studies of Gilli et al. (2005) and Yang et al. (2006). The work of Gilli et al. (2005) showed a difference between the spatial clustering in the CDFN and CDFS, this difference was explained with the presence of sheet like large scale structures in the CDFS. In particular they measured a correlation length r_0 $\sim$9 Mpc in the CDFS and $\sim$5 Mpc in the CDFN. When excising the two spikes at z=0.67 and z=0.73 in the redshift distribution of the CDFS sources, the difference between the two correlation function was not significant anymore. The work of Yang et al. (2006) measured together with a weak dependence of the AGN clustering, a weakly significant evolution of the clustering length with redshift consistent with the biasing evolution scenario modeled by Mo & White (1996). The works of Basilakos et al. (2004, 2005) measured a correlation length of $\sim$20 Mpc which is the highest ever recorded. By merging the results obtained with AGN and results from SNIa they were able, for the first time, to derive cosmological parameters using AGN clustering.

1.3 The importance of wide field surveys

X-ray pencil beam surveys such as the CDFN, the CDFS and the XMM-*Newton* Lockmann Hole are able to detect AGN down to very faint fluxes. The typical areas covered in these surveys are <0.2 deg^2, which at a median redshift of 1.5 corresponds to a spatial dimension of $\sim$10 Mpc. On this scale the clustering is dominated by the chance occurrence of single large scale structure in the field. Clustering results in pencil beam surveys are therefore strongly affected by the so called "cosmic variance". As introduced in the previous section, Cowie et al. (2002) and Manners et al. (2003) measured significant source counts variation in *Chandra* deep fields and attributed this fluctuation to the LSS. Following Peebles (1980), the variance of counts in cells mainly depends on a Poissonian term which is less important in large samples and a term that depends on the strength of the angular clustering. The result of different authors presented in the previous section appear therefore not stable. Indeed, their samples suffer for low counts statistics, so that the Poisson term is dominant with the data in hands. One can conclude that to obtain unbiased information on the clustering properties of AGN wide field surveys are required where, cosmic variance effects introduced by small scale clustering and low counts statistics are smeared out.

The difference among the X-ray clustering results are attributable to many variables which may alterate the estimation of the correlation function. Assuming that the cosmic variance is a dominant source of fluctuation, which is difficult to model, an important step to estimate correlation functions is the production of a realistic random sample of sources which reproduces all the selection effects of the survey. In X-ray observations the sensitivity strongly varies across the field; in fact the resolution of X-ray telescopes degrades with the off-axis angle together with a loss of sensitivity because of the vignetting. This yields a high sensitivity on axis and and a low sensitivity off axis. In addition, because of the lower reflectivity for high energy photons (i.e. $E >2$ keV) there is a strong drop of the sensitivity above 2 keV and a larger vignetting. The final result is a complicate pattern of sensitivities across the field of view which is difficult to model. The situation is additionally complicated by the presence of gaps between the chips of the CCD detectors. In order to overcome such limitations, astronomers designated wide field X-ray surveys which, with appropriate tiling of the telescope's field of view, are able to flatten the sensitivity variation on large solid angles. A step forward for

X-ray surveys is the XMM-COSMOS survey, it aimed to reach 2 deg^2 of moderately deep XMM-*Newton* observations with homogeneous sensitivity (for a detailed overview of the XMM-COSMOS survey see Chap. 2).

As mentioned above, a wide field survey can sample various scales in the Universe and thus overcome the Poisson effect of cosmic variance. By removing this effect from the survey, one can construct a reference logN-logS of X-ray sources. On one side this precise determination can play a role to discriminate between different models of the XRB population (see e.g. Gilli, Comastri & Hasinger 2006): on the other hand, having a reliable logN-logS is important when creating random samples of X-ray sources with the purpose of estimating the correlation function. Most of the works cited above, generally use logN-logS estimated from their surveys to generate random samples. But due to the cosmic variance this can lead to a biased estimation of the correlation introduced by using a logN-logS which is biased itself by small scale structures. Plionis (2007) argued that the differences among correlation functions in various surveys are due to different depths of the surveys, cosmic variance and errors in modeling of the random samples. While the last two arguments can be easily validated the first argument strikes against the evidence of a constant comoving correlation length through cosmic times. On the other hand there Yang et al. (2006) measured a weak dependence of the clustering on X-ray luminosity which can explain such a explanation, but the data still do not allow to draw any final conclusion.

In this thesis a large part of the effort is concentrated on producing a reliable source catalog in XMM-COSMOS fields, which, using Monte Carlo simulations, allows the estimation of the AGN clustering with no or few selection biases. Another focus is on estimating a reliable logN-logS relation for constructing the right random sample for the estimation of the correlation function of AGN. Given the large area of the survey it is also possible to keep the cosmic variance under control by directly measuring its effects on the source counts on subfields of the survey.

1.4 AGN activity in dense environment

Clusters of galaxies host a wide diversity of galaxy populations, they are therefore perfect labs to study the galaxy evolution.

The first evidence of evolution of galaxies in dense environments is attributable to the observations of Butcher & Oemler (1984). They measured

an increase from z=o to z$\sim$0.5, of the fraction of blue star forming galaxies in Abell clusters greater than in the field. Over the same redshift range also the fraction of post-starburst galaxies increased more rapidly than in the field. The evolution of AGN in a high density environment such as clusters and groups, is still a matter of debate. There are indication that the fraction of AGN in clusters and groups is different from that in the field.

Optical studies of Dressler et al. (1985) suggested that AGN are extremely rare in dense environment. They measured a fraction of AGN in cluster fields of the order of $<$1% which is a factor 5 less that what observed in the field. The comparison with studies in other wavelengths gives confusing results. Branchesi et al. (2006) pointed out that radio galaxies are a factor 2 more frequent in clusters than in the field. Ledlow & Owen (1996), making use of the bivariate radio/optical luminosity function, suggest that the fraction and the evolution of radio-galaxies is independent from the environment. The bivariate luminosity function of a certain class of object A and B, measures the evolution of class A object as a function of the space density of class B objects. This method could be really fruitful for X-ray/optical studies when both populations of objects have a significant redshift completeness.

X-ray observations revealed that in fields around high-z galaxy clusters the X-ray source surface density is higher than in the field (see e.g. Henry & Briel 1991; Cappi et al. 2001; Cappelluti et al. 2005; Johnson et al. 2003; Ruderman & Ebeling 2005; Branchesi et al. 2007). In particular by studying a sample of 10 Chandra high-z clusters, Cappelluti et al. (2005) measured that 40% of the cluster fields show a factor $>$2 overdensities of X-ray sources when compared with field observations (see Fig. 1.2). As shown in Fig. 1.3 there are indications that the number density of AGN in/around clusters increases with redshift (Cappelluti et al. 2005; Branchesi et al. 2007), recent studies reveal that this increase is larger than what is observed in the field (Eastman et al. 2007). Ruderman & Ebeling (2005) studied the spatial distribution of a sample of 508 soft X-ray sources detected around 51 *Chandra* massive clusters in the redshift range 0.3-0.7. As in the works mentioned above, the X-ray source surface density in their sample is higher in the inner projected 3.5 Mpc than in the field. They also measured the AGN density profile in the cluster and, surprisingly, detected features in the curve. They detected a strong excess in the inner 0.5 Mpc mainly attributable to the central cD galaxies, a depletion zone around 1.5 Mpc and a secondary excess above 3 Mpc from the cluster center. The secondary excess has been explained with merger-driven accre-

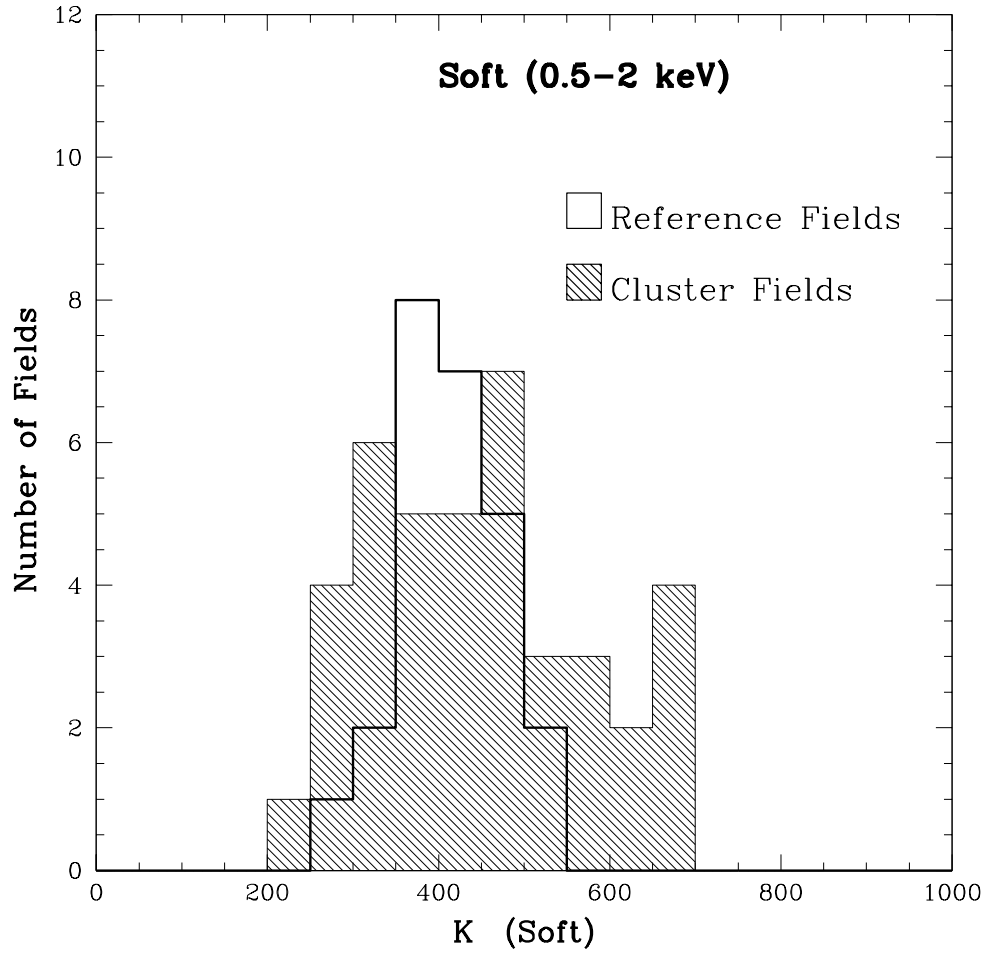

Figure 1.2: The distribution of the source surface density in the region of galaxy clusters and in the field as measured by Cappelluti et al. (2007).

tion onto SMBH at the edge of the cluster. This kind of triggering occurs most likely in low energy collisions favoured in the cluster-field transition region. The depletion zone is explained by the low probability of galaxy merging in regions of high velocity dispertion. Unfortunately, all the X-ray results mentioned above suffer from a lack of spectroscopic counterparts, they therefore need more detailed optical/IR follow-up campaigns to clarify the origin of the overdensities.

These X-ray sources in galaxy clusters typically have luminosities consistent with AGN or powerful starbursts. This suggests the presence of a remarkable population of obscured, or at least optically faint AGN in galaxy clusters. As mentioned above, optical observations are not able to detect low luminosity or

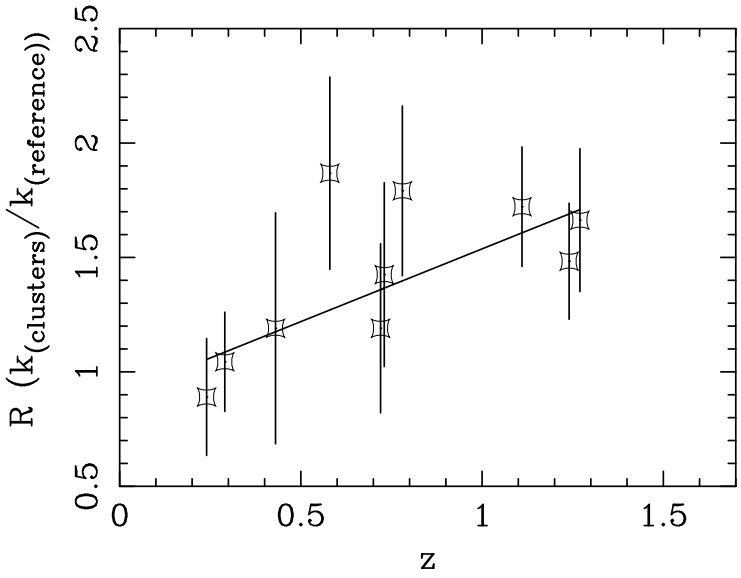

Figure 1.3: The X-ray source overdensity above the X-ray background in cluster field as a function of the redshift (Cappelluti et al. 2005).

obscured AGN, they therefore miss a significant fraction of AGN in clusters. As a confirmation of this, IR observations revealed that the AGN fraction in clusters is 5 times higher that what is observed in optical surveys (Martini et al. 2006). This fraction of AGN is however still too low to explain the amplitude of the X-ray source overdensity. More detailed IR observations suggested that some clusters show an AGN fraction on the order of 15% (Martini et al. 2007). As mentioned above, clustering studies revealed that AGN are very frequent in low-density galaxy groups. Indeed Jeltema et al. (2006) pointed out that intermediate redshift groups host at least one AGN and many show a higher fraction of active galaxies. Popesso & Biviano (2006) found an interesting anticorrelation of the AGN fraction in clusters and groups and the velocity dispersion of galaxies.

Many authors tried to use simulations to link the environment and the probability that a galaxy develops star formation or AGN activity. The conclusions are not yet definitive because of the large number of parameters that enter the problem. X-ray emission from galaxies is generated by different mechanisms. High luminosity emission is generally due to accretion onto SMBH, low luminosity emission (i.e. $L_X < 10^{42}$ erg/s) is generally associated to starformation activity. In the last case this is due a high number of X-ray binaries, young supernova remnants and hot plasma in starforming regions. Evrard (1991) modeled the interaction of the intercluster and interstellar

medium of galaxies (ICM & ISM) falling into the potential well of clusters. The ISM is expected to be compressed by the external pressure of the ICM which triggers star formation in the galaxy. The same ICM ram pressure strips the cold gas from the galaxy, which loses the reservoir to fuel AGN activity and star formation. The merging of groups and clusters is supposed to be responsible for starburst and AGN activity in clusters. N-Body simulations show that the rapid change of the tidal gravitational field during merging allows galaxies to exchange high quantities of cold gas to fuel the nuclear activity.

Numerical simulations by Moore et al. (1996) suggest that the galaxy falling into the cluster have a low probability of merging because of their high velocities. On the hand other successive high speed encounters can efficiently transfer cold gas. This scenario is called galaxy "harassment". AGN have been also taken into account to explain the quenching of star formation and the reddening of galaxies in clusters.

1.5 Open questions on the XRB

The discovery of the XRB opened an extraordinary window on the most energetic phenomena in the universe like the supermassive black holes in galaxy nuclei. Since the XRB is the result of all the accretion power in our Universe through the cosmic times, it is an unique tool to probe the evolution of black holes and cosmology.

- The number density of AGN varied in the history of Universe, we know that very luminous AGN (i.e. $L_{bol} > 10^{47}$ erg/s) were more frequent in the early Universe (i.e. $z > 2$) than now, the origin of this "cosmic downsizing" is still under debate.

- The presence of supermassive black hole in center of most of the galaxies rose the question on how star formation and gas accretion are related and what is their relative feedback. Which were the seeds of SMBHs?

- With the study of AGN clustering we now know that AGN are biased tracers of the matter distribution in the universe, the open question is what kind of matter overdensities host AGN and what is the relation

between the large scale environment and the triggering of AGN activity.

- AGN probe the high-z universe and therefore the first phases of life of galaxies.
- As mentioned above the majority of the soft X-ray background has been resolved into discrete sources while, the nature of the unresolved hard XRB is still matter of debate. Models predict that the high energy XRB is dominated by highly absorbed AGN which are difficult to detect also for the most advanced X-ray telescopes. These objects are still elusive and require wide research efforts to understand their nature.
- About 5% of the soft X-ray background is still unresolved, and there are evidences that this fraction could hide the thermal emission of the so called ”missing barions” which are supposed to live in the warm hot intergalactic medium (WHIM).
- Part of the XRB emission comes for galaxy clusters; the knowledge of their mass function and their number density in the universe are the basis to constrain and improve the cosmological parameters, especially the parameters of the elusive dark energy.

1.6 Overview of the thesis

In this thesis I focus, with an observational approach, on the relation between X-ray selected AGN and the environment in which they reside.

- In Chapter 2 I describe the XMM-COSMOS wide field survey, the data analysis with a particular attention to the source detection technique, the production of the selection function, the logN-logS measurement and a detailed analysis of the cosmic variance. The contribution of the XMM-COSMOS sources to the flux of the XRB will be discussed and the logN-logS relation will be compared with the expectations from population synthesis models.
- In Chapter 3 I focus on the measurement of the angular correlation function of AGN together with a detailed description of the technique. The deprojection of the 2D to the 3D correlation are discussed. The results

are compared with previous works and an I characterize the environment of AGN with particular attention to the characterization of the DM haloes in which they reside.

- In Chapter 4 I present the analysis of the cluster-AGN cross-correlation function in the ROSAT-NEP survey. Together with an exhaustive description of the technical methods I will try to interpretate the results in comparison with results on galaxies.

- In Chapther 5 I show the results of the measurement of the cross-correlation function cluster/groups-AGN making use of multiwavelength data in the COSMOS field. Given the good statistics, the data permit an interpretation of the results with the linear theory of biasing.

This work is part of COSMOS (Cosmic evolution survey) project based on an ACS HST treasury program. The survey involves more than 100 scientists all over the world and makes use of most advanced multiwavelength observing facilities. X-ray observations have been first performed with the XMM-Newton and later integrated by *Chandra*. The contents of Chapthers 2,3,4 have been already published on The Astrophysical Journal in September 2007, and on Astronomy and Astrophysics in January 2007.

Chapter 2

The XMM-COSMOS survey: source counts and cosmic variance

2.1 Introduction

The source content of the X–ray sky has been investigated over a broad range of fluxes and solid angles thanks to a large number of deep and wide surveys performed in the last few years using *ROSAT*, *Chandra* and XMM–*Newton* (see Brandt & Hasinger 2005 for a review). Follow–up observations unambiguously indicate that Active Galactic Nuclei (AGN), many of which are obscured, dominate the global energy output recorded in the cosmic X–ray background. The impressive amount of X–ray and multi-wavelength data obtained to date have opened up the quantitative study of the demography and evolution of accretion driven Supermassive Black Holes (SMBHs; Miyaji et al. 2000; Hasinger et al. 2005; Ueda et al. 2003; La Franca et al. 2005). At present the two deepest X–ray surveys, the *Chandra* Deep Field North (CDFN; Bauer et al. 2004) and *Chandra* Deep Field South (CDFS; Giacconi et al. 2001), have extended the sensitivity by about two orders of magnitude in all bands with respect to previous surveys (Hasinger et al. 1993; Ueda et al. 1999; Giommi et al. 2000), detecting a large number of faint X–ray sources. However, deep pencil beam surveys are limited by the area which can be covered to very faint fluxes (typically of the order of 0.1 deg^2) and suffer from significant field to field variance. In order to cope with such limitations, shallower surveys over larger areas have been undertaken in the last few years with both *Chandra*

(e.g. the 9 deg^2 Bootes survey (Murray et al. 2005), the Extended Groth strip EGS (Nandra et al. 2005), the Extended *Chandra* Deep Field South E-CDFS, (Lehmer et al. 2005; Virani et al. 2006) and the Champ (Green et al. 2004; Kim et al. 2004)) and XMM–*Newton* (e.g. the HELLAS2XMM survey (Fiore et al. 2003), the XMM–*Newton* BSS (Della Ceca et al. 2004) and the ELAIS S1 survey (Puccetti et al. 2006)).

In this context the XMM–*Newton* wide field survey in the COSMOS field (Scoville et al. 2007), hereinafter XMM–COSMOS (Hasinger et al. 2007), has been conceived and designed to maximize the sensitivity and survey area product, and is expected to provide a major step forward toward a complete characterization of the physical properties of X–ray emitting SMBHs. A contiguous area of about 2 deg^2 will be covered by 25 individual pointings, repeated twice, for a total exposure time of about 60 ksec in each field. In the first observing run obtained in AO3 (phase A), the pointings were disposed on a 5x5 grid with the aimpoints shifted of 15' each other, so as to produce a contiguous pattern of coverage. In the second run, to be observed in AO4 (phase B), the same pattern will be repeated with each pointing shifted by 1' with respect to phase A. The above described approach ensures a uniform and relatively deep coverage of more than 1 deg^2 in the central part of the field. When completed, XMM–COSMOS will provide an unprecedentedly large sample of about 2000 X–ray sources with full multi-wavelength photometric coverage and a high level of spectroscopic completeness. As a consequence, the XMM–COSMOS survey is particularly well suited to address AGN evolution in the context of the Large Scale Structure in which they reside. More specifically, it will be possible to investigate if obscured AGN are biased tracers of the cosmic web and whether their space density rises in the proximity of galaxy clusters (Henry & Briel 1991; Cappi et al. 2001; Gilli et al. 2003; Johnson et al. 2003; Yang et al. 2003; Cappelluti et al. 2005; Ruderman & Ebeling 2005; Miyaji et al. 2007; Yang et al. 2006; Cappelluti et al. 2006).

The X–ray reduction of phase A data along with a detailed analysis of the source counts in different energy bands are presented in this paper which is organized as follows. In Section 2 the data reduction procedure and the relative astrometric corrections are described. In Section 3 the source detection algorithms and technique are discussed. Monte Carlo simulations are presented in Section 4. The logN–logS relations and the analysis of the contribution of the XMM-COSMOS sources to the X-ray background are discussed in Section 5. The study of sample variance is presented in Section 6 and a

summary of the work is reported in Section 7. The strategy and the log of the observations of XMM–COSMOS are presented by Hasinger et al. (2007), the optical identifications of X-ray sources by Brusa et al. (2007), the analysis of groups and clusters by Finoguenov et al. (2007), the spectral analysis of a subsample of bright sources by Mainieri et al. (2007) and the clustering of X-ray extragalactic sources by Miyaji et al. (2007). Throughout the paper the concordance WMAP ΛCDM cosmology (Spergel et al. 2003) is adopted with H_0=70 km s^{-1} Mpc^{-1}, Ω_Λ=0.7 and Ω_m=0.3

2.2 EPIC Data cleaning

The EPIC data were processed using the XMM–*Newton* Standard Analysis System (hereinafter SAS) version 6.5.0. The Observational Data Files (ODF, "raw data") of each of the 25 observations, were calibrated using the SAS tools *epchain* and *emchain* with the most recent calibration data files. Events in bad columns, bad pixels and close to the chip gaps were excluded.
Both the EPIC PN and MOS event files were searched for high particle background intervals. The distribution of the background counts binned in 100 s intervals was obtained in the 12–14 keV band for the PN and in the 10–12 keV band for the MOS, which are dominated by particle background, and then fitted with a gaussian model. All time intervals with background count rate higher than 3σ above the average best fit value were discarded. In Fig. 2.1 an example of the application of this method to Field #6 is shown. Once the high energy flares were removed, the 0.3–10 keV background counts distribution was processed, with the same 3σ clipping method, in order to remove times during which low energy particle flares were important. These flares are not easily detected in the 12–14 keV band. As a result of this selection process the average time lost due to particle flares was $<$20% and 2 observations were completely lost (see Hasinger et al. 2007).
An important feature observable in the background spectrum of both MOS and PN CCDs is the Al Kα (1.48 keV) fluorescent emission. In the PN background two strong Cu lines are also present at $\sim$ 7.4 keV and $\sim$ 8.0 keV. Since these emission lines could affect the scientific results, the 7.2–7.6 keV and 7.8–8.2 keV energy bands in the PN and the 1.45–1.54 keV band (in PN and MOS) were excluded from the detectors events. Images were then created in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV energy bands with a pixel size of 4 arcsec. Single and double events were used to construct the PN images, while

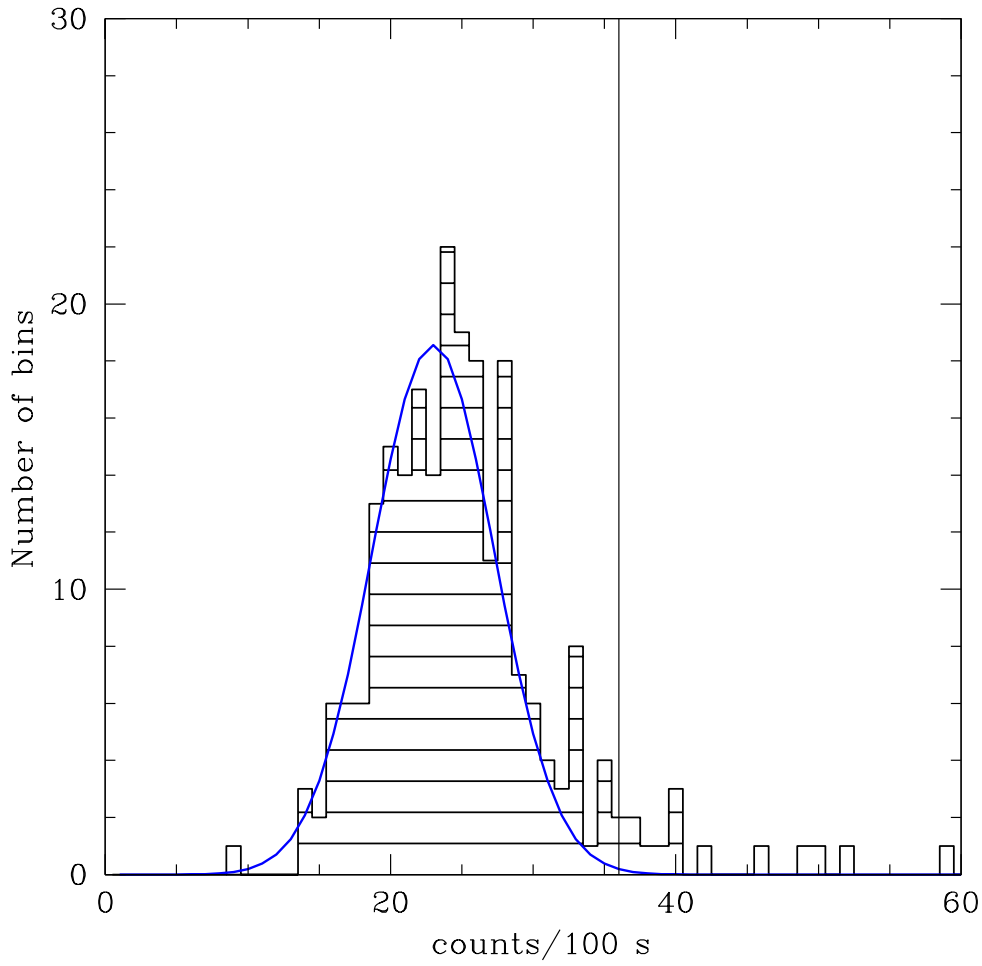

Figure 2.1: The background counts distribution in the PN observation of Field 6. The solid line represents the best gaussian fit to the distribution. The continuous vertical line represents the adopted 3σ cut above which the corresponding time intervals have been discarded.

MOS images were created using all valid event patterns. Out-of-Time (OOT) events appear when a photon hits the CCD during the read-out process in the IMAGING mode. The result is that the x position of the event/photon is known, while the y position is unknown due to the readout and shifting of the charges at this time. For this reason artificial OOT event files were created. A new y coordinate is simulated by randomly shifting the event along the read-out axis and performing the gain and CTI (charge transfer inefficiency) correction afterwards. For the PN, in full frame mode the OOT events constitute about 6.3% of the observing time. Those files were filtered in the same way as

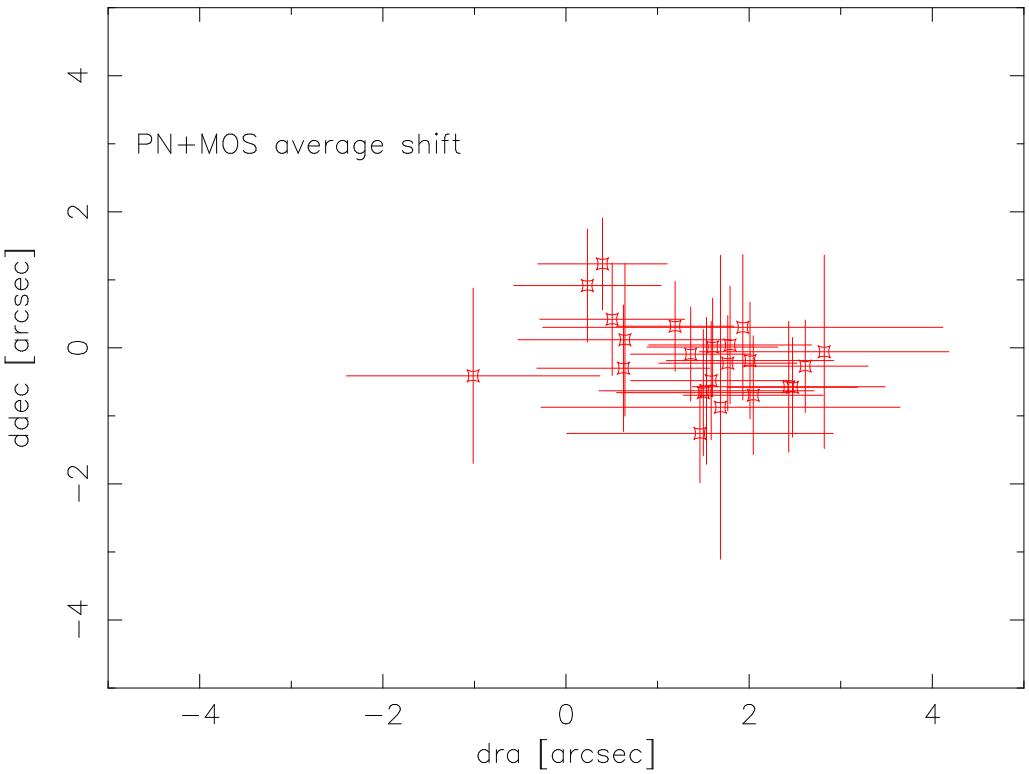

Figure 2.2: Shift between the PN+MOS mosaic and the MEGACAM catalog for each pointing of the XMM-*Newton* COSMOS field.

the event files and the produced images were subtracted from the event images. Images were then added in order to obtain PN+MOS mosaics. For each instrument and for each observation, spectrally weighted exposure maps were created using the SAS task *eexpmap*, assuming a power law model with photon index Γ=2.0 in the 0.5–2 keV band and Γ=1.7 in the 2–4.5 and 4.5–10 keV bands.

2.2.1 Astrometry correction

In order to correct the astrometry of our XMM–*Newton* observations for each pointing and for each instrument, the produced X-ray source list (see next section) was compared with the MEGACAM catalog of the COSMOS field (Mc Cracken et al. 2007) including all the sources with I magnitudes in the range 18-23. In order to find the shift between the two catalogues, an optical-X-ray positional correlation was computed using the likelihood algorithm included in the SAS task *eposcorr*. This task uses in a purely statistical way all possible counterparts of an X-ray source in the field to determine the most

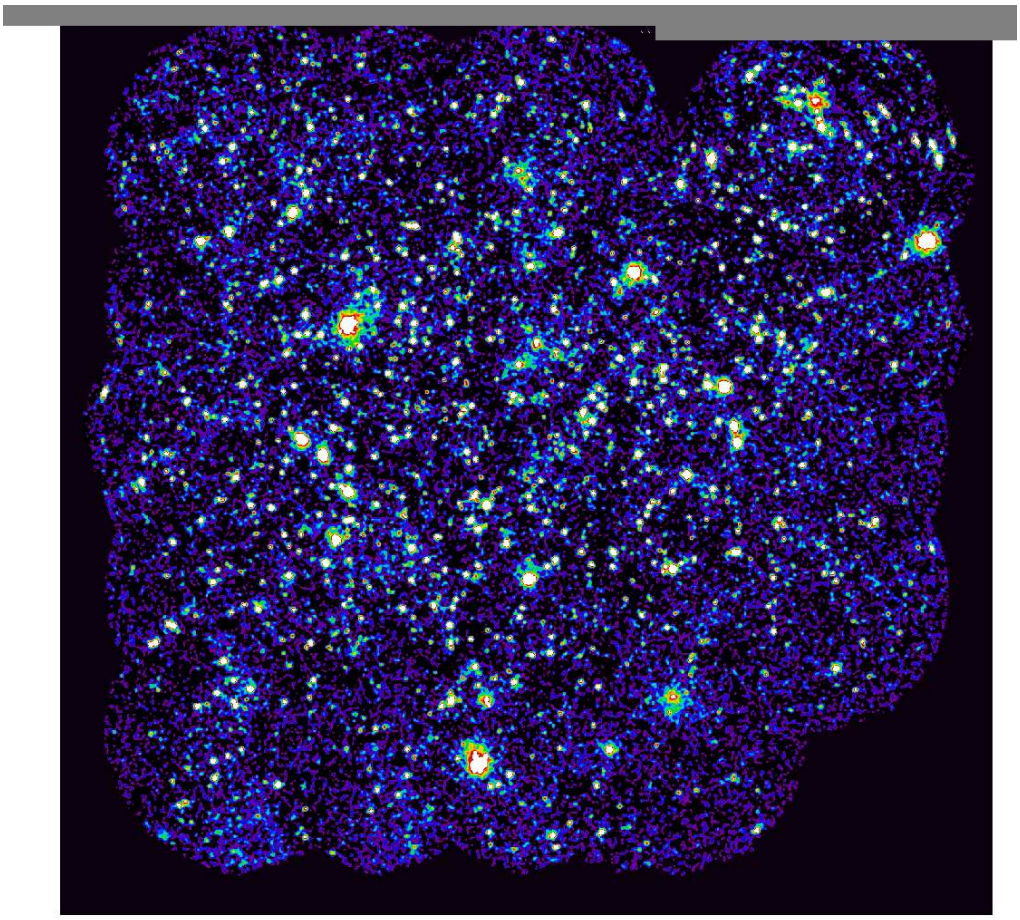

Figure 2.3: Signal–to–noise ratio map in the 0.5–2 keV band of the XMM–*Newton* raster scan in the COSMOS field. The stretch of the color map corresponds to [-0.1<S/N<1 per pixel]. The scale has been chosen to enhance the SNR contrasts. If S is the raw (sources + background) image and B is the model background image, then the SNR map is obtained by $SNR = \frac{S-B}{\sqrt{S}}$. The image was smoothed with a gaussian filter with σ=2 pixels. Negative values are places where the measured background is smaller than the model background.

likely coordinate displacement. This method is independent of actual spectroscopic identifications, but all post-facto checks using, for example, secure spectroscopic identifications, have demonstrated its reliability and accuracy. Using the magnitude range mentioned above, systematic effects introduced by bright stars and faint background objects are minimized. In the majority of the observations the shift between the three cameras turned out to be <

Energy Band keV	Γ	ECF cts $s^{-1}/10^{-11}$ erg cm^{-2} s^{-1}	S_{lim} erg cm^{-2} s^{-1}	N sources All	N sources Point-like	Single detection
0.5–2	2.0	10.45	7.2×10^{-16}	1307	1281	661
2–4.5	1.7	1.52	4.0×10^{-15}	735	724	89
4.5–10	1.7	1.21	9.7×10^{-15}	187	186	3

Table 2.1: Summary of the source detection results. From left to right, energy band, average ecf, limit flux, number of detected sources, number of pointlike sources number of single band detections.

1" (i.e. much smaller than the pixel size of the images used here). Since the shift between the EPIC cameras is negligible, a correlation between the joint MOS+PN source list and the optical catalog was calculated to derive the astrometric correction. For the 23 pointings presented here, the shifts between the optical and X-ray catalog are never larger than 3", with an average shift of the order of $\Delta\alpha \sim$1.4" and $\Delta\delta \sim$-0.17". The average displacement in the two coordinates between the PN+MOS mosaic X-ray positions and the MEGACAM catalog sources for each pointing of the XMM-COSMOS field is shown in Fig 2.2. The appropriate offset was applied to the event file of each pointing and images and exposure maps were then reproduced with the corrected astrometry.

2.3 EPIC source detection

2.3.1 Background modeling

In order to perform the source detection a sophisticated background modeling has been developed. In X-ray observations the background is mainly due to two components, one generated by undetected faint sources contributing to the cosmic X-ray background and one arising from soft protons trapped by the terrestrial magnetic field. For this reason two background templates were computed for each instrument and for each pointing, one for the sky (vignetted) background (Lumb et al. 2002) and one for instrumental and particles background (unvignetted). To calculate the normalizations of each template of every pointing, we first performed a wavelet source detection (see Finoguenov et al. 2007) without sophisticated background subtraction, then

we excised the areas of the detector where a significant signal due to sources was detected. The residual area is split into two parts depending on the value of the effective exposure (i.e. higher and lower than the median value). Using the two templates, we calculate the coefficients of a system of two linear equations from which we obtain the normalizations of both:

$$AM_v^1 + BM_{unv}^1 = C^1 \tag{2.1}$$

$$AM_v^2 + BM_{unv}^2 = C^2 \tag{2.2}$$

where A and B are the normalization factors, $M_v^{1,2}$ the vignetted templates in the region with effective exposure higher and lower than the median , $M_{unv}^{1,2}$ the unvignetted templates and $C^{1,2}$ are the background counts in the two regions. The region with effective exposure lower than the median (i.e. high vignetting, $>$7' off-axis) is dominated by the instrumental background, while the region with higher effective exposure is dominated by the sky background. Therefore with this method we have the advantage of better fitting the two components of the background. The standard method for estimating the background, based on the spline functions used in the XMM–*Newton* pipelines, returned in our case significant residuals. The excellent result of this technique can be seen in the signal-to-noise (SNR) map in Fig 2.3: despite the significant variations in exposure time and average background level from pointing to pointing, a rather homogeneous signal–to–noise ratio is achieved across the whole mosaic. It is worth noting that also pixels with negative values are shown in the map; these are located where the background model is higher than the measured background.

2.3.2 Maximum likelihood detection

In each pointing the source detection was conducted on the combined images of the different instruments in the three energy bands mentioned above using the SAS tasks *eboxdetect* and *emldetect*. As a first step, the sliding cell detection algorithm *eboxdetect* was run on the images in the three energy bands. In this procedure source counts were collected in cells of 5$\times$5 pixels adopting a low threshold in the detection likelihood (i.e. likemin=4). The source list produced by *eboxdetect* was then used as input for *emldetect*. For all the sources detected with the sliding cell method this task performs a maximum likelihood PSF fit. In this way refined positions and fluxes for the sources were determined. Due to the particular pattern of our observations (see Hasinger

et al. 2007), the same source could be detected in up to 4 different pointings. For this reason both *eboxdetect* and *emldetect* were run in raster mode. The source parameters (position and flux) were fitted simultaneously on all the observations where the source is observable, taking into account the PSF at the source position in each pointing. As likelihood threshold for the detection, we adopted the value det_ml=6. This parameter is related to the probability of a random Poissonian fluctuation having caused the observed source counts:

$$det\_ml = -lnP_{random} \tag{2.3}$$

In principle, the expected number of spurious sources could be estimated as the product of the probability for a random Poisson fluctuation exceeding the likelihood threshold times the number of statistically independent trials, N_{trial}. For a simple box detection algorithm N_{trial} would be approximately given by the number of independent source detection cells across the field of view. For the complex multi-stage source detection algorithm, like the one applied here, N_{trial} cannot be calculated analytically, but has to be estimated through Monte Carlo simulations. These simulations, which are discussed in Section 4, return a number of spurious sources of $\sim$2% at the likelihood level chosen. All the sources were fitted with a PSF template convolved with a beta model (Cavaliere & Fusco-Femiano 1976). Sources which have a core radius significantly larger than the PSF are flagged as extended (*ext* parameter > 0).

A total of 1307, 735 and 187 X-ray sources were detected in the three bands. Of these sources, twenty-six were classified as extended. The analysis of the X-ray extended sources in the COSMOS field is beyond the scope of this work; these sources are extensively discussed by Finoguenov et al. (2007). A total of 1281, 724 and 186 point-like sources were detected in the three bands down to limiting fluxes of 7.2×10^{-16} erg cm^{-2} s^{-1}, 4.7×10^{-15} erg cm^{-2} s^{-1} and 9.7×10^{-15} erg cm^{-2} s^{-1} respectively. The minimum number of net counts for the detected sources is $\sim$21, 17 and 27 in the three bands, respectively. A total of 1390 independent point-like sources have been detected by summing the number of sources detected in each band but not in any softer energy band. The number of sources detected only in the 0.5–2 keV, 2–4.5 keV, 4.5–10 keV bands are 661, 89 and 3, respectively.

From the count rates in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands the fluxes were obtained in the 0.5–2 keV, 2–10 keV and 5–10 keV bands respectively using the energy conversion factors (ECF) listed in Table 2.1, together with a summary of the source detection. The ECF values have been

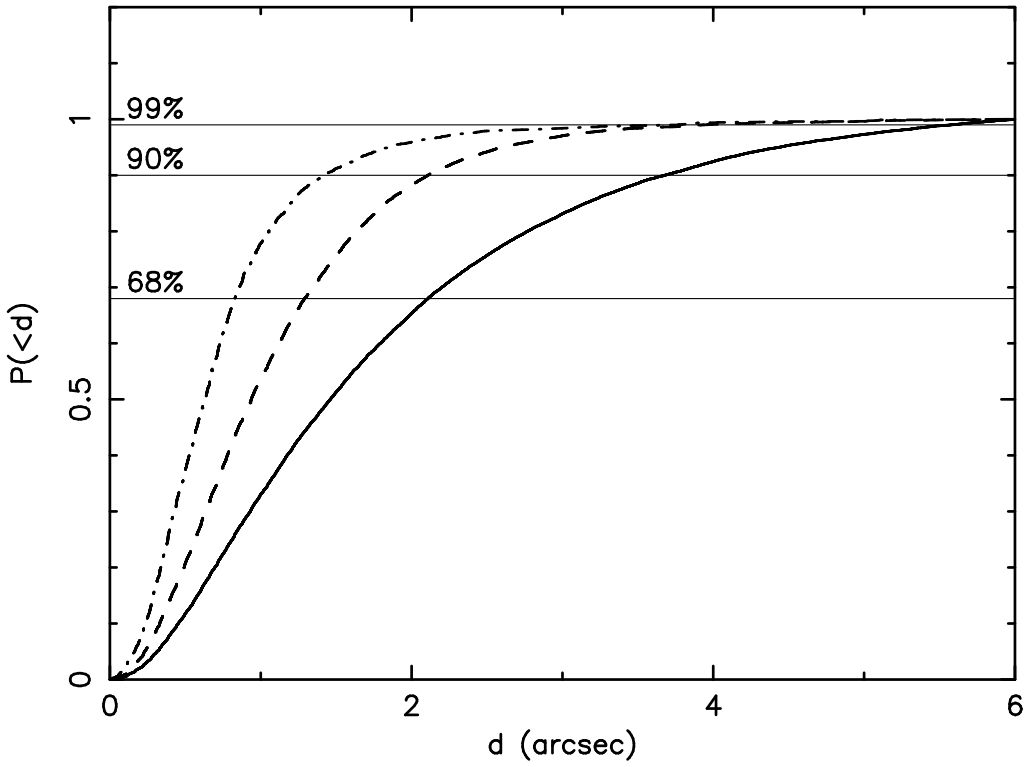

Figure 2.4: The cumulative probability to detect a true source with det_ml>6 in a circle of a given radius in the 0.5–2 (*continuous line*), 2–4.5 (*dashed line*) and 4.5–10 keV (*dashed – dotted line*) energy bands. The 68%, 90% and 99% levels are plotted as horizontal lines.

computed using the most recent EPIC response matrices in the corresponding energy ranges. As a model, we assumed power-law spectra with N_H =2.6$\times 10^{20}$ cm^{-2}, (corresponding to the average value of N_H over the whole COSMOS field (Dickey & Lockman 1990)) and the same spectral indices used to compute the exposure maps without considering any intrinsic absorption (see Table 2.1). It is worth noting that the spectral indices and the absorptions of the individual sources can be significantly different from the average values assumed here. In particular Mainieri et al. (2007) found that the spectral indices Γ of the XMM-COSMOS sources are in the range 1.5$\div$2.5, in the CDFS Tozzi et al. (2006) measured an average photon index $<\Gamma>\sim$1.75 and similar values were obtained by Kim et al. (2004) in the CHAMP survey. The mean spectrum assumed here is therefore consistent with the values measured up to now. By changing the spectral index of $\pm$0.3 the ECFs change of 2%, 12% and 4% in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV, respectively.

2.4 Monte Carlo simulations

In order to properly estimate the source detection efficiency and biases, detailed Monte Carlo simulations were performed (see, e.g., Hasinger et al. 1993; Loaring et al. 2005). Twenty series of 23 XMM–*Newton* images were created with the same pattern, exposure maps and background levels as the real data. The PSF of the simulated sources was constructed from the templates available in the XMM–*Newton* calibration database. The sources were randomly placed in the field of view according to a standard 0.5-2 keV logN-logS distribution (Hasinger et al. 2005). This was then converted to a 2–4.5 keV and 4.5–10 keV logN-logS assuming that all the sources have the same intrinsic spectrum (a power-law with spectral index $\Gamma = 1.7$). We then applied, to the simulated fields, the same source detection procedure used in the real data. Schmitt & Maccacaro (1986) showed that with the threshold adopted here for source detection, which corresponds roughly to the Gaussian 4.5-5σ, the distortion of the slope of the logN-logS due to Poissonian noise is $<$3% for a wide range of slopes. Therefore, the uncertainties introduced by using a single logN-logS as base for the simulations are negligible. A total of 30626, 13579 and 3172 simulated sources were detected in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands down to the same limiting fluxes of the observations. For every possible pair of input-output sources we computed the quantity

$$R^2 = \left(\frac{x-x_0}{\sigma_x}\right)^2 + \left(\frac{y-y_0}{\sigma_y}\right)^2 + \left(\frac{S-S_0}{\sigma_S}\right)^2, \tag{2.4}$$

where x,y and S are the position and flux of the detected source and x_0,y_0 and S_0 are the corresponding values for all the simulated sources We then flag as the most likely associations those with the minimum value of R^2. The distribution of the positional offsets is plotted in Fig. 2.4 for each energy band analyzed.

We find that 68% of the sources are detected within 2.1", 1.3" and 0.8" in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands, respectively. Since the detection software fits the position of the source using the information available for the three bands together, we expect to be able to detect sources with an accuracy of the order of, or somewhat better than, that shown in Fig. 2.4. As in Loaring et al. (2005), we then define a cut-off radius r_{cut} of 6". Sources with a displacement larger than r_{cut} from their input counterpart are classified as spurious. These account for 2.7%, 0.5%, 0.6% of the total number of sources in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands, respectively. Source confu-

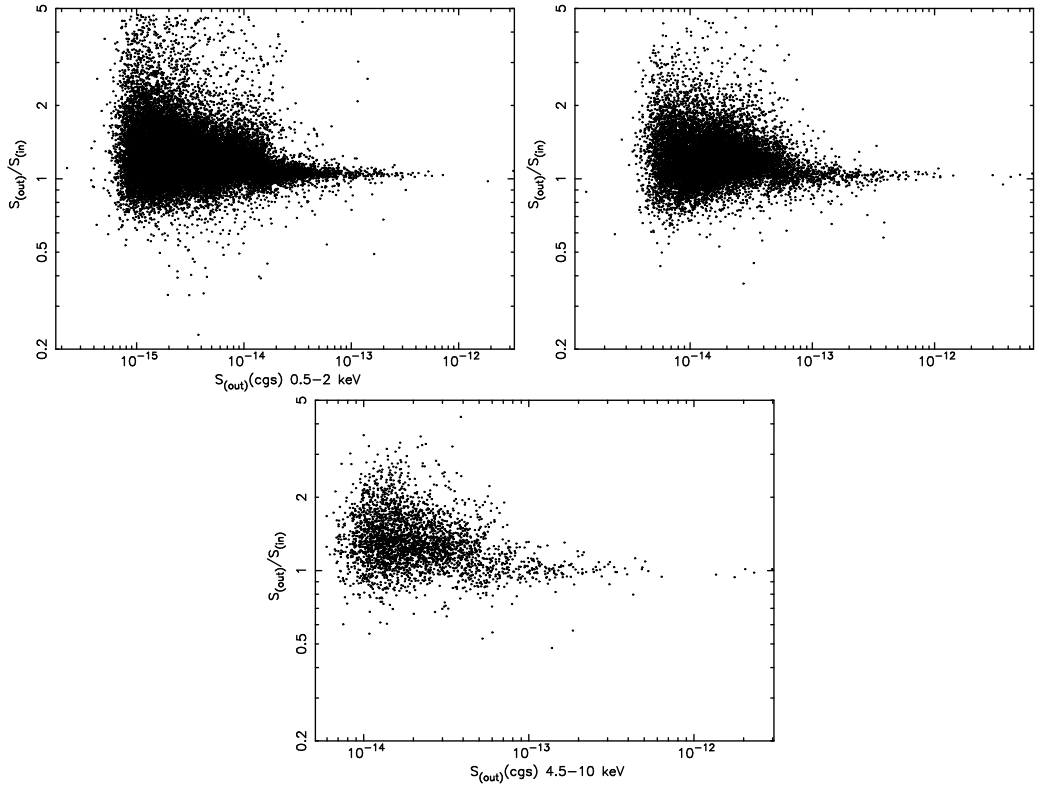

Figure 2.5: The ratio S_{out}/S_{in} as a function of the output detected flux in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands, respectively*Top middle and bottom panel*.

sion occurs when two or more sources fall in a single resolution element of the detector and result as a single detected source with an amplified flux. In order to determine the influence of the source confusion we adopted the method described in Hasinger et al. (1998). We define as "*confused*" sources those for which $S_{out}/(S_{in}+3*\sigma_{out})>1.5$ (where σ_{out} is the 1σ error on the output flux). The fraction of "*confused*" sources is 0.8%, 0.15% and <0.1% in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV bands, respectively.
The photometry was also tested; the ratio of output to input fluxes in the simulation is plotted in Fig. 2.5.

At bright fluxes this ratio is consistent with one, while at fainter fluxes the distribution of S_{out}/S_{in} becomes wider, mainly because of increasing errors, and skewed toward values greater than one. This skewness of the distribution can be explained mainly by two effects: a) source confusion and b) Eddington Bias (Eddington 1940). While source confusion, as defined above, affects only a small fraction of the sources, the Eddington bias results in a systematic

upward offset of the detected flux. The magnitude of this effect depends on the shape of the logN-logS distribution and the statistical error on the measured flux. Since there are many more faint than bright sources, uncertainties in the measured flux will result in more sources being up-scattered than down-scattered. Together with this, the fact that in the 4.5–10 keV band we are sampling a flux region in which the logN-logS is steeper than in the other bands (see Section 5), explains why such an effect is more evident in the 4.5–10 keV band.

Besides assessing the reliability of our source detection procedure, one of the aims of these simulations is to provide a precise estimation of the completeness function of our survey, known also as sky coverage. We constructed our sky-coverage (Ω) vs. flux relation by dividing the number of detected sources by the number of input sources as a function of the flux and rescaling it to the sky simulated area. Having analyzed the simulations with the same procedure adopted for the analysis of the data, this method ensures that when computing the source counts distribution (see next section) all the observational biases are taken into account and corrected. The Ω vs. flux relation relative to the 0.5–2 keV, 2–10 keV and 5–10 keV bands is plotted in fig. 2.6. The total sky area is 2.03 deg^2 and it is completely observable down to fluxes of $\sim$0.3, 1.3 and 2×10^{-14} erg cm^{-2} s^{-1} in the three bands, respectively. The sky coverage drops to 0 at limiting fluxes of $\sim7\times10^{-16}$ erg cm^{-2} s^{-1}, $\sim4\times10^{-15}$ erg cm^{-2} s^{-1}and $\sim9\times10^{-15}$ erg cm^{-2} s^{-1}, in the 0.5–2 keV, 2–10 keV and 5–10 keV bands, respectively.

2.5 Source counts

Once the sky coverage is known, the cumulative source number counts can be easily computed using the following equation:

$$N(>S) = \sum_{i=1}^{N_S} \frac{1}{\Omega_i} deg^{-2}, \tag{2.5}$$

where N_S is the total number of detected sources in the field with fluxes greater than S and Ω_i is the sky coverage associated with the flux of the ith source. The variance of the source number counts is therefore defined as:

$$\sigma_i^2 = \sum_{i=1}^{N_S} \left(\frac{1}{\Omega_i}\right)^2. \tag{2.6}$$

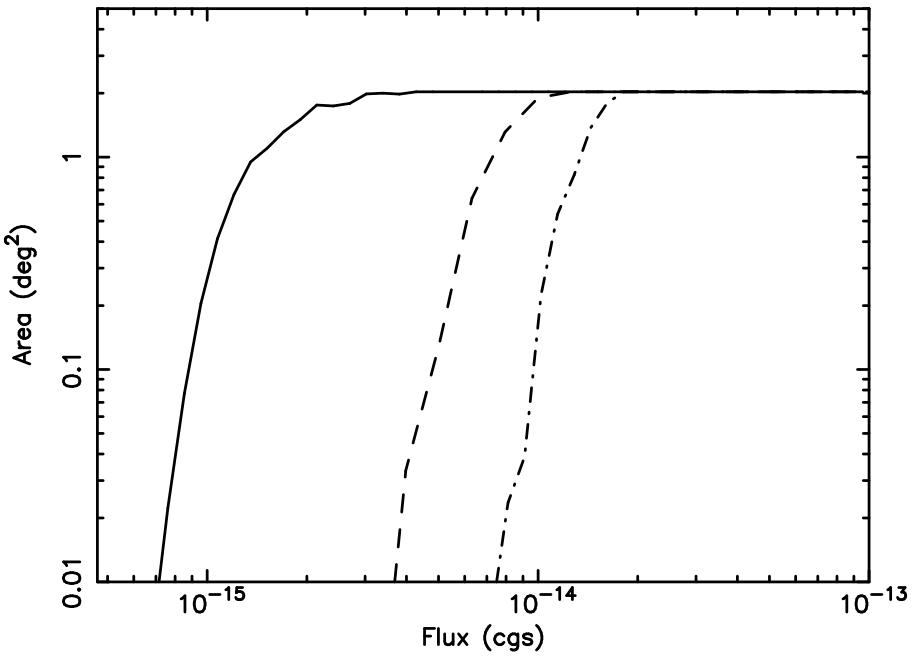

Figure 2.6: The sky coverage vs. flux relation in the 0.5-2, 2-10 and 5-10 keV energy bands (respectively continuous, dashed and dash-dotted line), resulting from the simulations described in the text.

Source number counts are reported in Table 2.2. The cumulative number counts, normalized to the Euclidean slope (multiplied by $S^{1.5}$), are shown in

Figures 2.7, 2.8 and 2.9, for the 0.5–2 keV, 2–10 keV and 5–10 keV energy ranges, respectively. With such a representation, the deviations from the Euclidean slope are clearly evident as well as the flattening of the counts towards faint fluxes. Source counts are compared with the findings of other deep and shallow surveys collected from the literature. The plotted reference results were selected in order to sample a flux range as wide as possible and at the same time to keep the plots as clear as possible. As discussed in the previous section, the sky coverage Ω was derived from realistic Monte Carlo simulations and therefore no further correction for the Eddington Bias is required.

In order to parameterize our relations, we performed a maximum likelihood fit to the unbinned differential counts. We assumed a broken power-law model for the 0.5–2 keV and 2–10 keV bands:

$$n(S) = \frac{dN}{ds} = \begin{cases} AS^{-\alpha_1} & S > S_b \\ BS^{-\alpha_2} & S \leq S_b, \end{cases} \tag{2.7}$$

log(S)	Ω	Counts
erg cm^{-2} s^{-1}	deg^2	deg^{-2}
0.5–2 keV		
-13.0	2.03	4.5±1.5
-13.5	2.03	18.8±3.1
-14.0	2.03	105.2±7.0
-14.5	2.03	327.0±12.7
-15.0	0.58	790±23.3
-15.1	0.12	931±53.0
2–10 keV		
-13.0	2.03	8.6±2.0
-13.5	2.03	57.0±5.3
-14.0	1.40	258.9±11.6
-14.3	0.13	600.1±34.2
5–10 keV		
-13.5	2.03	21.3±3.2
-14.0	0.35	111±11.0

Table 2.2: Cumulative number counts and Sky coverage.

where $A = BS_b^{\alpha_1-\alpha_2}$ is the normalization, α_1 is the bright end slope, α_2 the faint end slope, S_b the break flux, and S the flux in units of 10^{-14} erg cm^{-2} s^{-1} . Notice that using the maximum likelihood method, the fit is not dependent on the data binning and therefore we can make full use of the whole dataset. Moreover, the normalization A is not a parameter of the fit, but it is obtained by imposing the condition that the number of expected sources from the best fit model is equal to the total observed number. In the 0.5–2 keV energy band the best fit parameters are α_1=$2.60^{+0.15}_{-0.18}$, α_2=1.65±0.05, S_b=$1.55^{+0.28}_{-0.24}$ $\times 10^{-14}$ erg cm^{-2} s^{-1} and A=123. Translating this value of the normalization to that for the cumulative distribution at 2×10^{-15} erg cm^{-2} s^{-1} , which is usually used in the literature for *Chandra* surveys, we obtain A_{15} ∼450 which is fully consistent with most of previous works where a fit result is presented (Hasinger et al. 1993; Mushotzky et al. 2000; Hasinger et al. 2001; Baldi et al. 2002; Rosati et al. 2002; Bauer et al. 2004; Kim et al. 2004; Hasinger et al. 2005; Kenter et al. 2005), but significantly lower than that found in the CLASXS survey (Yang et al. 2004). In the 2–10 keV band the best fit param-

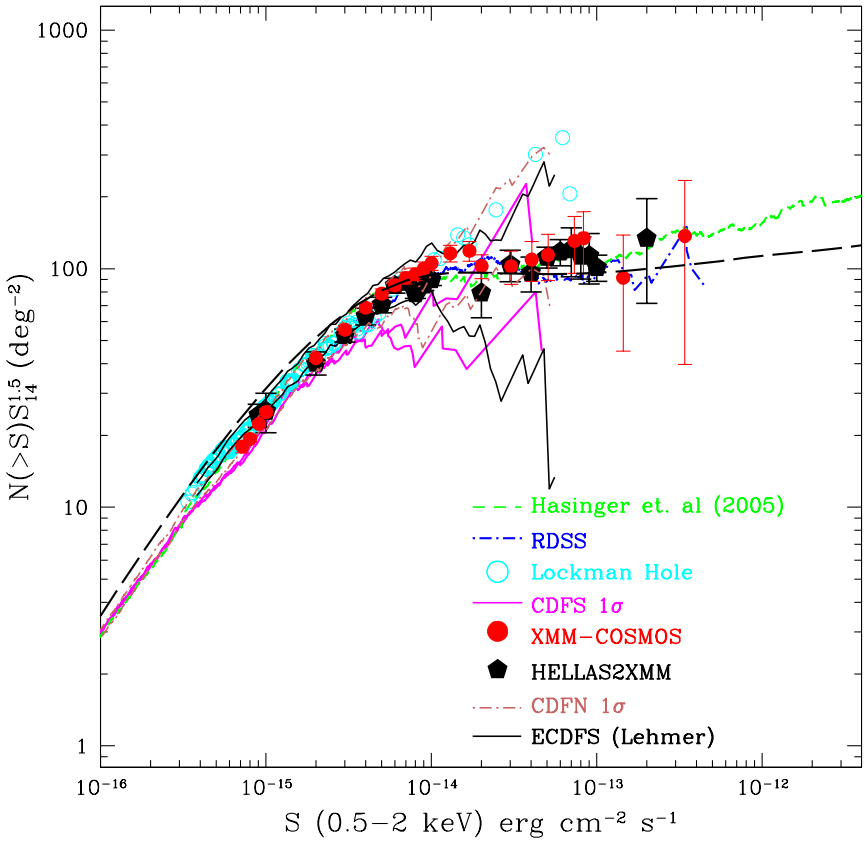

Figure 2.7: The 0.5-2 keV logN-logS of the XMM-COSMOS (*red dots*) sources compared with the ROSAT medium sensitivity survey (Hasinger et al. 1993) (*blue dot dashed line*), combined ROSAT, XMM–*Newton* , *Chandra* (Hasinger et al. 2005) (*green dashed line*), the *Chandra* deep field south 1σ error tie (Rosati et al. 2002) (*magenta continuous line*), the *Chandra* deep field north 1σ error tie (Bauer et al. 2004) (*pink dot − dashed line*), the 100 ks of the XMM–*Newton* Lockman hole (Hasinger et al. 2001) (*cyan circles*), the HELLAS2XMM (Baldi et al. 2002) (*black pentagons*) and the extended CDFS (Lehmer et al. 2005) (*black continuous line*) surveys. The overlayed black-dashed line represents the logN-logS predicted by the model of Gilli, Comastri & Hasinger (2006). The source number counts are plotted scaled by $S^{1.5}$ in order to highlight the deviation from the Euclidean behavior.

eters are α_1=2.43±0.10, α_2=1.59±0.33, S_b=1.02$^{+0.25}_{-0.19}$ ×10^{-14} erg cm^{-2} s^{-1} and A=266. The latest value translates into A_{15} ∼1250. Also in this band, our results are in agreement with previous surveys within 1σ, with the exception of

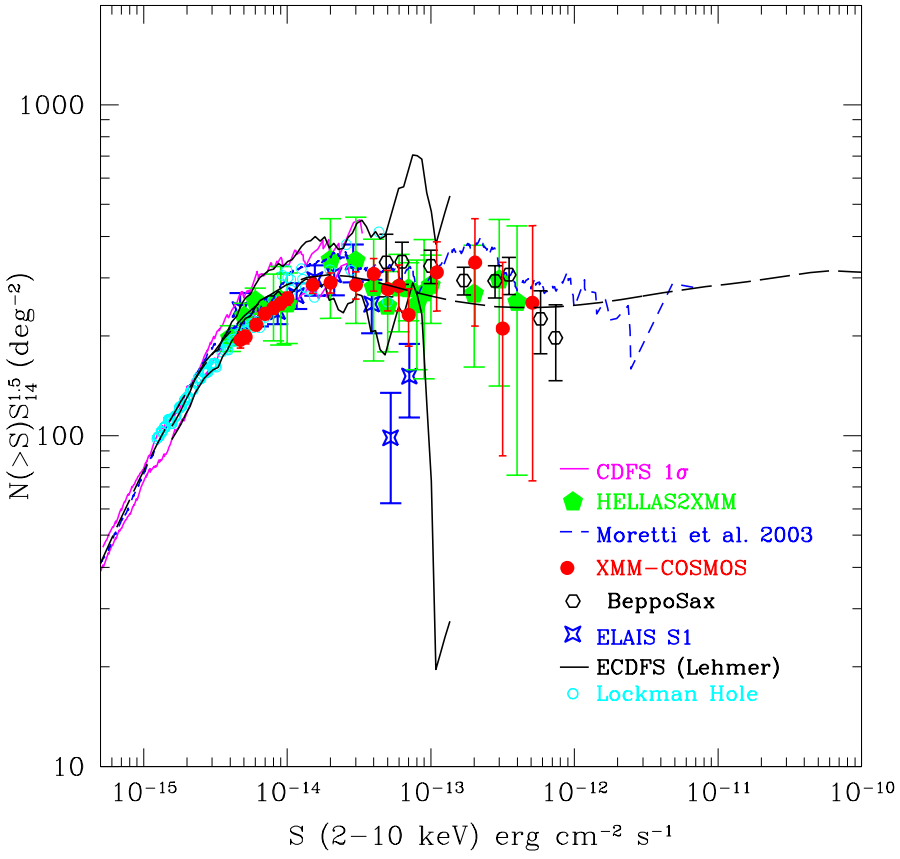

Figure 2.8: The 2-10 keV logN-logS of the XMM-COSMOS (*red dots*) sources compared with *Chandra*, XMM–*Newton* and ASCA (*blue dashed line*) (Moretti et al. 2003), HELLAS BeppoSAX (Giommi et al. 2000) (*black hexagons*) the Chandra deep field south 1σ error tie (Rosati et al. 2002) (*magenta continuous line*), the HELLAS2XMM (*green pentagons*) (Baldi et al. 2002), the ELAIS S1 (*blue stars*) (Puccetti et al. 2006), the extended CDFS (Lehmer et al. 2005) 1σ error tie (*black continuous line*) and the 100 ks of the Lockman hole (*cian open circles*) (Hasinger et al. 2001). The overlayed black-dashed line represents the logN-logS predicted by the model of Gilli, Comastri & Hasinger (2006). The source number counts are plotted scaled by $S^{1.5}$ in order to highlight the deviation from the Euclidean behavior.

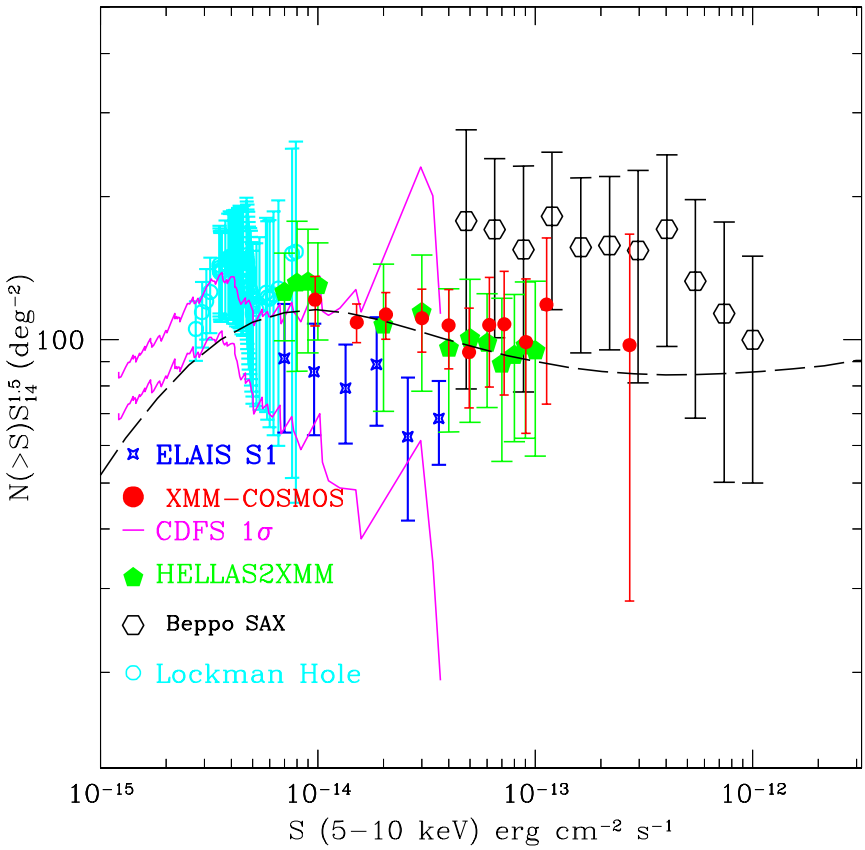

Figure 2.9: The 5-10 keV logN-logS of the XMM-COSMOS (*red dots*) sources compared with the HELLAS2XMM (Baldi et al. 2002) (*green pentagons*), the Chandra deep field south 1σ error tie (Rosati et al. 2002) (*magenta continuous line*) , the HELLAS-BeppoSAX data from Fiore et al. (2001) (*black hexagons*), the ELAIS S1 (*blue stars*) (Puccetti et al. 2006) and the 100 ks of the Lockman Hole (*cyan open circles*) (Hasinger et al. 2001). The overlayed black-dashed line represents the logN-logS predicted by the model of Gilli, Comastri & Hasinger (2006). The source number counts are plotted scaled by $S^{1.5}$ in order to highlight deviation from the Euclidean behavior.

the CLASXS survey which is ∼30% higher in this band (Yang et al. 2004). In the 5–10 keV energy bands, where the differential counts do not show any evidence for a break in the sampled flux range, we assumed a single power-law model of the form:

$$n(S) = \frac{dN}{ds} = AS^{-\alpha_1}, \tag{2.8}$$

for which the best fit parameters are found to be A=102 and α_1=2.36±0.1.

In the soft 0.5-2 keV (Fig 2.7) energy range a visual inspection of the various datasets suggests a remarkably good agreement between XMM-COSMOS and literature data[1].

In the 2–10 keV band the XMM-COSMOS counts bridge the gap between deep field observations (Rosati et al. 2002) and shallower large area BeppoSAX (Giommi et al. 2000) and XMM–*Newton* surveys (Baldi et al. 2002). At relatively bright fluxes ($> 10^{-14}$ erg cm^{-2} s^{-1}) the XMM-COSMOS logN–logS nicely matches previous measurements, though providing a much more robust estimate of the source counts thanks to the much smaller statistical errors.

A major step forward in the determination of X–ray source counts is achieved in the 5–10 keV band, where the previously existing data from different surveys show very significant differences. Thanks to XMM–COSMOS, a solid measure of the hard-X logN–logS in the flux interval 10^{-14} – 10^{-13} erg cm^{-2} s^{-1} is obtained for the first time. From Fig. 2.9 we notice that the normalization of the XMM–COSMOS logN–logS is slightly higher than (∼10%), although consistent at 1σ with, that measured by *Chandra* while, ELAIS S1 (Puccetti et al. 2006) source counts are 30% lower. However, in the overlapping flux range the latter is characterized by large errors due to the small number of relatively bright sources in the *Chandra* deep fields. Interestingly enough, the XMM–COSMOS counts match nicely, with smaller errors, those of the wide area HELLAS2XMM survey (Baldi et al. 2002), while the pioneering measurements of BeppoSAX (Fiore et al. 2001) are systematically higher than the counts from XMM-COSMOS.

[1]In particular, it is worthwhile to notice the good agreement between the XMM-COSMOS and the Hasinger et al. (2005) logN-logS, which has been used as input in our simulations. This good agreement can be considered as an a posteriori support of the reliability of those results from the simulations which depend on the assumed input logN-logS.

2.5.1 Resolved fraction of the X-ray background

One of the main aims of XMM-COSMOS with its large and medium–deep coverage, is to provide a solid census of the X-ray source population to be compared with observations and models of AGN evolution. According to recent synthesis models (see e.g.; Comastri et al. 1995; Gilli, Comastri & Hasinger 2006; Worsley et al. 2005) a high fraction of heavily obscured AGN is necessary to explain the spectral shape and the intensity of the X-ray background (XRB). We examine therefore which fraction of XRB is resolved into discrete sources in our survey.

As a first test, we computed the flux which XMM-COSMOS itself resolves into discrete sources by summing their fluxes weighted on the sky coverage in the 0.5–2 keV, 2–10 keV and 5-10 keV energy bands. As in Worsley et al. (2005) we used as reference value of the normalization at 1 keV of the XRB spectrum that of De Luca & Molendi (2004) which assumes that the spectral shape in the 1-10 keV band is a power-law with spectral index Γ=1.4 and a normalization at 1keV of 11.6 keV cm^{-2} s^{-1} keV^{-1}. The latter value corresponds to a flux of 0.80, 2.31 and 1.27$\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2} in the 0.5–2 keV, 2–10 keV and 5-10 keV energy bands, respectively. In the 0.5–2 keV band we measure a contribution of the sources to the XRB which corresponds to a normalization at 1 keV of $0.49\pm0.08\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2}. The corresponding values in the 2–10 keV and 5–10 keV bands are $0.92\pm0.22\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2} and $0.28\pm0.15\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2}. Therefore XMM-COSMOS resolves by itself $\sim$65%, $\sim$40% and $\sim$22% of the XRB into discrete sources in the 0.5–2 keV, 2–10 keV and 5–10 keV energy bands, respectively. It is worth noticing that the flux measured by De Luca & Molendi (2004) is the highest measured in literature in the 1–10 keV energy range (see e.g. Gilli, Comastri & Hasinger 2006,for a complete collection). It is also worth noticing that we computed the fraction of resolved XRB by assuming that all the sources have the same spectrum. Therefore, in our estimate, effects due to the broad absorption and spectral index distributions of AGNs are not included.

In Figs. 2.7, 2.8 and 2.9 we compared our logN-logS to those predicted by the recent XRB model of Gilli, Comastri & Hasinger (2006). This model makes use of the most recent observational constraints on the AGN populations and includes a conspicuous fraction of Compton thick AGN which, however, are not expected to significantly contribute to the XMM-COSMOS counts. In the 0.5–2 keV band a direct comparison of our data with the model shows a 1σ

agreement at the bright end. At the faint end the model predicts a slightly higher normalization when compared to most of the plotted data, including ours. A similar behavior is observed in the 2–10 keV band. It is worth noticing that in the model the average unabsorbed power-law spectral index of the sources is $<\Gamma>\sim1.8$ in the flux interval sampled by XMM-COSMOS (see Fig. 19 in Gilli, Comastri & Hasinger (2006)). Since in our data analysis we assumed $<\Gamma>=1.7$ we expect in this band a slight ($\sim$10%) underestimation of the fluxes when compared to those of the model. This effect is almost negligible (i.e. $<$5%) in the other bands investigated here.

By integrating our best fit 2–10 keV logN-logS between infinite and zero, and assuming that the slope of the "real" logN-logS remains constant down to low fluxes, we estimated the total contribution of AGNs to the XRB. We predict a total flux of AGNs in the XRB of 1.25×10^{-11} erg cm^{-2} s^{-1} deg^{-2}. This value is $\sim$40% lower than that measured by De Luca & Molendi (2004), and $\sim$ 25% smaller than those obtained by integration the model logN-logS of Gilli, Comastri & Hasinger (2006) which predicts a flux of $\sim1.66\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2}. This discrepancy between our predicted flux and that of Gilli, Comastri & Hasinger (2006), could arise by the fact that, in our measurement, we consider that all the sources have the same spectrum and from statistical uncertainty of the logN-logS parameters. By assuming an average spectral index $<\Gamma>=1.4$ for all our sources, we obtain a value for the total flux of the AGNs of $\sim1.48\times10^{-11}$ erg cm^{-2} s^{-1} deg^{-2} which is consistent, within the statistical uncertanties, with the prediction of the model. Considering the total flux of the XRB predicted by the model and our estimate from the logN-logS distributions, in the 2–10 keV band, XMM-COSMOS resolves the $\sim$55–65% of the total flux of the XRB into discrete sources.

It is interesting to observe how in the 5–10 keV band our data are in good agreement with the prediction of the model. This result is particularly important since in this band it is expected the major contribution of highly absorbed AGN, which are an important ingredient of the XRB models. A detailed analysis of the spectral properties of the brightest X-ray sources in XMM-COSMOS is presented by Mainieri et al. (2007).

2.6 Sample variance

The amplitude of source counts distributions varies significantly among different surveys (see e.g. Yang et al. 2003; Cappelluti et al. 2005,and references

therein). This "sample variance", can be explained as a combination of Poissonian variations and effects due to the clustering of sources (Peebles 1980; Yang et al. 2003). The variance of counts in cells for sources which are angularly correlated can be obtained with:

$$< (N - \mathcal{N}\Omega)^2 >= \mathcal{N}\Omega + \mathcal{N}^2 \int d\Omega_1 \, d\Omega_2 w(\theta_{1,2}) \tag{2.9}$$

where $\mathcal{N}$ is the mean density of objects in the sky, Ω is the cell size, and $w(\theta_{1,2})$ is the angular two–point correlation function. The first term of Eq. 2.9 is the Poissonian variance and the second term is introduced by the large scale structure. In order to determine whether the differences observed in the source counts of different surveys could arise from the clustering of X-ray sources, we estimated the amplitude of the fluctuations from our data, by producing subsamples of our survey with areas comparable to those of. e.g., *Chandra* surveys.

The XMM-COSMOS field and the Monte Carlo sample fields of Section 4 were divided in 4,9,16 and 25 square boxes. Making use of the 0.5–2 keV energy band data, we computed for each subfield, the ratio of the number of real sources to the number of random sources. In order to prevent incompleteness artifacts, we conservatively cut the limiting flux of the random and data sample to 5×10^{-15} erg cm^{-2} s^{-1}. At this flux our survey is complete over the entire area. In order to avoid artifacts introduced by the missing pointings in the external part of the field of view, we concentrated our analysis to the central $80'\times80'$. In Fig.2.10 we plot the ratio of the data to the random sample as a function of the size of the cells under investigation. The measured fractional standard deviations of the sample is reported in Table 2.3. Using Eq. 2.9 we computed the expected amplitude of source counts fluctuations with $w(\theta_{1,2})$ taken from Miyaji et al. (2007). They computed the X-ray two–point correlation function in the XMM-COSMOS field and detected clustering signal with angular correlation length θ_0 $\sim$1.9" $\sim$0.8" and $\sim$6" in the 0.5–2 keV, 2–4.5 keV and 4.5–10 keV band, respectively. The observed slope is γ=1.8 in all the energy bands.

The predicted fractional standard deviations are therefore 0.13, 0.19, 0.23 and 0.28 on scales of 0.44 deg^2, 0.19 deg^2, 0.11 deg^2 and 0.07 deg^2, respectively. These values are in good agreement with those observed in the sub-samples of our dataset as shown by the value of the fitted χ^2 to the counts in cell fluctuations (see Table 2.3). As shown in Table 2.3, at this limiting flux and on the areas considered here the main contribution to the source counts fluctuations

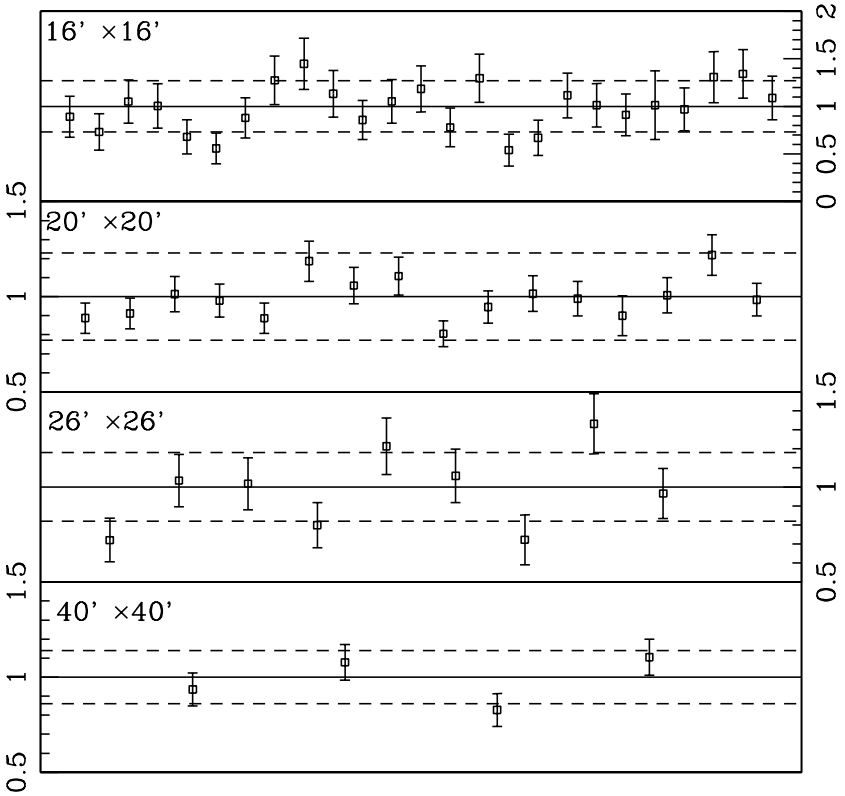

Figure 2.10: The counts in cell fluctuations within the XMM-COSMOS field. The data are normalized to a random distributed field in boxes of 40',26',20' and 16' of side, respectively. The dashed lines represent the 1σ expected fluctuation.

is from the Poissonian noise. At the flux limit assumed here, the ratio σ_{cl}/σ_p increases from ∼0.5 on the smallest scale (16 x 16 arcmin) to ∼0.85 on the largest scale (40 x 40 arcmin). This ratio scales as:

$$\sigma_{cl}/\sigma_p \propto \mathcal{N}^{0.5}\theta_0^{(\gamma-1)/2}a^{(3-\gamma)/2}, \tag{2.10}$$

where $\mathcal{N}$ deg^{-2} is the surface density of the sources, θ_0 (deg) is the angular correlation length and a (deg) is the size of the cell. In order to estimate at which flux limit fluctuations introduced by the large scale structure are predominant, we estimate that σ_{cl}/σ_p would be ∼1 on the smallest scale, corresponding to a Chandra ACIS field of view (16 x 16 arcmin) at a surface density of the order of ∼900 deg^{-2}, corresponding to a 0.5-2 keV flux S∼ 8×10^{-16} erg

cm^{-2} s^{-1}. At even fainter fluxes the dominant contribution to the total expected source counts fluctuations on this area (σ_{exp}) comes from the large scale structure, therefore the contribution of statistical fluctuations becomes less important. With the same procedure, we can estimate the total expected fluctuations (σ_{exp}) and the relative importance of σ_p and σ_{cl} also for the hard band (5-10 keV), even if in this band we do not have enough statistics to divide our field in sub-samples. Using the formal best fit for θ_0 in this band found by Miyaji et al. (2007) (θ_0=6"), we find that at the faintest 5-10 keV flux (S$\sim$10^{-14} erg cm^{-2} s^{-1}) sampled by the XMM-COSMOS survey ($\mathcal{N} \sim 110$ deg^{-2}) the ratio σ_{cl}/σ_p is smaller than one on all the four scales here analyzed, with a total expected standard deviation of the fluctuations ranging from $\sim$0.20 on the largest scale to $\sim$0.40 on the smallest scale. These values for σ_{exp} are significantly larger than those shown in Table 3 for the soft band, because in the hard band the surface density of sources is lower and the angular correlation length is higher than in the soft band.

This analysis is at least qualitatively consistent with Figures 8 and 10, which show a significantly larger dispersion in the data from different surveys in the hard band than in the soft band. Moreover, the results here discussed are also consistent with the observed fluctuations in the deep Chandra fields (see, for example, Bauer et al. 2004). Large area, moderately deep surveys like XMM-COSMOS are needed to overcome the problem of low counting statistics, typical of deep pencil beam surveys, and, at the same time, to provide a robust estimate of the effect of large scale structure on observed source counts.

As a final consideration, we tried to compute the expected intrinsic variance of XMM-COSMOS. This estimate must be made with care since we have only one sample on this scale. Assuming that the angular correlation function of Miyaji et al. (2007) was universal, the residual uncertainties on the source counts are estimated to be $<$5-6% in the 0.5–2 keV and 2–10 keV bands, and of the order of the 10% in the 5–10 keV band.

2.7 Summary

The data analysis of the first run of observations of the XMM-*Newton* COSMOS wide field survey has been presented. A total of 1390 point-like sources are detected on a contiguous area of about 2 deg^2 down to fluxes of 7.2$\times$10^{-16} erg cm^{-2} s^{-1}, 4.0$\times$10^{-15} erg cm^{-2} s^{-1}and 9.7$\times$10^{-15} erg cm^{-2} s^{-1}in the 0.5–2

0.5–2 keV S_{lim}=5×10^{-15} erg cm^{-2} s^{-1}					
Area arcmin^2	σ_{obs}	σ_p	σ_{cl}	σ_{exp}	
$40'\times40'$	0.09 ± 0.04	0.10	0.09	0.13	4.21/3
$26'\times26'$	0.20 ± 0.05	0.15	0.10	0.19	8.93/8
$20'\times20'$	0.21 ± 0.04	0.20	0.11	0.23	16.63/15
$16'\times16'$	0.24 ± 0.02	0.25	0.12	0.28	25.15/24

Table 2.3:

keV, 2–10 keV and 5–10 keV energy bands, respectively. The detection procedure was tested through Monte Carlo simulations which confirmed the high level of accuracy in the determination of the source properties (aperture photometry and positioning) and allowed us to keep statistical biases under control.

A robust estimate of X–ray source counts at both soft and hard energies, obtained thanks to the large number of sources detected in the XMM–COSMOS survey, is presented in this paper. The differential logN–logS was fitted with a broken power-law model in the 0.5–2 keV and 2–10 keV energy bands, and with a single power-law model in the 5–10 keV energy band. In the soft 0.5–2 keV band, already extensively covered by ROSAT, XMM–*Newton* and *Chandra* surveys over a range of fluxes encompassing those sampled by the COSMOS survey, our results are in excellent agreement with previous analysis (see Fig 2.7), providing an independent evidence of the validity of our data analysis procedure.

The large number of X–ray sources of the COSMOS survey allowed us to constrain with unprecedented accuracy the logN-logS parameters in the 2–10 and 5–10 keV energy ranges over a range of fluxes which were previously poorly constrained. Most importantly, in the hard 5–10 keV band, we were able to fill the gap between the deep *Chandra* surveys in the CDFS and CDFN and shallower large area surveys. The deviations from other surveys, which are, however, less than 30%, have been explained in terms of low counting statistics of pencil beam surveys, and partially by the effect of large scale structure. The major step forward in the determination of hard X–ray source counts achieved thanks to the XMM–COSMOS survey will provide an important reference point for the study of the AGN demography and evolution

especially with applications to obscured AGN. More specifically, the evolutionary properties of the obscured AGN can be tightly constrained, since they are indeed very sensitive, according to the most recent model of the X–ray background (Gilli, Comastri & Hasinger 2006), to the shape of the hard X–ray source counts around the break flux, which is precisely where the COSMOS data play a key role. In this context, we compared our results to the most recent predictions of the model by Gilli, Comastri & Hasinger (2006), finding a remarkable agreement between data and model.

The second pass of the XMM–*Newton* observations in the COSMOS field (600 ks) has already started, and is expected to significantly increase the total number of X–ray sources. The results of the full XMM–COSMOS survey including the complete (AO3+AO4) source catalogue will be the subject of a future paper. It is anticipated that, when completed, the XMM–COSMOS survey will provide a number of X–ray sources over a large enough contiguous area and down to moderately deep fluxes that it will make possible the study of AGN evolution and their connection with the large scale structure in which they reside with unprecedented detail. The COSMOS field has been granted 1.8 Msec observation with *Chandra* in its central square degree (C-COSMOS, P.I. Martin Elvis). The joint Chandra and XMM–*Newton* observation will provide an unprecedented lab for AGN physics.

Chapter 3

Angular Clustering of the X-ray Point Sources

3.1 Introduction

Results from recent X-ray surveys have made very significant contributions to understanding formation and evolution of supermassive blackholes (SMBHs) at galaxy centers. In particular, studies of X-ray luminosity function and its evolution have been providing the most reliable current estimates of the accretion history to SMBH. One of the most important findings in recent years is that luminous active galactic nuclei (AGNs) arise earlier in the history of the universe than lower luminosity ones (Ueda et al. 2003; Hasinger et al. 2005; Barger et al. 2005; La Franca et al. 2005). This suggests that more massive SMBHs have been formed earlier in the universe, and reside quiescently at the centers of giant elliptical galaxies in the later epochs, while more numerous, less massive SMBHs have been formed and accreted later in the history of the universe.

Clustering properties of AGNs and their evolution with redshift provide yet additional clues to understanding the accretion processes onto the SMBHs. These give clues to environments of AGN activities. In the framework of the Cold Dark Matter (CDM) structure formation scenario, clustering properties or the bias of AGNs over a sufficiently large scale $b_{AGN} = (\delta\rho/\rho)_{AGN}/(\delta\rho/\rho)_{mass}$ may be related to the typical mass of dark matter halos in which they reside, (Mo & White 1996; Sheth et al. 2001). At the same time, the mechanisms of triggering the AGN activity, which might be closely related to galaxy interactions and/or merging (Menci et al. 2004; Di Matteo et al. 2005), yield a

clustering of AGNs and can therefore be infered from the clustering analysis.

Since strong X-ray emission is a typical feature of an AGN activity, X-ray surveys provide most efficient means of constructing comprehensive complete samples of AGNs without contamination from the light in the stellar population of host galaxies. In particular, surveys in the harder ($E > 2$ keV) X-ray band such as available from *XMM-Newton* are very efficient in finding not only unobscured AGNs, which are relatively easy to select also in the optical bands, but also obscured ones, which are difficult to select with optical selection criteria alone. This is important because most of the accretion ($\sim 80\%$; Comastri et al. 1995; Gilli et al. 2001; Ueda et al. 2003) occurs in AGNs obscured by gas (in X-ray bands) and dust (in the optical bands). While one approach in investigating the environment of AGNs is to measure AGN overdensities around known clusters of galaxies (Cappi et al. 2001; D'Elia et al. 2004; Cappelluti et al. 2005), a more common and direct measure can be obtained by calculating auto-correlation functions (ACFs) of well-defined samples of X-ray selected AGNs.

While small number statistics limits the accuracy of the clustering measurements of X-ray selected AGNs, there are a number of reports on the detection of the correlation signals. Samples based on the *ROSAT* All-Sky survey have mainly constrained the clustering properties of type 1 local AGNs at $z < 0.3$. The correlation lengths resulting from the angular (Akylas et al. 2000) and 3D (Mullis et al. 2004) analyses of these samples are 6-7 h^{-1} Mpc [1]. Due to the wide redshift distribution, it is more difficult to obtain clustering signal in deeper X-ray surveys before redshifts for a complete set of X-ray sources are obtained. Nevertheless, Basilakos et al. (2004, 2005) measured strong angular correlation signals in their *XMM-Newton*/2dF survey, which covers a total area of 2 deg^2 over two fields. They obtained a correlation length of $\sim 7.5h^{-1}$ Mpc in physical units for both soft and hard X-ray selected sources, suggesting a clustering evolution which is fixed in the proper coordinate between z$\sim$ 0 and z$<$ 1. At much fainter X-ray fluxes, Gilli et al. (2005) analyzed the projected-distance correlation function $w(r_p)$ for the X-ray sources with spectroscopic redshifts in the Chandra Deep Fields North (CDF-N) and South (CDF-S). They found significantly different clustering properties in these two fields, suggesting a cosmic variance effect. Recently, Yang et al. (2006) made detailed analysis on their 0.4 deg^2 *Chandra* Large Area Synoptic X-ray Survey (CLASXS) supplemented by CDF-N. With spectroscopic redshifts for a

[1] Throughout this paper we adopt $(H_0, \Omega_M, \Omega_\Lambda) = (100h\,kms^{-1}Mpc^{-1}, 0.3, 0.7)$

good portion of the sources, they explored the clustering properties in different redshift and luminosity bins as well as intrinsic absorption bins. They found the evolution of bias with redshift but they did not find significant dependence in the clustering properties of X-ray selected AGNs based on either luminosity or intrinsic absorption.

One of the main aims of the COSMOS (Scoville et al. 2007) project is to trace the evolution of the large-scale structure of the universe with an unprecedented accuracy and redshift baseline. The *XMM-Newton* Survey, covering the entire COSMOS field (XMM-COSMOS; Hasinger et al. 2007), is one of the most extensive *XMM-Newton* Survey programs conducted so far. In the first-year *XMM-Newton* observations, about 1400 X-ray point sources have been detected and cataloged (Cappelluti et al. 2007) (hereafter C07), which are dominated by AGNs at redshifts $0.7 < z < 2$ (Brusa et al. 2007; Trump et al. 2007).

In this paper, we report the first results of our investigations on the large scale structure through an angular auto-correlation (ACF) function analysis of the X-ray point sources detected in XMM-COSMOS, as a preview of more detailed studies in the near future. Our future studies include the derivation of the direct three-dimensional correlation function using redshift information already available for a large portion of the X-ray sources and the analysis of the cross-correlation of the X-ray sources with galaxies in the multiwavelength COSMOS catalog.

The outline of the paper is as follows. In Sect. 3.2, we explain the selection of our samples of X-ray sources to be used in the correlation function analysis, which are subsets of those described in C07. Details of the calculations, including the ACF estimator, the random sample, and power-law fits are presented in Sect. 3.3. The de-projection to the three-dimensional correlation function is presented in Sect. 3.4. The results are discussed in Sect. 3.5. We summarize our conclusions in Sect 3.6

3.2 Sample Selection

Our samples consist of the X-ray sources detected in the first-year *XMM-Newton* observations of the COSMOS field. The source detection, construction of the sensitivity maps, and source counts are described in C06. The X-ray source catalogs in three energy bands, corresponding to energy channels of 0.5-2 (SFT), 2-4.5 (MED) and 4.5-10 (UHD) keV are used.

Band (keV)	Number (cts s^{-1})	$CR_{lim,min}$–$CR_{lim,max}$ erg s^{-1} cm^{-2}	$S_{x,lim}$ range (deg^{-2})	Area
SFT	1037	$7.0\,10^{-4}$–$2.2\,10^{-2}$	$6.7\,10^{-16}$–$2.1\,10^{-14}$	1.43
MED	545	$7.0\,10^{-4}$–$2.5\,10^{-2}$	$4.6\,10^{-15}$–$1.6\,10^{-13}$	1.56
UHD	151	$9.0\,10^{-4}$–$2.2\,10^{-2}$	$8.7\,10^{-15}$–$1.8\,10^{-13}$	1.25

Table 3.1: X-ray sources and Sensitivity limits. Flux range corresponding to the *CR* limits. The conversions have been made, following C06, to fluxes in 0.5-2.0, 2.0-10, 5.0-10 keV assuming power-law spectra of photon indices $\Gamma = 2.0$, 1.7, and 1.7 for the SFT, MED, and UHD bands respectively.

For the angular ACF studies, we have applied further selection criteria to the C06 sources to minimize the effects of possible systematic errors in the sensitivity maps. The applied criteria for this kind of analysis should be stricter than those adopted for the derivation of the $\log N - \log S$ function, because localized systematic errors may cause spurious clustering of X-ray sources. In order to do this, we have scaled up the original sensitivity map to:

$$CR_{lim,acf} = \max(a\,CR_{lim,C06} - b,\, CR_{lim,min}), \tag{3.1}$$

where $CR_{lim,C06}$ is the limiting count rate (in cts s^{-1}) in the original C06 sensitivity map and $CR_{lim,acf}$ is the sensitivity map used for the ACF analysis. After a number of trials, the scaling coefficients (a,b) have been set to (1.33,1×10^{-4}), (1.40, 0.) and (1.44, 4×10^{-4}) for the SFT, MED and UHD bands respectively. We have excluded the area where $CR_{lim,acf}$ exceeds $CR_{lim,max}$ (low exposure areas close to the field borders). Those X-ray sources with *CR*'s below $CR_{lim,acf}$ at the source position have been excluded from the ACF analysis. After these screenings, the numbers of sources for the ACF analysis are 1037, 545, and 151 for the SFT, MED, and UHD bands respectively. While the sensitivity in the soft band is the best among the three bands, some X-ray sources are hard enough that they are detected in only MED and/or UHD bands. These are mainly highly obscured AGNs. Out of the 545 (151) MED (UHD) band sources after the screening process, 59 (13) have not been detected in the SFT band, and only one UHD sources have escaped the detection in the MED band (before the screening process). The numbers of the X-ray sources, values of $CR_{lim,min}$,$CR_{lim,max}$ and the total areas used for the ACF analysis are summarized in Table 3.1.

3.3 Angular Correlation Function Calculation

3.3.1 The ACF calculation

In calculating the binned ACF, we have used the standard estimator by Landy & Szalay (1993):

$$w_{est}(\theta_i) = (DD - 2DR + RR)/RR, \tag{3.2}$$

where DD, DR, and RR are the normalized numbers of pairs in the i-th angular bin for the data-data, data-random, and random-random samples respectively. Also we use the symbols D and R to represent the data and random samples respectively. Expressing the actual numbers of pairs in these three combinations as $n_{pair,DD}(\theta_i)$, $n_{pair,DR}(\theta_i)$ and $n_{pair,RR}(\theta_i)$, the normalized pairs are expressed by:

$$\begin{aligned} DD &= n_{pair,DD}(\theta_i)/[N_D(N_D-1)] \\ DR &= n_{pair,DR}(\theta_i)/(N_D N_R) \\ RR &= n_{pair,RR}(\theta_i)/[N_R(N_R-1)] \end{aligned} \tag{3.3}$$

where N_D, and N_R are the numbers of sources in the data and random samples respectively. The number of objects in the random sample has been set to 20 times of that in the data sample. This makes the variance of the second and third terms of Eq. 3.2 negligible in the error budget of w_{est}.

Our *XMM-Newton* observations have varying sensitivity over the field. In order to create a random sample, which takes the inhomogeneity of the sensitivity over the field into account, we have taken the following steps.

1. Make a random sample composed of N_R objects, where N_R is an integer times N_D.

2. For each random object, assign a count rate from a source from the data sample. The assignments are made in sequence so that the CR distributions of the random and the data sample objects are exactly the same.

3. For each random object, assign a random position in the field. If the sensitivity-map value at this position ($CR_{lim,acf}$ from Eq. 3.1, in units of counts s^{-1}) is larger than the assigned CR, find a different position. Repeat this until the position is sensitive enough to detect a point-source with the assigned CR.

As a check on this procedure, we have also calculated ACFs using two other methods of generating random samples. The second method is to assign the count rates to the random sources drawn from a $\log N - \log S$ relation (e.g. Moretti et al. 2003), instead of copying the count rates of the data sample. Then the source is placed at a random position in the field. If the sensitivity limit at this position is higher than the assigned CR, this random source is rejected. Another, but more sophisticated and computationally demanding method is to generate random sources based on the $\log N - \log S$ relation, down to a flux level much lower than the sensitivity limit of our observations. These sources are then fed into a simulator, taking into account *XMM-Newton* instrumental effects, including position-dependent PSF, exposure maps, and particle background. The entire first-year XMM-COSMOS image has been simulated and the same source detection procedure as that applied to the actual data has been applied on the simulated data. A random sample R is generated from 10 simulated XMM-COSMOS fields. Using the above two methods for random sample generations did not alter the results significantly. In the following analysis, we show the results obtained by the first method.

3.3.2 Error Estimation and Covariance Matrix

We have estimated the errors using the variance of the calculated ACF by replacing D in Eq. 3.2 by random samples. A random sample with the same number of objects and the same set of count rates as D has been drawn independently from R in Eq. 3.2. We denote the random sample as a replacement of D during the error search by R' to distinguish from R. For each angular bin, a 1 σ error has been calculated as a standard deviation from resulting ACFs calculated from N_{run} =80 different R' samples, which is then multiplied by a scaling factor $[1+w(\theta_i)]^{1/2}$ (hereafter referred to as "the scaled random errors"). This scaling factor corrects for the difference between the errors in the null-hypothesis case, obtained from R's, and those in the presence of the correlation signal. This is in line with the observation that the error in each bin of the ACF calculated using the Landy & Szalay (1993) estimator is approximately a Poisson fluctuation of the number of data-data pairs in the bin, i.e., $\sigma \sim [1+w(\theta)]/\sqrt{n_{pair,DD}}$. The standard deviation of the null-hypothesis ACF, obtained by replacing D by an R', is $\sigma_{ran} \sim 1/\sqrt{n_{pair,R'R'}}$. The scaling factor can be obtained by using the relation, $n_{pair,DD} \approx [1+w(\theta)]n_{pair,R'R'}$.

In correlation functions, the errors in different angular bins are not inde-

pendent from one another and correlations among the errors have to be taken into account when we make function fits. Thus we have also estimated full covariance matrix in order to represent the correlations among errors by

$$\begin{aligned} M_{cov,ij} &= \sum_k (w_{R,k}(\theta_i) - \langle w_R(\theta_i)\rangle) \\ &\times (w_{R,k}(\theta_j) - \langle w_R(\theta_j)\rangle)/N_{run} \\ &\times (1+w_{est}(\theta_i))^{1/2}(1+w_{est}(\theta_j))^{1/2} \end{aligned} \tag{3.4}$$

where $w_{R,k}(\theta_i)$ is the ACF value for the k-th random run (k runs through 1 to $N_{run} = 80$), $\langle w_R(\theta_i)\rangle$ is their mean value, evaluated at the center of the i-th angular bin θ_i and $w_{est}(\theta_i)$ is from Eq. 3.2. The square roots of the diagonal elements of Eq. 3.4 are the scaled random errors discussed above. The covariance matrix calculated in Eq. 3.4 is used later in Sect. 3.3.4. Strictly speaking, Eq. 3.4 only takes into account the correlations of errors for the random cases, but not the correlation of errors due to clustering. One way of explaining this is that removing/adding one source (by a Poisson chance) affects multiple angular bins and this is represented by non-diagonal elements of Eq. 3.4. On the other hand, correlation of errors due to large scale structure or the cosmic variance is not represented by this. If we observe another part of the sky, we sample different sets of large scale structures such as filaments and voids. Since the existence or non-existence of one such structure affects multiple angular bins, there should be additional contribution to the non-diagonal elements of the correlations among errors in different angular bins. Eq. 3.4, based on many random samples, thus includes the former type of the correlation of errors but not the latter. One way to include also the latter effect is to use the Jackknife re-sampling technique, as was done by e.g. Zehavi et al. (2004). However, the Jackknife re-sampling requires to divide the sample into many statistically independent regions, which is not practically possible in our case.

3.3.3 The binned ACF results

The ACFs have been calculated for the three bands in logarithmically equally spaced bins with $\Delta\log\theta = 1/6$. The results are shown in Fig 3.1, where the upper panels, composed of two layers of logarithmic plots with positive and negative parts ($\log|w(\theta)| > -2.8$ respectively), are attached together. The lower panels show fit residuals for the best-fit functions described in Sect. 3.3.4. Changing the bin size did not change the clustering amplitude significantly.

The ACFs are presented with the scaled random errors. Positive signals have been detected down to $\theta \sim 0.5$' in the 0.5-2 keV band and $\theta \sim 1$' in the other bands. At the smallest scales, correlation signal is negative, probably due to confusion effects, where two sources separated by a distance comparable to or closer than the point spread function (PSF) cannot be detected separately in the source detection procedure and may well be classified as one extended source. In our current sample, the sources that have been classified as extended have been removed from the sample. The effect of this is discussed in detail in Sect. 3.3.5. In the 0.5-2 and 2-4.5 keV bands, positive signals extend out to $\sim 20'$. Negative signals are seen at the largest angular scales, probably due to the integral constraint as discussed below.

3.3.4 Power-law Fits

In order to make a simple characterization of our ACF results, we have fitted the ACF with a power-law model of the form:

$$w_{mdl}(\theta) = A\,\theta^{1-\gamma} \tag{3.5}$$

where γ is the slope index of the corresponding three-dimensional correlation function. We use the normalization A as a fitting parameter rather than the angular correlation length $\theta_c = A^{1/(\gamma-1)}$, since this gives much better convergence of the fit.

The fits are made by minimizing χ^2_c. The subscript c denotes that the correlations between errors have been taken into account through the inverse of the covariance matrix:

$$\chi^2_c = \Delta^T M^{-1}_{cov} \Delta, \tag{3.6}$$

where Δ is a vector composed of $w_{est}(\theta_i) - w_{mdl}(\theta_i) + C$, M_{cov} is the covariance matrix calculated in Eq. 3.4 with $N_{run} = 80$, and C is a constant to compensate for the integral constraint as discussed below.

Due to the finite area and the construction of w_{est} in Eq. 3.2, the estimated angular correlation function satisfies the integral constraint (e.g Basilakos et al. 2005; Roche & Eales 1999):

$$\int\int w_{est} d^2\Omega = 0. \tag{3.7}$$

This constraint usually results in w_{est} underestimating the true underlying angular correlation function by the constant C. Under an assumption that

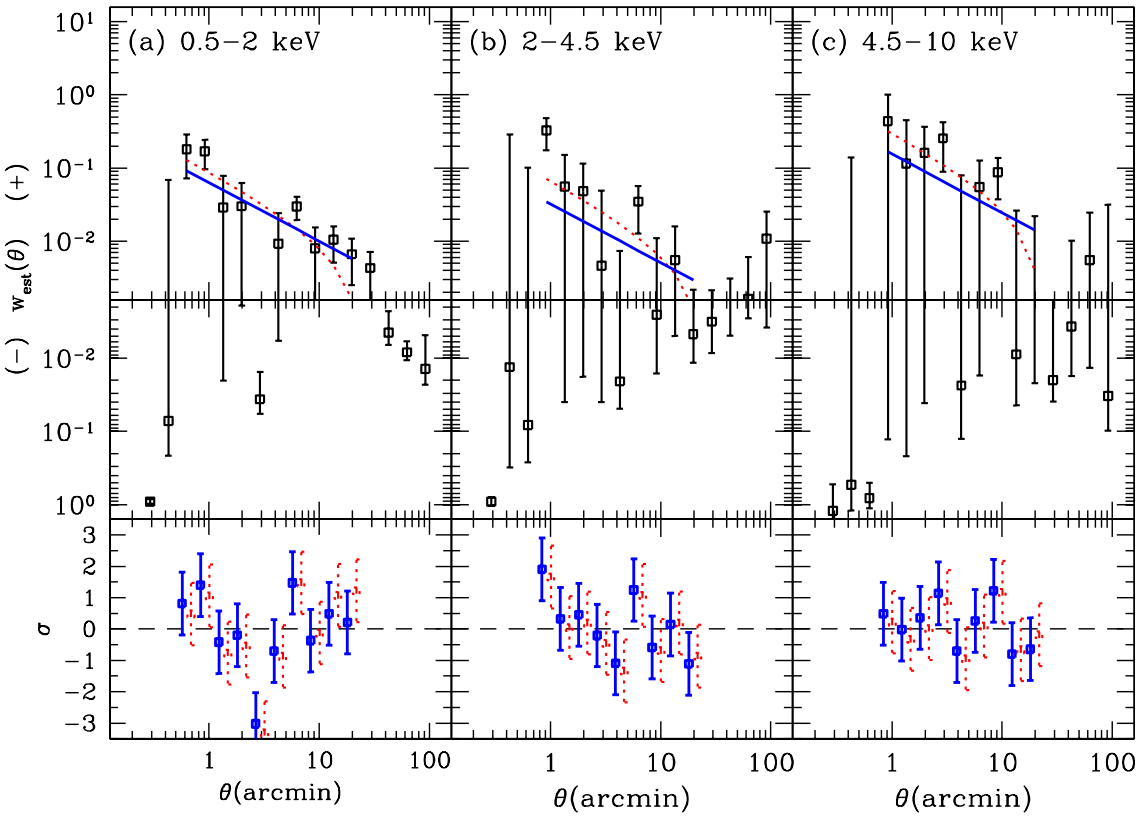

Figure 3.1: The binned estimated angular correlation functions $w_{est}(\theta)$ are plotted for the X-ray sources detected in the first-year *XMM-Newton* data in three standard energy bands as labeled. The vertical scale is logarithmic, where the positive and negative parts ($\log|w_{est}(\theta)| > -2.8$ respectively) are attached together. The 1σ errors are the diagonal components of Eq. 3.4, i.e., the scaled random errors. The blue-solid and red-dotted lines (colors in the electronic version only) show the best-fit power-law models for $\gamma - 1 = 0.8$ without and with an integral constraint respectively. The models are plotted in the range where the fits are made. Fit residuals in terms of σ has been also plotted in the lower panels for the two models (slightly offset for visibility) in the same line styles (colors) as the models.

the true underlying $w(\theta)$ is a power-law and is extended to the scale of the survey area, one can include C in the fitting process, where C can be uniquely determined by θ_c and γ by imposing the integral constraint (Roche & Eales 1999),

$$C = \sum_i A\theta_i^{1-\gamma} RR(\theta_i) / \sum_i RR(\theta_i), \tag{3.8}$$

where the sums are over angular bins and $RR(\theta_i)$ is the number of random-random pairs in the i-th angular bin. The above assumption is not necessarily true. If residual systematic errors in the sensitivity maps are the main cause of the negative values at large angular separations, the determination of C shown above is not valid. However, this gives an approximate estimate of the degree of the underestimation by the integral constraint. This sets an limitation to the our angular ACF analysis, where the estimated C values are not negligible compared with the amplitude of the ACF signal. We have made fits with or without including the integral constraint.

Because of the limited signal-to-noise ratio, we were not able to constrain A and γ simultaneously. Thus we have calculated the best-fit values and 1σ confidence errors for the amplitude for two fixed values of $\gamma - 1 = 0.8$ and 0.5. The former value is for the canonical value for local galaxies (e.g. Peebles 1980; Zehavi et al. 2004), the latter is approximately the slope found for X-ray sources in the *Chandra* Deep Fields (Gilli et al. 2005). The angular fit results are summarized in Table 3.2. In this table, fits with different bands and parameters are identified with a Fit ID.

The angle range for the fits are $\theta_{min} < \theta < \theta_{max}$ and the boundaries are also shown in Table 3.2. For fit ID's S1-S4, M1-M4, and U1-U4, we have set $\theta_{min} = 0.5'$ (SFT) or $0.7'$ (MED,UHD), which is the minimum at which ACF is still positive, and below which the ACF goes negative due to the *XMM-Newton* PSF. Likewise, we set $\theta_{max} = 24'$, which is about the maximum scale where the ACF is still positive. The best-fit models for $\gamma - 1 = 0.8$ are overplotted in Fig. 3.1 in the bin ranges included in the fits.

As another choice, we have set θ_{min} and θ_{max} in such a way that the range approximately corresponds to the projected comoving distance range of 1-16 h^{-1} Mpc (Fit ID S5,S6,M5,M6,U5, and U6) at the effective median redshift of the sample ($\bar{z}_{eff}$ discussed below in Sect. 3.4. The rationale for the maximum scale is that, in our subsequent analysis, the correlation functions are converted to the root mean square (rms) density fluctuation with a $8h^{-1}$ Mpc-radius sphere (therefore the relevant maximum separation is $16h^{-1}$) in dis-

Fit ID	Band	A	$\gamma-1$	θ_c	C	θ_{min}	θ_{max}
(keV)	(arcmin$^{1/(\gamma-1)}$)	(arcsec)	(arcmin)	(arcmin)			
S1	SFT	$0.063 \pm .008$	0.8	1.9 ± 0.3	0	0.5	24
S2	SFT	$0.093 \pm .012$	0.8	3.1 ± 0.5	$8\,10^{-3}$	0.5	24
S3	SFT	$0.034 \pm .004$	0.5	$0.07 \pm .02$	0	0.5	24
S4	SFT	$0.078 \pm .010$	0.5	$0.37 \pm .09$	$1.5\,10^{-2}$	0.5	24
S5	SFT	$0.059 \pm .009$	0.8	1.7 ± 0.3	0	1.6	24
S6	SFT	$0.089 \pm .015$	0.8	2.9 ± 0.6	$7\,10^{-3}$	1.6	24
M1	MED	$0.032 \pm .015$	0.8	$0.8^{+0.5}_{-0.4}$	0	0.7	24
M2	MED	$0.071 \pm .027$	0.8	2.2 ± 1.0	$5\,10^{-3}$	0.7	24
M3	MED	$0.013 \pm .008$	0.5	$.010^{+.016}_{-.009}$	0	0.7	24
M4	MED	$0.048 \pm .020$	0.5	$.14^{+.14}_{-.09}$	$9\,10^{-3}$	0.7	24
M5	MED	$0.021 \pm .016$	0.8	$.47^{+.49}_{-.40}$	0	1.6	24
M6	MED	$0.044 \pm .029$	0.8	$1.2^{+1.1}_{-0.9}$	$3\,10^{-3}$	1.6	24
U1	UHD	$0.15 \pm .05$	0.8	5.6 ± 2.3	0	0.5	24
U2	UHD	$0.32 \pm .09$	0.8	14 ± 5	$3\,10^{-2}$	0.7	24
U3	UHD	$0.075 \pm .024$	0.5	$0.34^{+0.25}_{-0.18}$	0	0.7	24
U4	UHD	$0.23 \pm .07$	0.5	$3.2^{+2.2}_{-1.6}$	$5\,10^{-2}$	0.7	24
U5	UHD	$0.080 \pm .032$	0.8	$2.5^{+1.3}_{-1.2}$	0	2.4	35
U6	UHD	$0.17 \pm .06$	0.8	$6.5^{+3.0}_{-2.7}$	$3\,10^{-2}$	2.4	35

Table 3.2: Results of the Power-law fits. One σ errors are shown. The effects of PSF merging (Sect. 3.3.5) have not been taken into account.

cussing bias parameters. The rationale for the minimum scale is to minimize the effects of non-linearity in discussing the bias parameters and typical halo masses.

3.3.5 Effects of Source Merging due to PSF

The *amplification bias*, due to which the estimated ACF from sources detected in a smoothed image (e.g. by a finite PSF) is amplified with respect to the true underlying ACF, has been first noted and discussed by Vikhlinin & Forman (1995) in the context of the clustering of X-ray sources. This is caused by merging of multiple sources which are separated by distances comparable to or closer than the PSF. The effect of this bias depends on the true underlying angular correlation function and the number density of the sources. In principle, full simulations involving PSF smoothing and the source detection process are required to estimate the amount of this bias. Basilakos et al. (2005) took a simplified approach in estimating this effect on their ACF from their *XMM-Newton*/2dF survey. In order to estimate the size of the effect, they used particles sampled from a cosmological simulation. They simulated the *XMM-Newton* sources by merging all the particle pairs closer than 6". They then compared the angular ACFs from the particles themselves and the simulated *XMM-Newton* sources. As a result, they estimated that the measured angular correlation length is overestimated by 3-4% due to the amplification bias.

In our case with XMM-COSMOS, we have explicitly excluded sources that are classified as extended by the source detection procedure (C06). This causes most of the source pairs closer than ~ 20" to disappear from the sample, since these pairs are classified as single extended sources. Because the exclusion of these sources can suppress the estimated angular correlation function, we use the term "PSF merging bias" rather than the "amplification bias". Pairs of sources that are closer than ~ 4 arcseconds are, however, detected as a single point source. We have applied a similar approach to Basilakos et al. (2005) in estimating the effects of the PSF merging bias. We have sampled particles from the COSMOS-Mock catalog extracted from the Millennium simulation (Kitbichler, M., priv. comm.) over ~ 2 deg^2 of the sky. Redshift, cosmological intrinsic redshift, and magnitudes in various photometric bands are provided for each mock galaxy in the catalog. We use the mock catalog to estimate the effect of the PSF merging bias in the *angular correlation function*. Thus the

selected objects from the mock catalog for our simulation do not have to physically represent to the actual X-ray selected AGNs. For our present purpose, we have chosen the mock galaxies in a redshift interval (roughly in the range $0.4 < z < 0.8$) and a magnitude range in such a way that the amplitude of the resulting angular ACF and the source number densities roughly match those of the X-ray samples. We have then created a simulated XMM-COSMOS catalog as follows: 1) source pairs with separations smaller than 4 arcseconds are merged into single sources and 2) pairs that are between 4 to 20 arcseconds from each other are eliminated. We repeated this experiment 19 times and compared the mean angular ACFs from the original particles and that from simulated XMM-COSMOS by making power-law fits to the mean ACFs. As a result we found that the ACF amplitudes measured using the sources in our source detection procedure on the XMM-COSMOS data are underestimated by 15% and 8% for the SFT and MED bands respectively. Corrections for this effect have not been applied for the values in Table 3.2, but are considered in further discussions. The effects is negligible in the UHD band, due to the relatively low number density of the sources detected in this band, which made the average distance among neighboring sources much larger than the PSF.

3.4 Implication for 3-D Correlation Function and Bias

3.4.1 De-Projection to Real Space Correlation Function

The 2-D ACF is a projection of the real-space 3-D ACF of the sources $\xi(r)$ along the line of sight. In the following discussions and thereafter, r is in comoving coordinates. The relation between the 2-D (angular) ACF and the 3-D ACF is expressed by the Limber's equation (Peebles 1980,e.g.,). Under the usual assumption that the scale length of the clustering is much smaller than the distance to the object, this reduces to:

$$w(\theta)N^2 = \int \left(\frac{dN}{dz}\right)^2 \times \int \xi(\sqrt{[d_A(z)\theta]^2 + l^2}\,(1+z)) \quad \left(\tfrac{dl}{dz}\right)^{-1} dl\, dz, \tag{3.9}$$

where $d_A(z)$ is the angular distance, N is the total number of sources and dN/dz is the redshift distribution (per z) of the sources. The redshift evolution of the

3-D correlation function is customarily expressed by

$$\xi(r,z) = (r/r_{c,0})^{-\gamma}(1+z)^{-3-\varepsilon+\gamma}, \tag{3.10}$$

where $\varepsilon = -3$ and $\varepsilon = \gamma - 3$ correspond to the case where the correlation length is constant in physical and comoving coordinates respectively. In these notations, the zero-redshift 3-D correlation length $r_{c,0}$ can be related to the angular correlation length θ_c by:

$$\begin{aligned} r_{c,0}^{\gamma} &= (N^2/S)\,\theta_c^{\gamma-1}, \\ S &= H_\gamma \int \left(\tfrac{dN}{dz}\right)^2 \left(\tfrac{c\,d\tau(z)}{dz}\right)^{-1} d_A^{1-\gamma}(1+z)^{-3-\varepsilon} dz \\ H_\gamma &= \tfrac{\Gamma[(\gamma-1)/2]\Gamma(1/2)}{\Gamma(\gamma/2)}, \end{aligned} \tag{3.11}$$

where $\tau(z)$ is the look back time. Note that all dependence on cosmological parameters are included in $d_A(z)$ and $\tau(z)$. We also define the comoving correlation length:

$$r_c(\bar{z}_{eff}) = r_{c,0}(1+\bar{z}_{eff})^{(-3-\varepsilon+\gamma)/\gamma} \tag{3.12}$$

at the effective redshift $\bar{z}_{eff}$, which is the median redshift of the contribution to the angular correlation (the integrand of the second of Eq. 3.11).

An essential ingredient of the de-projection process is the redshift distribution of the sources. At this stage, we do not yet have individual redshifts of a comprehensive set of the *XMM-Newton* complete sample. Thus we use expected distributions from the X-ray luminosity functions and AGN population synthesis models. We use the model by Ueda et al. (2003) (Luminosity-dependent density evolution or LDDE) for all bands. In calculating the redshift distribution, we have used the sensitivity map in units of CR and the actual *XMM-Newton* response function in each band. We also use Hasinger et al. (2005) type 1 AGN soft X-ray luminosity function (SXLF) for the 0.5-2 keV for comparison. The redshift distributions of the X-ray sources predicted by these models are plotted in Fig. 3.2. Both Ueda et al. (2003) and Hasinger et al. (2005) used samples with a very high identification completeness with redshifts measurements ($> 90\%$), at least down to the flux limits sampled by the XMM-COSMOS survey. Thus the effect on the expected redshift distribution due to the identification incompleteness is negligible. . In calculating the three dimensional correlation functions, we use the fits with and without integral constraints. The angular correlation amplitude A has been multiplied by a correction factor due to PSF merging as discussed in Sect. 3.3.5. Also we use the fits with $\gamma = 1.8$. The calculated $r_{c,0}$ and the comoving correlation

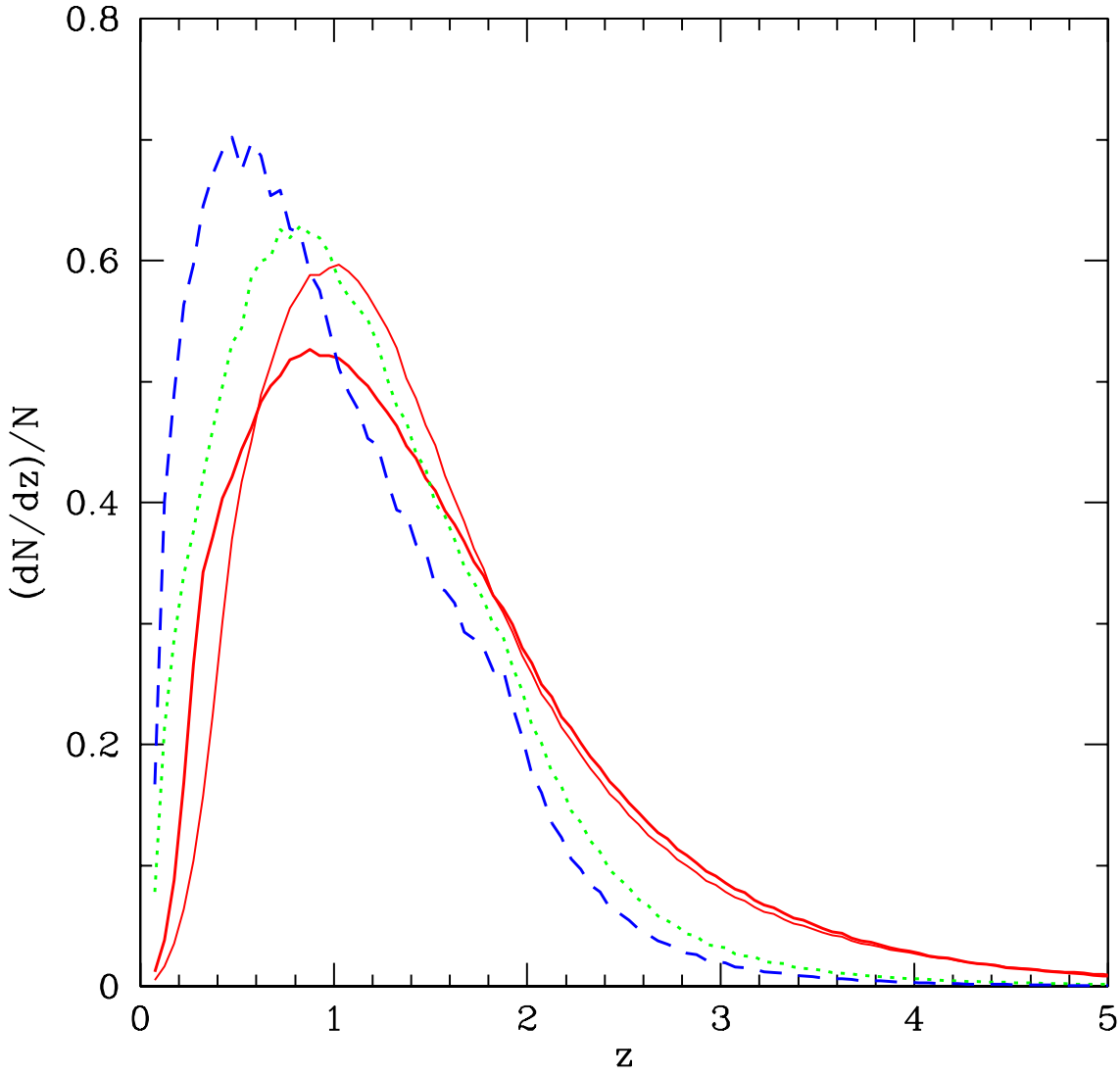

Figure 3.2: The model redshift distributions of the COSMOS-XMM sources in the 0.5-2 keV (solid lines), 2-4.5 keV (dotted line), and 4.5-10 keV (dashed line) respectively. The thicker and thinner solid lines correspond to the 0.5-2 keV band redshift distributions based on Ueda et al. (2003) model and Hasinger et al. (2005) soft X-ray luminosity function respectively.

Fit ID	γ	ε	$\bar{z}_{eff}$	$r_{c,0}$	$r_c(\bar{z}_{eff})$	Model
S1	1.8	-1.2	1.07	9.8±0.7	9.8±0.7	U03
S1	1.8	-1.2	1.11	9.4±0.7	9.4±0.7	H05
S1	1.8	-3.0	1.42	4.3±0.3	10.4±0.7	U03
S2	1.8	-1.2	1.07	12.1±0.9	12.1±0.9	U03
S5	1.8	-1.2	1.07	9.4±0.8	9.4±0.8	U03
S6	1.8	-1.2	1.07	11.8±1.1	11.8±1.1	U03
M1	1.8	-1.2	0.87	$5.8^{+1.4}_{-1.7}$	$5.8^{+1.4}_{-1.7}$	U03
M1	1.8	-3.0	1.13	$2.9^{+0.7}_{-0.8}$	$6.1^{+1.5}_{-1.8}$	U03
M2	1.8	-1.2	0.87	$9.0^{+1.8}_{-2.1}$	$9.0^{+1.8}_{-2.1}$	U03
M5	1.8	-1.2	0.87	$4.6^{+1.7}_{-2.5}$	$4.6^{+1.7}_{-2.5}$	U03
M6	1.8	-1.2	0.87	$6.9^{+2.2}_{-3.1}$	$6.9^{+2.2}_{-3.1}$	U03
U1	1.8	-1.2	0.60	$11.9^{+2.1}_{-2.4}$	$11.9^{+2.1}_{-2.4}$	U03
U1	1.8	-3.0	0.88	$6.6^{+1.1}_{-1.2}$	$12.5^{+2.2}_{-2.5}$	U03
U2	1.8	-1.2	0.60	19±3	19±3	U03
U5	1.8	-1.2	1.60	$8.4^{+1.7}_{-2.0}$	$8.4^{+1.7}_{-2.0}$	U03
U6	1.8	-1.2	1.60	$12.7^{+2.3}_{-2.7}$	$12.7^{+2.3}_{-2.7}$	U03

Table 3.3: Three Dimensional Correlation Lengths. Errors are 1 σ. Model designations– U03:Ueda et al. (2003), H05:Hasinger et al. (2005)

length at the effective median redshift $r_c(\bar{z}_{eff})$ are listed in Table 3.3 for selected results. The errors on $r_{c,0}$ and $r_c(\bar{z}_{eff})$ have been calculated for fixed γ and ε.

3.4.2 Bias and Comparison with Other Works

In order to estimate the bias parameter of the X-ray sources with respect to the underlying mass distribution, we calculate the rms fluctuation of the distribution of the X-ray sources in the sphere with a comoving radius of $r_{max} = 8h^{-1}$ for the power-law model (e.g. §59 of Peebles 1980),

$$\begin{aligned} \sigma^2_{8,AGN} &= \int\int \xi(|\mathbf{r_1}-\mathbf{r_2}|)dV_1 dV_2/V^2 \\ &= (r_{max}/r_c)^{-\gamma} J_2 \qquad (3.13) \\ J_2 &= 72/[(3-\gamma)(4-\gamma)(6-\gamma)2^{\gamma}]. \qquad (3.14) \end{aligned}$$

As discussed above, we have used the results from Fit ID's S5, S6, M5, M6, U5,and U6, where the fits are made to the angle range corresponding to

$\approx$ 1-16 h^{-1} Mpc at z_{eff}. The corresponding quantity of the underlying mass distribution at $z = 0$, σ_8 is one of the commonly used parameters in cosmology (Spergel et al. 2003). In order to compare our results with other similar works on a common ground, we calculated $\sigma_{8,AGN}$ from power-law representations from literature and plotted them versus the effective redshift of each sample[2]

For this comparison, we have used the best fit correlation lengths and slopes (r_c, γ) from literature to estimate $\sigma_{8,AGN}$ values and their 1σ errors. Since each reference has a different method of presenting results, we take the following strategy in calculating $\sigma_{8,AGN}$ and its 1σ error.

(a) If the referenced article gives confidence contours in the (r_c, γ) space, we calculate $\sigma_{8,AGN}$ values for the nominal case as well as at each point in the $L = L_{min} + 1$ contour, where L (with the best-fit value L_{min}) is either χ^2 or a statistical estimator that varies as χ^2, e.g. Cash C-estimator for Yang et al. (2006). The error range on $\sigma_{8,AGN}$ is determined by the minimum and maximum values calculated from the points along the contour.

(b) If no confidence contour in the (r_c, γ) space is given and there is a fit result with a fixed γ, we use this fit to calculate $\sigma_{8,AGN}$. The 1σ error in r_c is propagated from that of r_c.

(c) If the article gives only best-fit (r_c, γ) values, and 1σ errors on both parameters, the error of $\sigma_{8,AGN}$ has been propagated from those of r_c and γ, neglecting possible correlation of errors between these two parameters. In this case, we may well have over/under estimated the errors on $\sigma_{8,AGN}$.

If the (r_c, γ) values are given in multiple evolution models, we use the one where the correlation length is fixed in the comoving coordinates (i.e. $\varepsilon = -3 - \gamma$). At about the effective median redshift of the sample, however, the correlation lengths calculated assuming different values of ε do not differ significantly. In the case of our present work, we see this by comparing the $r_c(\bar{z}_{eff})$ values for $\varepsilon = -1.2$ and $\varepsilon = -3$ cases in Table 3.3. Also the value of $\sigma_{8,AGN}$ is insensitive to the assumed value of γ. The change of $\sigma_{8,AGN}$ is less than 0.1 between the assumed γ of 1.8 and 1.5 for our results in all the three

[2]Some authors give the median redshift of the number distribution of the X-ray sources, while we and some others give median redshift of the contribution to the clustering signal. We denote the former by $\bar{z}$ and the latter by $\bar{z}_{eff}$. We do not make a distinction between these in Fig. 3.3.

bands. The results from literature we use for this comparison and the details of the conversion to $\sigma_{8,AGN}$ are described below, roughly in order of redshift.

Grazian et al. (2004) calculated the correlation function of 392 optically-selected QSOs from the Asiago-ESO/RASS QSO Survey (AERQS) with $\bar{z}_{eff} = 0.062$. They found the nominal values of $r_c = 8.6 \pm 2.0 h^{-1}$ with $\gamma = 1.56$. Also in the low-redshift end, Akylas et al. (2000) calculated the correlation function of X-ray selected AGNs from the *ROSAT* All-Sky Survey (RASS) with a median redshift of $\bar{z} = 0.15$. Their correlation length for $\gamma = 1.8$ of $r_c = 6.5 \pm 1.0 h^{-1}$ Mpc for the Einstein de-Sitter Universe is increased by 5% to convert it to our adopted cosmology. Mullis et al. (2004) found $r_c = 7.4^{+1.8}_{-1.9} h^{-1}$ Mpc for $\gamma = 1.8$ in their *ROSAT* North Ecliptic Pole Survey (NEPS) AGNs with median redshift for the contribution to the clustering signal of $\bar{z}_{eff} = 0.22$. Basilakos et al. (2004, 2005) in their 2 deg^2 *XMM-Newton* survey, with shallower flux limits than XMM-COSMOS, found $r_c = 16.4 \pm 1.3 h^{-1}$ at $\bar{z} = 1.2$ and $r_c = 19 \pm 3 h^{-1}$ at $\bar{z} = 0.75$ for the 0.5-2 and 2-8 keV respectively. A recent work by Puccetti et al. (2006) on the central ~ 0.6 deg^2 region of the ELAIS-S1 field , covered by four mosaiced *XMM-Newton* exposures with $\sim 50 - 60$ ks each, also measured angular ACFs of X-ray point sources. For fixed $\gamma = 1.8$, they found $r_c = 12.8 \pm 4.2 h^{-1}$ at $\bar{z} = 1.0$ and $r_c = 17.9 \pm 4.8 h^{-1}$ at $\bar{z} = 0.85$ for the 0.5-2 and 2-10 keV bands respectively.

The correlation functions on the deepest X-ray surveys on the *Chandra* Deep Fields- South (CDF-S; $\bar{z} = 0.84$) and North (CDF-N; $\bar{z} = 0.96$) by Gilli et al. (2005) gave, for fixed $\gamma = 1.4$, $r_c = 10.4 \pm 0.8 h^{-1}$ and $5.1^{+0.4}_{-0.5} h^{-1}$ Mpc respectively. We use the results from their AGN samples. For all of the above samples, the errors on the $\sigma_{8,AGN}$ have been calculated using method (b).

An extensive redshift-space correlation function was made by Yang et al. (2006), who made use of the data from a combination of the CLASXS and CDF-N surveys, with a significant portion of the X-ray sources having measured spectroscopic redshifts. We have used their (s_0, γ) confidence contours, where s_0 is the redshift-space comoving correlation length, in the four redshift bins with median redshifts of $\bar{z} = 0.45$, 0.92, 1.26, and 2.07 to estimate $\sigma_{8,AGN}$ using method (a). In this conversion, we have corrected for the redshift distortion by dividing the redshift-space $\sigma_{8,AGN}$ value by $\sqrt{1.3}$ (Marinoni et al. 2005; Yang et al. 2006). Extensive clustering studies of QSOs from the 2dF QSO redshift survey (2QZ) have been made using both the projected-distance correlation function approach (Porciani et al. 2004) and the redshift-space three-dimensional correlation function approach (Croom et al. 2005). We have con-

Fit ID	$\bar{z}_{eff}$	$\sigma_{8,AGN}$	b_{AGN}	$\log M_{halo}$ $h^{-1}M_{\odot}$
Without integral constraint...				
S5	1.07	$1.58^{+.12}_{-.13}$	3.5 ± 0.3	13.6 ± 0.1
M5	0.87	$0.82^{+.27}_{-.42}$	$1.7^{+0.6}_{-0.9}$	<13.3
U5	0.60	$1.42^{+.26}_{-.32}$	$2.6^{+0.5}_{-0.6}$	13.5 ± 0.2
With integral constraint...				
S6	1.07	$1.63^{+.13}_{-.14}$	3.7 ± 0.3	13.6 ± 0.1
M6	0.87	$1.19^{+.34}_{-.50}$	$2.5^{+0.7}_{-1.0}$	$13.3^{+0.3}_{-0.7}$
U6	0.60	$2.08^{+.34}_{-.41}$	$3.8^{+0.6}_{-0.8}$	13.9 ± 0.2

Table 3.4: Estimated $\sigma_{8,AGN}$ and Bias. The error on M_{halo} reflects the statistical error on b_{AGN} only.

verted the nominal r_0–γ values and confidence contours in three redshift bins at $\bar{z}_{eff}$=1.06, 1.51 and 1.89 by Porciani et al. (2004) to $\sigma_{8,AGN}$ using method (a). In converting the Croom et al. (2005)'s (s_0, γ) results in 10 redshift bins ranging from $z = 0.5$ to 2.5, we have used method c) and the redshift distortion correction has been made in the same way as we have done to the Yang et al. (2006) results.

Figure 3.3(a) shows the $\sigma_{8,AGN}$ values as a function of the look back time $\tau(z)$ for our default cosmology from our analysis results both without and with integral constraints. We also overplot $\sigma_{8,AGN}$ values calculated from the results found in literature as detailed above. In order to compare them with those of the underlying mass distribution, we have also plotted the $\sigma_8 D(z)$ values from the linear theory (e.g. Carroll et al. 1992; Hamilton 2001), normalized to 0.75 at z=0 (Spergel et al. 2003) [3] This curve has been shown to accurately represent the distribution of dark matter particles in the ΛCDM Hubble Volume Simulation (Marinoni et al. 2005). Figure 3.3(b) shows the inferred bias parameters $b_{AGN} = \sigma_{8,AGN}(z)/[\sigma_8 D(z)]$. The values of $\sigma_{8,AGN}$ and b_{AGN} from this work are shown in Table 3.4.

[3]We use the latest value of σ_8 as of writing this paper obtained from http://lambda.gsfc.nasa.gov/product/map/current/parameters.cfm for the Λ-CDM model derived from all datasets.

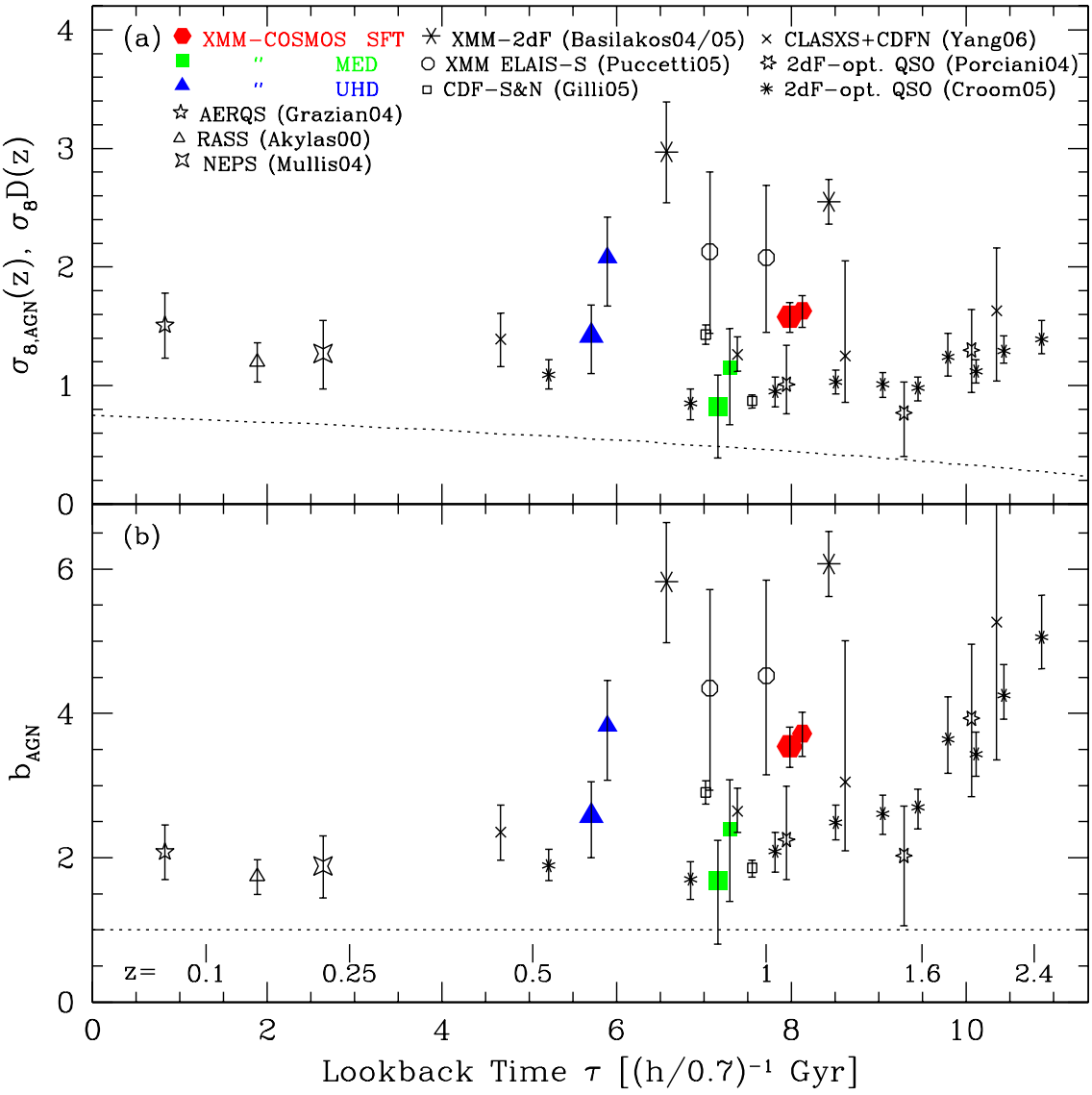

Figure 3.3: (a) The $\sigma_{8,AGN}$ of the X-ray sources/AGNs inferred by the power-law fits to the correlation functions from this work and literature are plotted against the look back time corresponding to the effective median redshift of the samples. All error bars indicate 1σ errors. The results from this work for the fits without integral constraints are plotted with large solid symbols as labeled (with colors in the electronic version). Those with integral constraints are also plotted with smaller symbols and at positions slightly shifted rightward for visibility. The $\sigma_{8,AGN}$ values calculated from various results in literature (see text) are also shown as labeled (abbreviated as the first author followed by the last two digits of the publication year). The dotted line shows $\sigma_8 D(z)$ for the mass in the linear theory normalized to 0.9 at z=0. (b) The bias parameter $b_{AGN} = \sigma_{8,AGN}(z)/[\sigma_8 D(z)]$ are plotted as a function of the effective redshift. The meaning of the symbols are the same as panel (a). We also show redshift ticks at the bottom part of the figure.

3.5 Discussion and Prospects

In this work, we used all the point sources above the scaled sensitivity threshold without further classification of the sources. We analyzed our results assuming that all the X-ray sources are AGNs. This, in practice, is a good approximation. Our preliminary identifications of the sources indicate that out of the 1037, 545, and 151 sources selected for the ACF analysis for the SFT, MED and UHD bands, 20, 5, and 1 are apparent Galactic stars respectively. Removing these sources from the analysis changed the results very little. Also our results are not likely to be heavily affected by the contamination of clusters/groups, since these sources are extended by > 20" (e.g. Finoguenov et al. 2007) and are likely to be classified as extended by the source detection procedure, hence removed from our sample.

As seen in Fig. 3.3, with an exception of the MED band, our analysis without integral constraints gives somewhat larger $\sigma_{8,AGN}$ values than those obtained from results using 2dF optically selected QSOs by Porciani et al. (2004) and Croom et al. (2005), but in general agreement with the values from *Chandra* CLASXS+CDF-N by Yang et al. (2006), CDF-S by Gilli et al. (2005), and *XMM-Newton* results from Puccetti et al. (2006). Most likely due to the cosmic variance over a small FOV, Gilli et al. (2005)'s result on CDF-N gave a significantly smaller correlation amplitude than their own CDF-S values as well as our results. The angular ACFs from a shallower *XMM-Newton* survey by Basilakos et al. (2004, 2005) gave significantly larger $\sigma_{8,AGN}$ values than other works in both 0.5-2 keV and 2-8 keV bands. The reason for their distinctively large value is unclear.

One of the interesting questions in investigating clustering properties of X-ray selected AGNs is to investigate whether there is any difference in the environments of obscured and unobscured AGNs. Applying the population synthesis model of Ueda et al. (2003) to our sensitivity maps, only $\sim 17\%$ of the sources detected in the SFT band at $z \sim 1$ are obscured AGNs with $N_H > 10^{22}$ cm^{-2}. The fraction increases to $\sim 40\%$ in the MED and UHD bands. A comparison of bias parameters between SFT band and MED band, which have similar $\bar{z}_{eff}$ values, seems to show a lager bias parameter for the SFT sample. However, with the combination of statistical uncertainties and uncertainties in modeling the integral constraint in the MED band, we can only conclude that the bias of the obscured AGNs is not stronger than that of unobscured AGNs. In other works, Gilli et al. (2005); Yang et al. (2006) as well as Basi-

lakos et al. (2004, 2005) did not find any statistical difference between the clustering properties of these two. Further studies involving the second-year XMM-COSMOS data, which in effect doubles the *XMM-Newton* exposure over the COSMOS field, and redshift information of individual objects will probe into this problem further. Also with the accepted C-COSMOS program totaling 1.8 Ms of *Chandra* exposure, we will be able to probe the correlation functions to a much smaller scale, enabling us to investigate the immediate neighbor environments of these AGNs.

Our measured bias parameters based on the rms fluctuations in the $8h^{-1}$ Mpc radius sphere are in the range $b_{AGN} = 1.5 - 4$. The clustering properties of dark matter halos (DMH) depend on their mass (Mo & White 1996; Sheth et al. 2001) and we can estimate the typical mass of the DMHs in which the population of AGNs represented by our sample reside, under the assumption that the typical mass halo is the main cause of the AGN biasing. Following the approach of Yang et al. (2006) and Croom et al. (2005) who utilized the model by Sheth et al. (2001), we roughly estimate that the typical mass of DMH is $\sim 10^{13} - 10^{14}\ M_{odot}$ for our SFT and UHD samples (see Table 3.4). These are an order of magnitude larger than those estimated by Porciani et al. (2004) and Croom et al. (2005), probably reflecting the large bias parameters from our results.

One of the largest uncertainties in our analysis lies in the treatment of the integral constraint, because its effect is not negligible in our case compared with the ACF amplitudes in the range of our interest. Fig. 3.3 shows that our results based on the fits with integral constraint, under an assumption that the fitted power-law behavior of the underlying $w(\theta)$ extends to the scale of the entire FOV, give a somewhat larger correlation amplitudes. This assumption may not be true. Also the apparent negative $w_{est}(\theta)$ values at $\theta > 30^{'}$ in Fig. 3.1 may well be caused by remaining systematic errors. Thus the interpretation of the angular correlation functions, where the signals are diluted by the projection along the redshift space, has a major limitation in correctly taking the integral constraint into account.

The situation will improve when redshift information on individual X-ray sources becomes available for a major and comprehensive set of the X-ray sources. This will enable us to calculate three-dimensional correlation functions or projected-distance correlation functions in a number of redshift bins. With the line-of-sight dilution effect suppressed, we will be able to obtain a larger amplitude in the correlation signal, making the analysis much less

subject to the uncertainties in the integral constraint. With optical followup programs underway on the COSMOS field through Magellan and zCOSMOS projects (Trump et al. 2007; Lilly et al. 2007), we are obtaining spectroscopic redshifts from a major fraction of the X-ray sources. At the time of writing this paper, we have been able to define a sample of 378 XMM-COSMOS detected AGNs with measured spectroscopic redshifts ($\sim 30\%$ of the X-ray point sources), with a median redshift of $z \sim 1$. Our preliminary analysis of these sources based on the projected-distance correlation function $w(r_p)$ gives a comoving correlation length of $r_c \approx 8h^{-1}$ Mpc and $\gamma = 1.6$, which is fully consistent with our results without the integral constraints. The results of a full analysis utilizing the redshift information will be presented in a future paper (Gilli et al. in preparation).

Extensive multi-wavelength coverage and the availability of a galaxy catalog also enables us to investigate the cross-correlation function between X-ray selected AGNs and galaxies. By cross-correlating the X-ray selected AGNs with three orders of magnitude larger number of galaxies, we will be able to investigate the environments of the AGN activity in various redshifts with much better statistics.

3.6 Conclusions

We have presented the first results on the angular correlation functions (ACFs) of the X-ray selected AGNs from the XMM-COSMOS survey and reached the following conclusions.

1. A significant positive angular clustering signals has been detected in the 0.5-2 (SFT) bands in the angle range of $0.5'$-$20'$, while in the 2-4.5 (MED) and 4.5-10 keV (UHD) bands, the positive signals are 2 and 3σ respectively. The robustness of the estimated correlation functions has been verified using different methods of generating random samples.

2. Power-law fits to the angular correlation function have been made, taking into account the correlation of errors. Correctly taking the integral constraint into account is a major limitation on interpreting the angular correlation function. For fits with fixed $\gamma - 1 = 0.8$ and without (with) the integral constraint term, we found correlation lengths of $\theta_c = 1.9" \pm 0.3"$, $0.8"^{+0.5"}_{-0.4"}$ and $6" \pm 2"$ ($3.1" \pm 0.5"$, $2.2" \pm 1.0"$, and $14" \pm 5"$) for the SFT, MED, and UHD bands respectively.

3. Due to *XMM-Newton* PSF, most of the source pairs closer than ~ 20" are classified as single extended sources, and therefore excluded from the sample. This causes a bias in angular correlation function measurements. We have estimated this effect (the PSF merging bias) by simulations and found that the estimated ACF underestimates the amplitude of the true underlying ACF by $\sim 15\%$ and $\sim 8\%$ for the SFT and MED bands respectively

4. Using Limber's equation and the expected redshift distributions of the sources, we have found comoving correlation lengths of $r_c \approx 9.8 \pm 0.7$, $5.8^{+1.4}_{-1.7}$, and 12 ± 2 h^{-1} Mpc for $\gamma = 1.8$ at the effective redshifts of $\bar{z}_{eff} \approx$ 1.1, 0.9, and 0.55 for the SFT, MED, and UHD bands respectively for the fits without integral constraints, while 20%-90% larger correlation lengths have been obtained for the fits with integral constraints.

5. Using the fits in the angles corresponding to a projected distance range of 1-16h^{-1} Mpc at the effective median redshift of the sample, we have calculated the rms fluctuations of the X-ray source distributions. Comparing them with that of the mass distribution from the linear theory, we find that the bias parameters of the X-ray sources are in the range $b_{AGN} = 1.5 - 4$ at $0.5 < z < 1.2$.

6. If the bias mainly reflects the typical mass of dark matter halos in which these X-ray AGNs reside, their typical masses are $10^{13} - 10^{14}$ $M_{\odot}$.

7. Further investigations utilizing redshifts of individual X-ray sources and/or involving cross-correlation function with galaxies taking advantage of the wealth of multiwavelength data are being conducted. The approved Chandra observations (C-COSMOS) on this field will enable us to probe into the clustering in much smaller scales and therefore into immediately neighboring environments of AGNs.

Chapter 4

The Soft X-ray Cluster-AGN cross-correlation function in the NEP survey

4.1 Introduction

The current paradigm of galaxy formation assumes that all types of galaxies reside in dark matter (DM) haloes, and that the properties of these haloes determine to some extent the properties of the galaxies inside (White & Rees 1978). In general, the clustering amplitudes of haloes depend on halo mass. The relation between the clustering properties of both DM haloes and galaxies (biasing) should thus tell us something about the physical processes leading to the formation and evolution of galaxies. Clusters of galaxies are the highest peaks in the global mass distribution of the Universe and should follow a direct and simple biasing scheme – mainly related to the underlying primordial Gaussian random field (Kaiser 1987). A simple means towards a better understanding of galaxy biasing is thus provided by studies of the relative biasing between galaxies and clusters of galaxies.

As a first step, previous investigations estimated two-point statistics like the *auto*-correlation function (Mullis et al. 2004, Gilli et al. 2005, Basilakos et al. 2005, Yang et al. 2006). They could show that, in fact, AGNs trace the underlying cosmic large-scale structure. In addition, the large-scale structure of X-ray selected galaxy clusters could be studied in some detail (e.g. Schuecker et al. 2001), but without investigating the link between clusters and AGNs.

In the present paper, we concentrate on the study of the relative clustering

between X-ray selected AGNs and galaxy clusters. Our work improves on most previous work on the large-scale structure of X-ray selected AGNs in two important aspects. First, with the exception of Mullis et al. (2004), our sample is the only one that is spectroscopically complete (99.6%). Gilli et al. (2005) used the CDFS (35%) and the CDFN (50%). The Basilakos et al. (2005) sample had almost no spectroscopic redshifts. Yang et al. (2006) used the CLASXS sample (52% complete) and the CDFN (56% complete).

Another motivation for our work is that over the last several years, X-ray observations revealed that a significant fraction of high-z clusters of galaxies show overdensities of AGNs in their outskirts (i.e. between 3 h_{70}^{-1} Mpc and 7 h_{70}^{-1} Mpc from the center of the cluster) (Henry et al., 1991; Cappi et al., 2001; Ruderman & Ebeling 2005, Cappelluti et al., 2005, and references therein). These overdensities were however detected in randomly selected archive targeted observations of galaxy clusters. While these overdensities are highly significant (up to 8σ) when compared to cluster-free fields, the incompleteness of the samples does not allow drawing any conclusion about the average clustering properties of AGNs around clusters. The majority of the sources making these overdensities have no spectroscopical identification and therefore any information on their spatial clustering is lost. More recently Branchesi et al. (2007) showed that at high-z the source surface density of AGNs significantly increases even in the central regions of the clusters. These results imply that further progress will come from studying the three dimensional spatial distribution of AGNs around clusters. A natural way to characterize this specific type of clustering is given by the three-dimensional *cross*-correlation of AGNs and galaxy clusters, the computation of which needs complete redshift information for all objects, which is rare in X-ray surveys.

In this respect, the *ROSAT* North Ecliptic Pole (NEP) survey (Henry et al., 2001,2006; Voges et al., 2001) is one of the few X-ray surveys covering a sufficiently large volume with an almost complete follow-up identification of AGNs and clusters (i.e. 440 sources spectroscopically identified of 442 detected). This survey thus provides a very useful basis for more precise investigations of the relative clustering properties of these two types of objects.

We organized the present paper in the following way. In Sect. 2, we describe the ROSAT-NEP survey data which we use for our investigations of the spatial distribution of X-ray selected AGNs and galaxy clusters. For the statistical analysis we estimate their cross-correlation. A useful estimator for this statistic and the mock samples needed for its determination are described

in Sects. 3 and 4, respectively. The results are presented in Sect. 5, and are discussed in Sect. 6. In this paper, we assume a (concordance) Friedmann-Lemaitre Universe characterized by the Hubble constant given in units of $h_{70} = H_0/(70\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$, the normalized cosmic matter density $\Omega_\mathrm{m} = 0.3$, and the normalized cosmological constant $\Omega_\Lambda = 0.7$. Unless otherwise stated, errors are reported at the 1σ confidence level.

4.2 The data

The *ROSAT* NEP survey covers a region of 9×9 deg^2 around the North Ecliptic Pole ($17^h15^m < \alpha < 18^h45^m$, $62^o < \delta < 71^o$) observed with the PSPC proportional counter as part of the *ROSAT* All Sky Survey (Henry et al., 2001; Voges et al., 2001) with a flux limit of 2×10^{-14}erg cm^{-2} s^{-1} in the 0.5–2 keV energy band. 442 X-ray sources were detected and 440 optically identified. Spectroscopic redshift information is available for 219 AGNs and 62 clusters of galaxies. The clusters have redshifts $z \leq 0.81$ with a median of 0.18 and the AGNs have $z \leq 3.889$ with a median of 0.4 (Fig. 4.1). For the purpose of this work we selected all the clusters and the 185 AGNs with $z \leq 1$ (Gioia et al. 2003).
Such a dataset was used also by Mullis et al. (2004) for the calculation of the 3-D auto correlation function of X-ray selected AGNs. Mullis and collaborators find significant clustering on scales smaller than $\sim 43\ h_{70}^{-1}$ Mpc with a correlation length of $\sim 10.4\ h_{70}^{-1}$ Mpc, and a slope of the correlation best-fit power law of γ=1.8.

4.3 Cluster-AGN spatial cross-correlation

The cross-correlation function ξ_{CA} of clusters and AGNs is defined by the joint probability to find, at a distance r, one cluster in the infinitesimal comoving volume element δV_C and one AGN in the comoving volume element δV_A,

$$\delta P = n_C\, n_A\, [1 + \xi_{CA}(r)]\, \delta V_C\, \delta V_A, \tag{4.1}$$

where n_C and n_A are the mean comoving number densities of clusters and AGNs, respectively. In calculating the differential cross-correlation in redshift space we used an adapted version of the Landy–Szalay estimator (Landy &

Szalay, 1993; see also e.g. Blake et al., 2006),

$$\xi_{CA}(s) = \frac{D_C D_A - R_C D_A - R_A D_C + R_A R_C}{R_A R_C}, \tag{4.2}$$

where $D_C D_A$, $R_C D_A$, $R_A D_C$ and $R_A R_C$ are the normalized number of pairs in the i-th redshift space separation s bin for the clusters data-AGNs data, cluster random-AGNs data, AGNs random-clusters data and clusters random-AGNs random samples, respectively. Using the symbols D and R to represent the data and random samples, respectively, and C and A to identify clusters and AGNs, respectively, the normalized pairs are expressed by

$$D_C D_A = \frac{n_{pair,D_A D_C}(s_i)}{(N_{D_A} \times N_{D_C})},\ R_C D_A = \frac{n_{pair,R_C D_A}(s_i)}{(N_{R_C} \times N_{D_A})},$$
$$R_A D_C = \frac{n_{pair,R_A D_C}(s_i)}{(N_{R_A} \times N_{D_C})},\ R_A R_C = \frac{n_{pair,R_A R_C}(s_i)}{(N_{R_A} \times N_{R_C})}.$$

Here, $N_{D_A}, N_{D_C}, N_{R_A}, N_{R_C}$ are the total numbers of AGNs and clusters in the data and in the randomly generated samples, respectively. The quantities n_{pair} represent, adopting the symbolism used above, the actual number of pairs measured in the random and data samples as a function of the redshift space separation s_i. The distances were computed assuming for the position of the clusters the centroid of the X-ray emission while for AGNs the optical positions were adopted. In order to have a good signal-to-noise ratio (SNR) the data were grouped in logarithmic bins of $\Delta\log(sh_{70}^{-1}\ \mathrm{Mpc})$=0.15.

4.4 Random Samples

Equation 4.2 indicates that an accurate estimate of the distribution function of the random samples is crucial in order to obtain a reliable estimate of ξ_{CA}. Several effects must be taken into account when generating a sample of objects in a flux limited survey. Simulated AGNs were randomly placed within the *ROSAT* NEP survey area. In order to reproduce the flux distribution of the real sample, we followed the method of Mullis et al. (2004). In practice since the cumulative AGNs logN-logS source counts distribution can be described by a power law, $S = kS^{-\alpha}$, with $\alpha = 1.3$, the differential probability scales as $S^{-(\alpha+1)}$. Using a transformation method (e.g. Press et al. 1986, chapter 7) we see that the random flux above a certain X-ray flux S_{lim} is distributed as $S = S_{lim}(1-p)^{\frac{1}{\alpha}}$, where p is a random number uniformly distributed between 0 and 1 and S_{lim}=2$\times 10^{-14}$ erg cm^{-2} s^{-1}, i.e. the flux limit of the NEP survey.

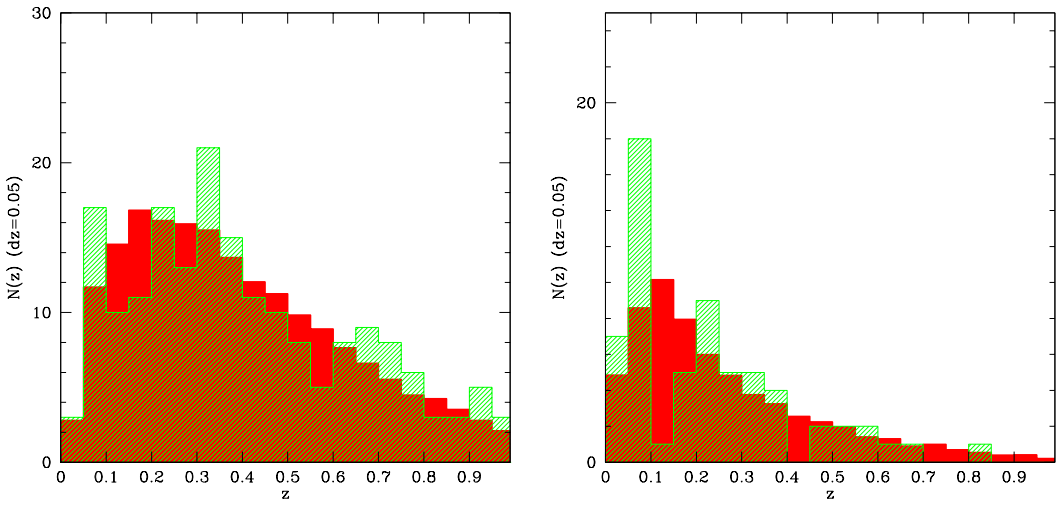

Figure 4.1: *Top Panel*: The redshift distribution of the AGNs in the NEP survey (*shaded histogram*) and of the randomly generated AGNs with the same selection effects (*filled histogram*). *Bottom Panel*: The redshift distribution of the galaxy clusters in the NEP survey (*shaded histogram*) and of randomly generated clusters (*filled histogram*). The deviations at low z between the data and the random sample are mainly caused by the NEP supercluster of galaxies (Mullis et al., 2001).

All random AGNs with a flux lower than the flux limit map (see Fig. 4 in Henry et al., 2006) at the source position were excluded. In order to assign a redshift to these "sources" we computed the predicted redshift distribution at the position of each accepted source. Once the flux limit at the position where the source was randomly placed is known, and denoting with $\phi(L,z)$ the luminosity function, then the number of sources per redshift interval dz is given by

$$N(z)dz = \int_{L_{min}}^{\infty} \int_{z}^{z+dz} \phi(L,z)\, dV(z)\, dL\, dz, \tag{4.3}$$

where L_{min} is the minimum luminosity observable at redshift z with a local flux limit S_{lim} and $dV(z)$ the differential comoving volume element. The k-correction does not play any role in the calculations since we assumed an average spectral index Γ=2 (as in Mullis et al. 2004) for all AGNs. For the luminosity function $\phi(L,z)$ we took the luminosity-dependent density evolution (LDDE) best fit model of Hasinger et al. (2005). A redshift was then randomly assigned to each source via Monte Carlo integration of the predicted redshift

distribution. For galaxy clusters we applied the same procedure assuming a slope of the logN-logS distribution of α=1.3, as luminosity function an A-B evolving Schechter model (Rosati et al. (2002)) with the parameters obtained by Mullis et al. (2004b) and using a sensitivity map specific for NEP clusters (Henry et al. (2006)). Since for clusters of galaxies the k-correction is not negligible, we assigned to the random clusters an intrinsic spectrum according to a MEKAL spectral model with a plasma temperature kT=3 keV and a metallicity $Z = 0.3Z_{\odot}$. We also applied to L_{min} a "size-correction" according to the results of Henry et al. (2006) in order to compensate for the missing flux in the X-ray photometry aperture and the variation of the angular dimensions of the object with z. Such a procedure was repeated until we populated the survey volume with 37200 random AGNs and 12600 random clusters (i.e. 200 times more objects than in the real data sample). The redshift distribution of the random cluster and AGN samples are plotted in Fig. 4.1 together with the real data. To obtain a realistic estimate of the uncertainties of the cross correlations we used the bootstrap resampling technique described by e.g. Ling, Frenk and Barrow (1986).

4.5 Results

We present the spatial cross-correlation function between clusters and AGNs in Fig. 4.2. A positive clustering signal is detected in the distance interval $s \leq 50\ h_{70}^{-1}$ Mpc. In order to test the strength of the clustering we performed a canonical power-law fit,

$$\xi_{CA}(s) = \left(\frac{s}{s_0}\right)^{-\gamma}, \tag{4.4}$$

with s_0 and γ as free parameters. The fit can be done using the coarsely binned data in Fig. 4.2 and minimizing the χ^2 statistics. However, this approach is extremely sensitive to the size and distribution of the bins. In order to overcome this problem we performed a standard maximum likelihood power-law fit to unbinned data. The comoving separation s was parsed in very small bins so that there are either 0 or 1 data pairs for any given interval. In this regime Poisson probabilities are appropriate. In order to perform the fit, we need to find the predicted probability distribution of the cluster-AGN pairs for each value of γ and s_0. We calculated the number of predicted pairs by replacing $\xi_{CA}(s)$ in Eq. 4.2 with the model given by Eq. 4.4 and using $D_C D_A(s)$ (hereinafter $\lambda(s)ds$) as variable. We can then use the separations of all the

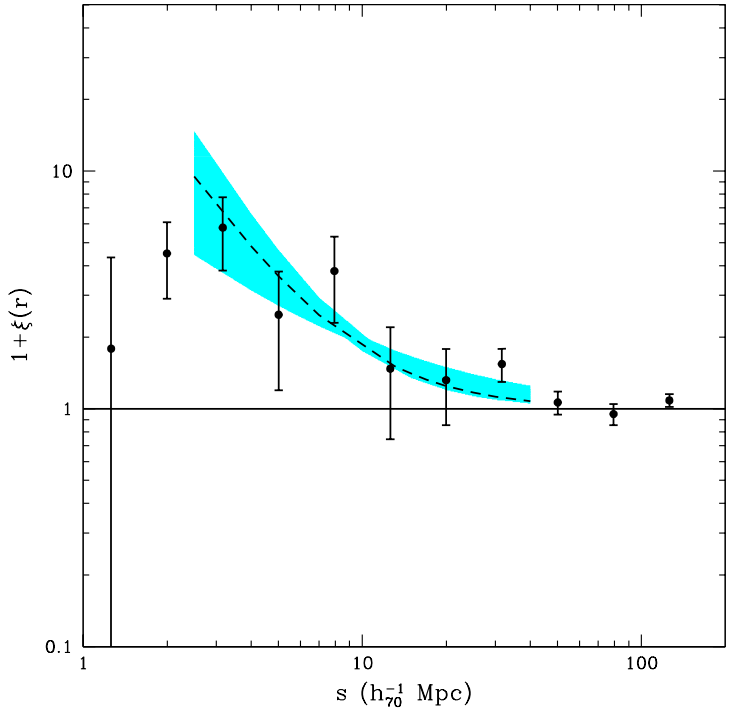

Figure 4.2: The Cluster-AGN soft X-ray cross correlation function plus one. The error bars are quoted at 1σ level. The dashed line represents the best fit maximum-likelihood power-law fit s_0=8.7$^{+1.2}_{-0.3}$ h_{70}^{-1} Mpc and γ=1.7$^{+0.2}_{-0.7}$. The shaded region illustrates the 1σ confidence region of the power-law fit in the distance range in which it was performed.

N cluster-AGN pairs to form a likelihood function. This is defined as the product of the probabilities of having exactly one pair in the interval ds at each separation s_i of the N pairs times the probability of having no pairs in all the other differential intervals. This for all s in a chosen range $(s_a - s_b)$, in our case where $\xi_{CA}(s)$ can be reasonably represented by a power law. Assuming Poisson probabilities we thus obtain the likelihood

$$\mathcal{L} = \prod_i^N \lambda(s_i)ds \exp \sum_{j \neq i} \lambda(s_j)ds, \tag{4.5}$$

where $\lambda(s_i)ds$ is the expected number of pairs in the interval ds, and the index j runs over all the elements ds which do not contain pairs. We then define the usual quantity $S = -2\ln\ \mathcal{L}$ and drop the terms independent of model parameters (see e.g. Schuecker & Boehringer (1998); Croft et al. (1997)) leading

to

$$S = 2\int_{s_a}^{s_b} \lambda(s)\, ds - 2\sum_{i}^{N} \ln[\lambda(s_i)]. \tag{4.6}$$

In order to check in which range of separations we can conduct our analysis, we transformed the angular separations between clusters and AGNs into an average redshift space separation. We assumed that all the sources were at $< z >$=0.38 (i.e. the median redshift of the cluster and AGN sample). The result is plotted in Fig. 4.3. As one can see there are no real pairs with separation $< 2\ h_{70}^{-1}$ Mpc. For this reason we decided not to consider points at separation lower than 2.5 h_{70}^{-1} Mpc. These points in the cross-correlation function are mainly introduced by the parameters $R_C D_A$, $R_A D_C$ in Eq. 4.2 which could have smaller separations than the real data since the random sample includes neither the extended emission of the clusters nor the broadening due to the PSF of pointlike sources. This prevents us also from overestimates of the correlation length introduced by the amplification bias (see e.g Vikhlinin & Forman (1995)). For this reason and since on scales larger than 50 h_{70}^{-1} Mpc

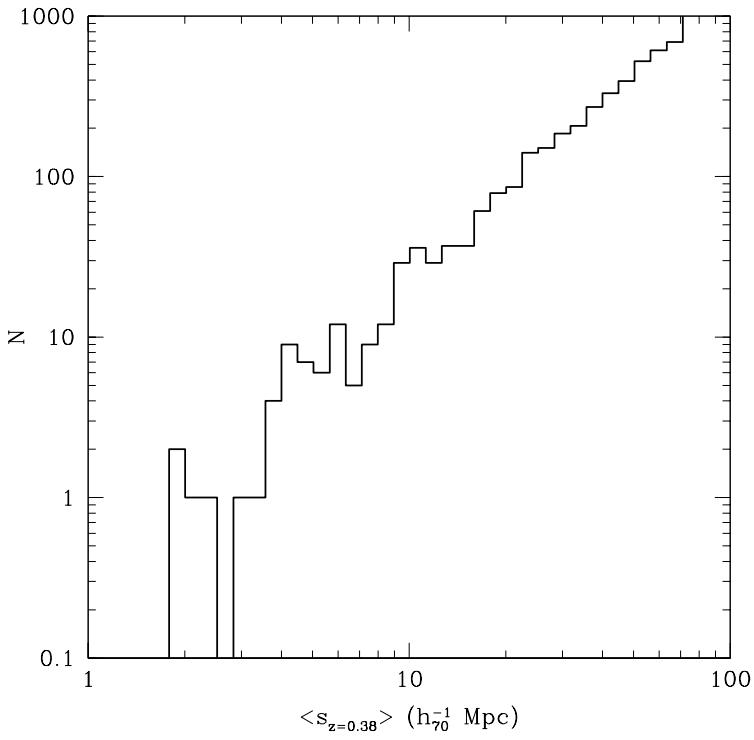

Figure 4.3: The distribution of the angular radial separations translated into redshift space distances at $< z >$=0.38

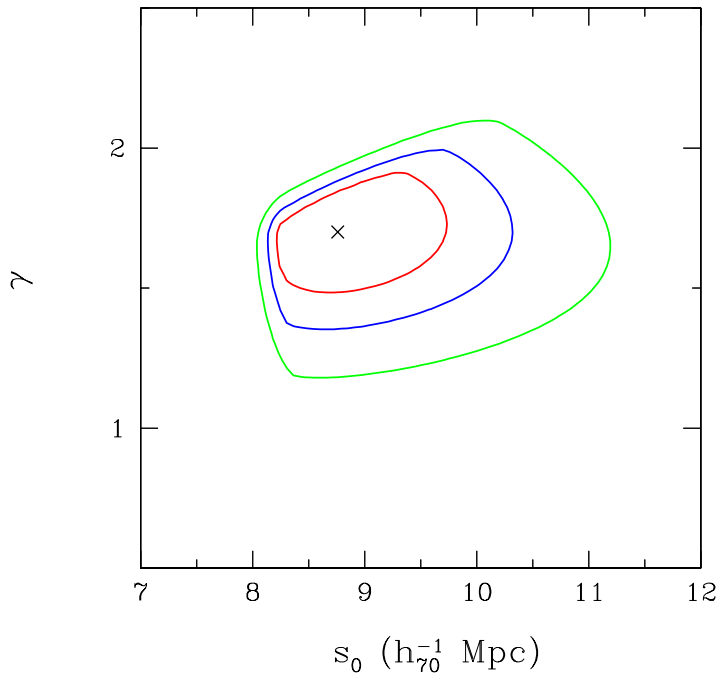

Figure 4.4: The 1σ, 2σ and 3σ confidence contours in the s_0, γ space for the power-law fit to $\xi_{CA}(s)$ for two interesting parameters. The cross represents the best fit values.

there is no evidence of signal, the fit was performed over the distance range 2.5–50 h_{70}^{-1} Mpc. In Fig. 4.4 we show the results of the maximum-likelihood power-law fit to $\xi_{CA}(s)$ for the *ROSAT* NEP survey. The 1, 2 and 3σ were obtained at ΔS levels of 2.3, 6.2 and 11.8 from the minimum value of S. The best fit parameters obtained are s_0=$8.7^{+1.2}_{-0.3}$ h_{70}^{-1} Mpc and γ=$1.7^{+0.2}_{-0.7}$ where the uncertainty is at the 1σ confidence level. With γ fixed to 1.8 (i.e. a typical value found in galaxy-galaxy correlation function) we find $s_0 \sim 8.5 h_{70}^{-1}$ Mpc, a similar value was obtained by extending the fitting region to 60 h_{70}^{-1} Mpc and restricting it to the 2.5–40 h_{70}^{-1} Mpc.

The integral constraint (Peebles (1980)), which is a systematic shift in correlation functions introduced by the limited volume observed, was computed following the prescription of Roche et al. (1993). This can be obtained numerically with a fit by assuming that the correlation function is represented by a power law of fixed index (here we used γ=1.7) on all scales sampled by the survey. The underestimate of $\xi_{CA}(s)$ due to the integral constraint is found to

be of order 2%.
In order to determine the stability of these results the procedure was repeated first by separating the field in two subfields twice, the North and South, and West and East parts of the survey. The fluctuations due to sample variance are found to be smaller than the typical amplitude of the uncertainties. A similar result is obtained by recomputing the $\xi_{CA}(s)$ for clusters with $z \leq 0.18$ (i.e. the cluster median redshift) and $z > 0.18$. The dependence of $\xi_{CA}(s)$ on the cluster X-ray luminosity (L_X) was evaluated dividing the cluster sample into two subsamples with $L_X \leq (or \geq) 3.8 \times 10^{43}$ h_{70}^{-2} erg s^{-1}. Though no significant L_X dependent behaviour was detected, we cannot yet conclude that there is no luminosity or redshift dependent cross-correlation length because of the low-statistics of the sub-samples. The stability of the result was also checked fitting the data with the likelihood estimator used by Mullis et al. (2004) returning no significant deviations from our results at 1-2 σ level.

4.6 Discussion

We presented here the first direct evidence of spatial clustering of soft X-ray selected AGNs around X-ray selected clusters of galaxies. Indirect evidence was presented by Henry et al. (1991), Cappi et al. (2001), Cappelluti et al. (2005) (and references therein). These authors found significant X-ray point source overdensities (about a factor 2) around distant clusters of galaxies when compared to cluster-free fields. If the overdensities were at the cluster redshift they would arise at scales smaller than $\sim 7\ h_{70}^{-1}$ Mpc. Since the correlation function is proportional to $\left(\frac{\delta\rho}{\rho}\right)^2$, a ξ_{CA}=1 implies an overdensity of a factor 2 with respect to a randomly distributed field. We can conclude that, since the correlation length found in this work reflects the scale of the overdensities known up to now, we observe a physical overdensity (of at least a factor 2) of AGNs around clusters between 2 and $\sim 8\ h_{70}^{-1}$ Mpc from the center of the clusters.
Because of the shallowness of the NEP survey, the AGN surface density (i.e. $<$30 deg^2 in the central region) does not allow detection of such a correlation via overdensity analysis since it would be dominated by small number statistics. In fact, from our results we expect to detect AGNs overdensities on scales $<$7-8 h_{70}^{-1} Mpc from the center of clusters. At $< z > \sim 0.18$ (i.e. the median z of the cluster sample of the NEP survey) these overdensities arise

on scales of $\sim$0.6 deg^{-2} which are easily resolved by the NEP survey. However to significantly detect these overdensity on single clusters, a conspicuous number of sources is necessary to disentangle real overdensities from shot noise. This problem could be easily resolved by high angular resolution telescopes like *Chandra* and partially by XMM-*Newton*. In this direction deep and wide [1] *Chandra* and XMM-*Newton* surveys like COSMOS, which will return an AGN surface density of up to $\sim$ 2700 deg^{-2}, would allow seeing in a 0.015 deg^2 region (i.e. the size of an ACIS-I chip) a population of at least 40 AGNs belonging to the cluster environment (i.e. assuming an average overdensity of a factor 2, we expect $\sim$40 sources belonging to the cluster and $\sim$40 to the background). As an example, at the limiting flux of the C-COSMOS survey, the AGN population of a cluster at z=1 would be observed with a 0.5–2 keV limiting luminosity $L_{min} \sim 10^{42}$ erg s^{-1}. At z=1 the size of an ACIS chip (i.e. 8 arcmin) corresponds to a linear dimension of $\sim 4\ h_{70}^{-1}$ Mpc. Having 40 AGNs in a sphere with this radius, corresponds to a space density of AGN with $L_X > L_{min}$ of $\sim 0.15\ h_{70}^3$ Mpc^{-3}. We can however state that, according to the result presented here, clusters of galaxies could be detected by AGN overdensities (rather than galaxy overdensities) if the depth of the survey would provide an AGN surface density sufficient to overcome the Poisson noise on the AGN number. In general in order to understand the galaxy evolution in dense environments the measure of cross correlation between clusters and different kind of galaxies is an important tool. We already know that infrared dusty galaxies avoid dense environments therefore showing a large cross-correlation length and weak clustering signal in the small separations region (Sánchez et al., 2005). We also know that blue galaxies avoid low-z rich clusters cores (Butcher & Oemler, 1984). It is therefore important to compare the cluster-AGN cross correlation length to that of cluster and different galaxy type. Mo et al. (1993) computed the cross-correlation function of Abell clusters and QDOT IRAS galaxies. They found an average correlation length and a slope in agreement with the results presented here. Moreover Lilje & Efstathiou (1988) showed that the cross-correlation function of Abell clusters with Lick galaxies is positive on scales $\leq 29\ h_{70}^{-1}$ Mpc with a slope $\gamma \sim 2.2$ and a correlation length of $\sim$ 12.6 h_{70}^{-1} Mpc. These results are also in agreement within 1 σ with our findings on AGNs. These first comparisons already sug-

[1] The expected fluxes limit for the C-COSMOS and XMM-COSMOS survey are 1-2$\times 10^{-16}$ erg cm^{-2} s^{-1} and $\sim 6\times 10^{-16}$ erg cm^{-2} s^{-1}, respectively. These surveys will cover 0.9 deg^2 and 2.1 deg^2, respectively

gest that AGNs are clustered around galaxy clusters just like galaxies. As a final check we compared our ξ_{CA} to the X-ray cluster-galaxy cross-correlation function (hereinafter CGCCF) computed by Sánchez et al. (2005). They used the X-ray selected clusters of the REFLEX survey (Böhringer et al. 2002) and the galaxies from the APM survey (Maddox et al. (1990)) limited to b_j=20.5 mag. They found that the CGCCF behaves like a broken power-law with a cut-off distance of $\sim 2\ h_{70}^{-1}$ Mpc with a steeper slope at small distances. We can define the following approximate biasing relations:

$$\xi_{CA}(s) = b_C b_A \xi_\rho(s),\ \ \xi_{CG}(s) = b_C b_G \xi_\rho(s). \tag{4.7}$$

Here $\xi_\rho(s)$ is the autocorrelation of matter, b_G, b_A and b_C are the bias factors relative to galaxies, AGNs and clusters, respectively. By dividing the two equations we can then derive $\frac{b_A}{b_G}(s)$. In order to perform this operation several effects must be taken in account.

- The bias of REFLEX clusters could be slightly different from that of NEP clusters since they are differently distributed in redshift and the surveys have different limiting fluxes and X-ray luminosity distributions. Schuecker et al. (2001) showed with the power spectrum that even having a large sample of clusters as REFLEX the error on the bias determination is still high. Since in the NEP survey we expect an even higher uncertainty it is a reasonable approximation to consider the ratio $\frac{b_{NEP}}{b_{REF}}$ consistent to 1.

- Sánchez et al. (2005) computed ξ_{CG} in real space while we work in redshift space. In order to evaluate the effect of redshift space distortion on ξ_{CG} we used the results of Croft et al. (1999). Their Figs. 3 and 5 show ξ_{CG} computed both in redshift and real space. From that work we estimate that our relative bias is affected by a $\sim$30% overestimate below 10 h_{70}^{-1} Mpc and of $\sim$10% between 10 h_{70}^{-1} Mpc and 20 h_{70}^{-1} Mpc. Their work also indicates that ξ_{CG} does not depend on the richness of the clusters used in the calculation and that the correction for the scale independent biasing on large scales (Kaiser (1987)) can be neglected when compared to the size of our uncertainties.

The ratio $\frac{b_A}{b_G}(s)$ is plotted in Fig. 4.5 as a function of the distance from the center of the cluster. The shaded region shows the value of our measurement that implies that $\frac{b_A}{b_G}(s)$=1 when taking into account the difference between

real and redshift space measurements discussed in the previous paragraph. The ratio is consistent with 1 on almost all scales. We cannot exclude, within the errors, much different values of the relative bias. Our data suggest an average relative bias consistent with unity but allow an upper limit of $\sim$6 (at 1σ) at separations $s < 50\ h_{70}^{-1}$ Mpc. For separations $s > 10\ h_{70}^{-1}$ Mpc no lower limits bigger than zero can be given. On larger scales the error increase thus it is difficult to draw any conclusion. At large separations the power-law shape of ξ_{CG} becomes uncertain, this makes a comparison of our data with those of Sánchez et al. (2005) less meaningful. We cannot exclude a significant antibiasing of AGNs when compared to galaxies, especially at low separations. Though the amplitude of the uncertainties of our data still allows a fluctuation in the relative biasing of more than a factor 2, we can conclude with a precision of 1σ that the probability for a galaxy to become an AGN is constant in the range of separations sampled in this work and that AGNs can be considered as good tracers of the dark matter distribution as are galaxies. New deep and wide field surveys such as XMM-COSMOS (Hasinger et al. (2007); Cappelluti et al. (2007)) and *Chandra*-COSMOS (P.I.: Martin Elvis) with their spectroscopically identified >2000 AGNs and $\sim$100 clusters will improve the precision of this work and will allow us to investigate the behaviour of AGNs also in the inner regions of the clusters. In this region the observations suggest that the AGNs source density is affected by other physical effects such as merging (Ruderman & Ebeling (2005), Branchesi et al. 2006) or ram pressure phenomena

4.7 Summary

We derived for the first time the soft X-ray spatial cross-correlation function between clusters and AGNs using the data of the *ROSAT* NEP survey. A strong clustering signal was detected on scale s <50 h_{70}^{-1} Mpc. The best power-law fit parameters are s_0=8.7$^{+1.2}_{-0.3}$ h_{70}^{-1} Mpc and γ=1.7$^{+0.2}_{-0.7}$. In this work we observed that the source density of AGNs is higher near clusters than in the field. This result confirms earlier findings of overdensities of AGNs around clusters reported by many authors and improves the evidence connecting the overdensities to the large scale structure of the Universe. We also derived the relative bias between AGNs and galaxies which is consistent with one on almost all scales investigated here. This result still allows, within the errors, a factor 2 fluctuation. New wide field surveys (such as XMM-COSMOS) per-

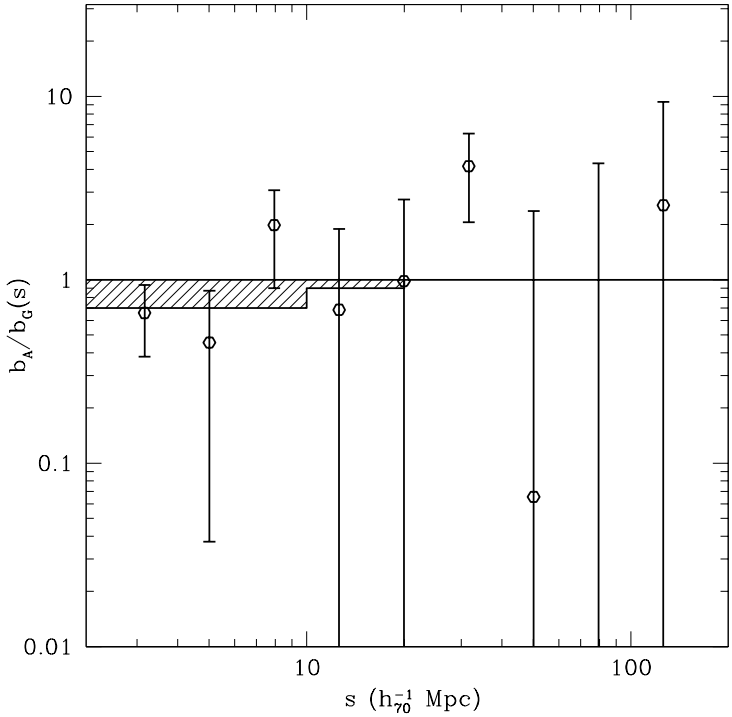

Figure 4.5: The ratio between the observed *ROSAT* NEP ξ_{CA}(s) and the best fit ξ_{CG}(r) obtained by Sánchez et al. (2005). Errors are quoted at the 1σ level. The shaded region shows the expected level of $\frac{b_A}{b_G}(s)$=1 if the cross-correlation functions were compared in the same space.

formed with the new generation X-ray telescopes will be useful to enlarge the statistics, to better understand the physics of AGNs in clusters and to extend the analysis to the inner regions of clusters.

Chapter 5

Real space cross-correlation function of AGN with Galaxy Clusters and Groups in the COSMOS field

5.1 Introduction

In the previous Chapter, by measuring ξ_{CA}(s), we detected for the first time a strong physical correlation between clusters and AGN in the NEP survey. Because of the low statistics and the large PSF of ROSAT, the significance of the result is about 3σ and we did not obtain informations on AGN within 2.5 Mpc h^{-1} from the center of the clusters.

The real space Cluster/Group-AGN cross-correlation function, ξ_{CA}(r), is defined so that the probability dP of finding an AGN and a cluster in the real space co-moving volume elements δV_1 and δV_2 is given by eq. 4.1. ξ_{CA}(r) is therefore equivalent to the radially averaged overdensity profile of AGN centered on a typical cluster or group. Its determination and the direct comparison with analytical models of mass clustering, can return interesting results on the relation between the AGN activity and the mass density.

The completion of the COSMOS field XMM-observations of AO4 and AO6 is a step forward for this kind of studies. The COSMOS field has been observed in 55 XMM-*Newton* pointings for a total exposure time of 1.5 Ms. The exposure map of the XMM-COSMOS survey is shown in Fig. 5.1.

The total covered area is $\sim$2.13 deg^2 with a median exposure of $\sim$40 ks.

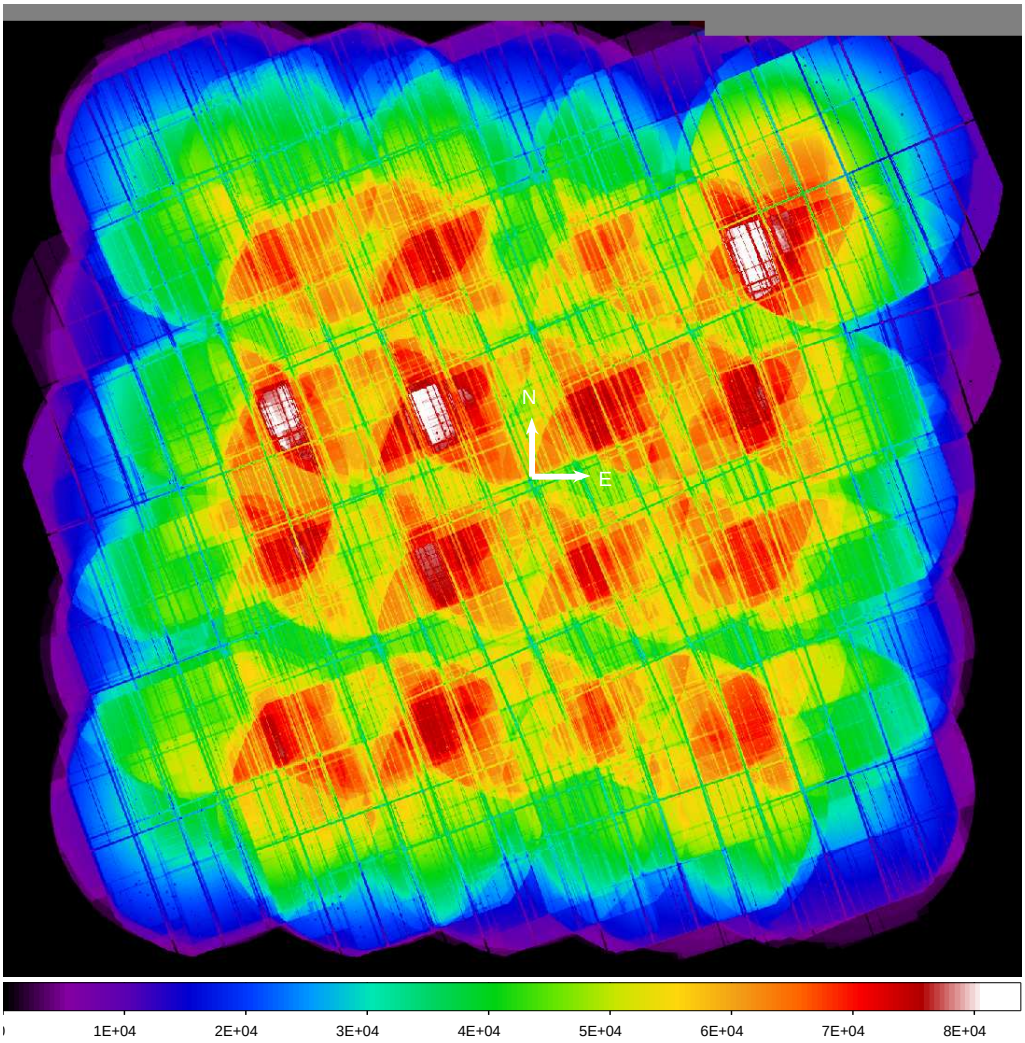

Figure 5.1: The exposure map of the 1.5 Ms XMM-COSMOS survey. The maximum exposure in the field is ∼83 ks.

Data reduction as well as source detection have been performed as described in Chapter 2. Sources with a detection likelihood "det_ml>10" have been included in the catalog in order to have a fraction of spurios detections $< 1\%$. A total of 1943 individual pointlike sources have been detected. In the single bands we detected 1672, 1128 and 254 sources in the 0.5–2 keV, 2–10 keV and 5–10 keV band, respectively. In the three bands our catalogs reach limiting fluxes of 6.2×10^{-16}, 3×10^{-15} and 5×10^{-15} erg cm^{-2} s^{-1}, respectively. Extended sources were searched with a wavelet source detection technique (for technical details see e.g. Finoguenov et al. 2007). A total of 140 extended sources have been detected down to a 0.5–2 keV limiting flux of 8×10^{-16} erg cm^{-2} s^{-1}. The variety of X-ray colors together with the emission from large

scale structures is shown with a false color image in Fig. 5.2. Optical identi-

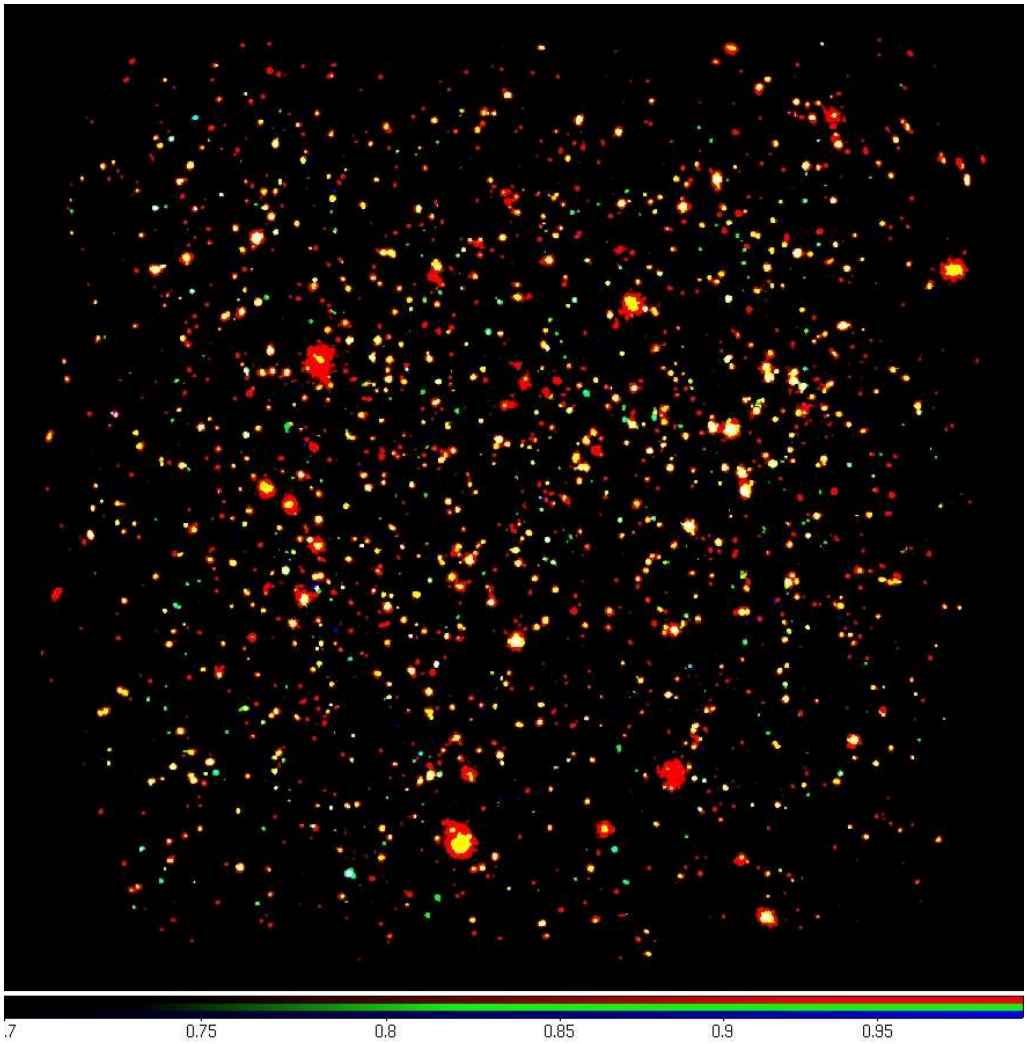

Figure 5.2: RGB X-ray image of the COSMOS field observed by XMM-*Newton*. The color are coded so that red, green and blue represent the 0.5-2 keV, 2-4.5 keV and 4.5-10 keV energy bands, respectively.

fications of pointlike source have been performed (Brusa et al. 2007,and additional unpublished identifications), by cross-correlating the XMM-COSMOS catalog with those from HST, Megacam, Subaru and Spitzer. About 84% of the XMM sources were found to have a unique optical/IR counterpart, $\sim$15% have 2 or more "bright" counterparts while, only 1% of the sources have no or more than 1 faint counterparts. Excluding the 110 with known SDSS and MMT spectroscopic informations, all the X-ray source have been followed up by the zCOSMOS project (Lilly et al. 2007) with VLT-VIMOS and by IMACS (most of the data are still not published, courtesy of S. Lilly, J. Trump). At the

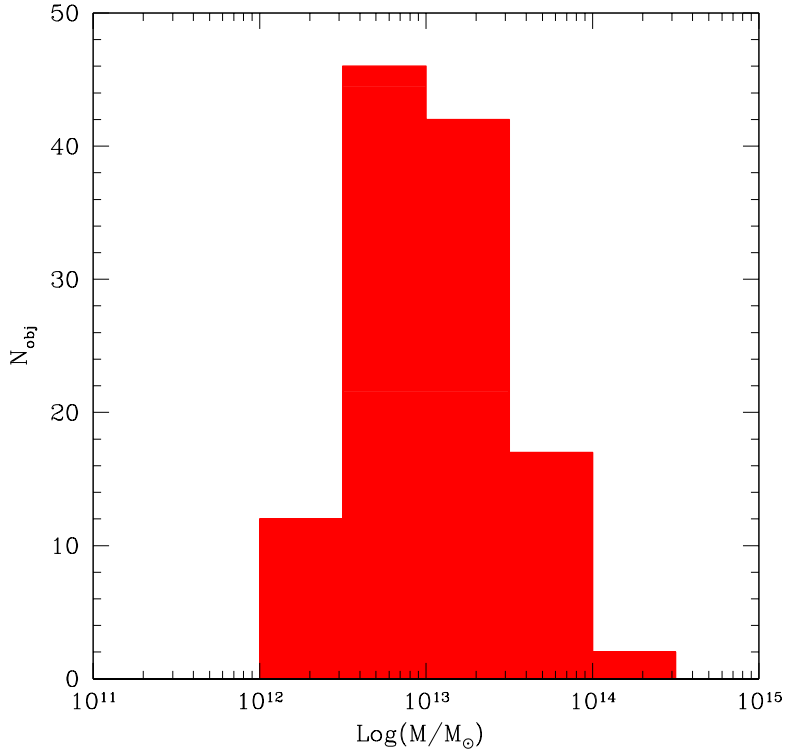

Figure 5.3: The mass distribution of the redshift confirmed clusters and groups in the XMM-COSMOS field.

moment of writing this thesis a total of about 600 X-ray sources have been spectroscopically identified as AGN and have a precise redshift information. 98 Galaxy clusters have been spectroscopically identified and all of them have a photometric redshift estimate. Through this chapter we will adopt a concordance, Lambda dominated, flat cosmological model with Ω_m=0.3, Ω_Λ=0.7 and H_0= 100h km/s/Mpc.

5.2 Sample Selection

In order to compute the cluster/groups AGN cross-correlation function we selected 98 clusters and groups in the XMM-COSMOS survey with spectroscopic redshift measurement. These objects have X-ray luminosities in the range 3×10^{41} erg/s $<L_{X_{0.5-2keV}} < 2 \times10^{44}$ erg/s.

An important information when studying cross-correlation function of clusters with any kind of objects is the total gravitational mass of the cluster. The observable parameters of clusters such as T, L_X and M are known to be linked

each other via well constrained scaling relations. In particular for the calculation of cluster masses Finoguenov et al. (2007) adopted:

$$M_{500} = 2.36 \times 10^{13} M_{\odot} \times T^{2.89} E(z)^{-1} \tag{5.1}$$

where M_{500} is the total gravitational mass (including dark matter), T is the temperature and E(z) is defined as:

$$E(z) = \sqrt{(\Omega_m (1+z)^3 + \Omega_\Lambda)} \tag{5.2}$$

The Mass distribution of the clusters selected in XMM-COSMOS with a spectroscopic redshift measurement is shown in Fig. 5.3. The median mass of the sources is $Log(M/M_{\odot}) \sim$13.

Active galactic nuclei have been selected with the same criterion as the clusters. Using the optical identification performed by Brusa et al. (2007), we selected sources with spectroscopic redshift information. The two samples contain therefore 98 clusters and groups and 567 AGN.

The redshift distributions of clusters and AGN are plotted in Fig. 5.4. The median redshift of the clusters is z=0.37 and z=0.95 for AGN. The resulting median redshift z_{eff}, contributing to the clustering[1] is z_{eff}=0.38.

5.3 Cross-correlation function

In order to derive ξ_{CA}, it is necessary to produce a random sample of AGN and clusters that reproduces, with good approximation, the selection effects of our survey. The redshift distributions of clusters and AGN have been smoothed with a Gaussian filter with σ_z=0.2. With such a choice, effects of large scale structures in the field have been smeared out. Indeed the width of the of the filter is larger than the known structures in the universe. Sources were then placed in redshift according to the smoothed redshift distribution.
Since the sensitivity of the survey varies across the field, selection effects introduce by sensitivity variations must be reproduced in the random samples, moreover the distribution of the spectroscopic mask on the field is difficult to reconstruct. Coordinates of the random particles were then randomly chosen among the known positions of the sources. In order to smear out the shot noise, random samples contain 50 times the number of sources of each catalog.

[1]This is defined as the median of the distribution $(\frac{dN_C}{dz})(\frac{dN_A}{dz})$

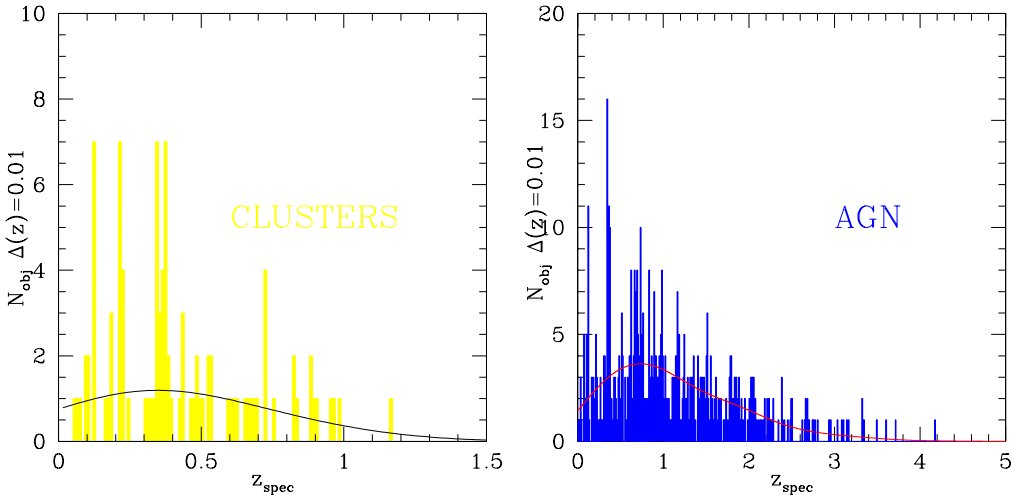

Figure 5.4: *Right panel*: In yellow, the redshift distribution of spectroscopically identified clusters and groups. The solid black line is the smoothed redshift distribution used for generating a random sample of clusters. *Left panel*: In blue, the redshift distribution of spectroscopically identified AGN. The solid black line represents the the smoothed redshift distribution used for generating a random sample of AGN.

If one does not randomize the x,y source positions in the random samples can, in principle, artificially dilute the clustering signal by smearing out the angular correlation. However, the majority of the AGN in XMM-COSMOS have a redshift larger than that of the bulk of the clusters. Since most these objects are not physically correlated with the lower-z clusters, the angular clustering is expected to be very weak. Therefore, the procedure described above is not expected to significantly bias the final result.

5.3.1 Redshift space cross-correlation function

The redshift space cross-correlation function has been determined using the Landy & Szalay (1993) estimator shown in eq. 4.2. The number of data-data, data-random and random-random source pairs were grouped into logarithmic bins of width $\Delta Log(s) = 1/6$ Mpc h^{-1}. Notice that in comparison with the ROSAT-NEP survey, the number of data-data pairs in the XMM-COSMOS is 5 times higher. Errors have been computed with a bootstrap resampling technique in the same way as in Chapter 4.

ξ_{CA}(s) is plotted in Fig. 5.5. A positive signal has been significantly de-

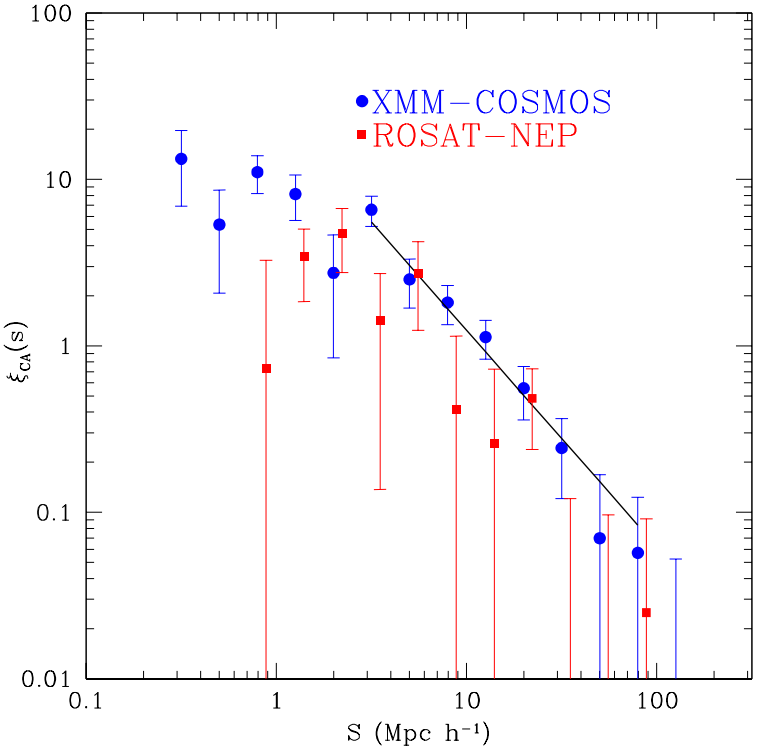

Figure 5.5: *Blue circles*: ξ_{CA}(s) in the XMM-COSMOS surveys compared with the results in the ROSAT-NEP survey (*Red squares*)

tected on scales s<70 Mpc h^{-1}. On scales smaller than 3 Mpc is evident a flattening of the signal. At small separations the determination of distances between galaxies using redshift is biased by peculiar velocities. This effect, called "finger of god effect", is known to reduce the clustering signal at small separations since the estimates of distance is diluted in the *z* direction (see e.g. Hawkins et al. 2003; Kaiser 1987). Moreover, on this scale, the X-ray emission of AGN is embedded in the diffuse emission of the galaxy cluster with the effect of reducing the sensitivity for detecting AGN. This effect is difficult to be included in the random samples with the spectroscopic completeness level of this work. Considering the mass distribution of the clusters in our sample, the measurable cluster emission extends to scales of the order of 0.5-1 Mpc h^{-1}.

ξ_{CA}(s) has been fitted with a canonical power-law in the form of eq. 4.4 using a χ^2 minimization technique. The fit has been performed in the separation range 2-60 Mpc h^{-1} in order to avoid strong redshift distortion effects. As a result we obtained s_0=11.8±2.1 Mpc h^{-1} and γ=1.38±0.15.

In Fig. 5.5 ξ_{CA}(s) is compared with the results obtained in the ROSAT-NEP survey. Although the NEP measurement is more noisy than that in the COSMOS field, one can notice than on average ξ_{CA}(s) has an higher normalization in XMM-COSMOS than in the ROSAT-NEP. In order to have a first estimate of the redshift distortion effects, we repeated the fit in the separation range 0.5-60 Mpc h^{-1}. On these scales we obtain s_0=9.9$\pm$1.9 Mpc h^{-1} and γ=0.85$\pm$0.10. The net effect is therefore a flattening of the correlation, due to the low-separation bins, and a drop of the correlation length.

5.3.2 Real space cross-correlation function

To determine the real space cluster-AGN cross-correlation function $\xi_{CA}(r)$, we follow the standard practice of computing the cross-correlation function in a two dimensional grid of pair separations parallel and perpendicular to the line of sight.

We define r_p (perpendicular) as the projected space separation between object at redshift z and another separated in the sky with an angle θ, so that r_p=yθ where y is the co-moving distance of the object and θ the angular separation of the objects. If s is the co-moving redshift space separation of the two objects the we define the line of sight separation (parallel)

$$\pi = \sqrt{s^2 - r_p^2}. \tag{5.3}$$

The logarithmic bins have a width of $\Delta Log(r_p,\pi)$=0.2 Mpc h^{-1}. As above the the two-dimensional cross-correlation function $\xi_{CA}(r_p,\pi)$ is estimated using the Landy & Szalay (1993) estimator of eq. 4.2. Eventual redshift errors and peculiar velocities affect the separation along the π-direction but not along the r_p direction. We then project $\xi_{CA}(r_p,\pi)$ along the π direction to obtain the projected cross-correlation function $w_p(r_p)$:

$$w_p(r_p) = 2\int_0^\infty d\pi \xi_{CA}(r_p,\pi). \tag{5.4}$$

In practice we integrate up to some cutoff value π=50 Mpc h^{-1} in order to not add noise to the integral. We checked that the result is stable by choosing a cutoff between 40 Mpc h^{-1} and 70 Mpc h^{-1}. The projected cross-correlation function is related to the real-space cross-correlation $\xi_{CA}(r)$ function via:

$$w_p(r_p) = 2\int_{r_p}^\infty \frac{r\xi_{CA}(r)}{(r^2 - r_p^2)^{1/2}} dr. \tag{5.5}$$

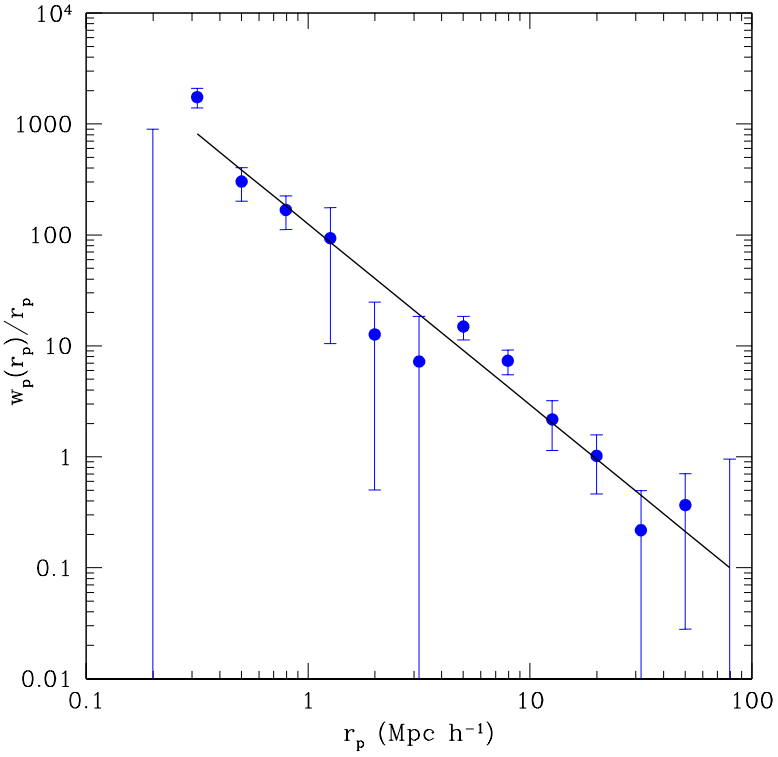

Figure 5.6: The projected cross-correlation function $w_p(r_p)/r_p$ (*Blue dots*) and the best fit power-law (*Black solid line*).

If $\xi_{CA}(r)$ follows the canonical power-law $\xi_{CA}(r)=(r/r_0)^{-\gamma}$, then

$$\frac{w_p(r_p)}{r_p} = \frac{\Gamma(1/2)\Gamma[(\gamma-1)/2]}{\Gamma(\gamma/2)} \left(\frac{r_0}{r_p}\right)^{\gamma}. \tag{5.6}$$

$w_p(r_p)$ has been fitted making use of eq. 5.6. The fit was performed in the range 0.3 Mpc $h^{-1} < r_p <$35 Mpc h^{-1} where the function shows an amplitude significantly higher than 0. As a result we obtained r_0=7.75$\pm$1.17 Mpc h^{-1} and γ=1.64$\pm$0.10. The projected cross-correlation function $w_p(r_p)/r_p$ and its best fit power-law are plotted in Fig. 5.6. As for $\xi_{CA}(s)$ errors have been computed with a bootstrap resampling technique.

In order to check whether the cluster-AGN cross-correlation function depends on the AGN luminosity, the sample of AGN has been then divided into two subsamples, one including all the sources with $L_{x_{0.5-2keV}}$ <43 erg/s h^{-2} and the other with $L_{x_{0.5-2keV}}$ >43 erg/s h^{-2} (hereinafter LL anf HL sample). The LL sample containts 212 sources while, the HL sample contains 355 sources. We then recomputed $w_p(r_p)$, for both the data sets, by making use of random

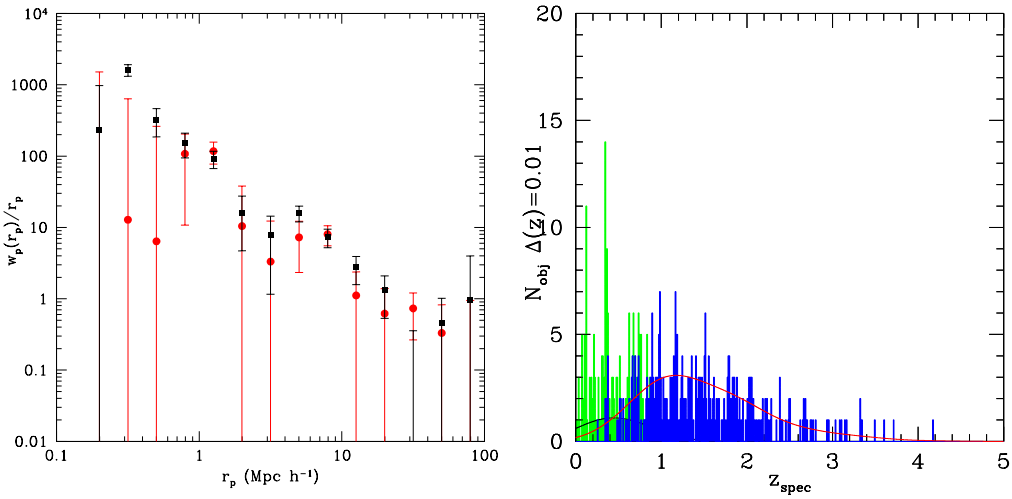

Figure 5.7: *Left Panel*:The projected cross-correlation function $w_p(r_p)/r_p$ for the LL sample (*Black squares*) and for the HL sample (*Red circles*). *Right Panel*: The redshift distribution of the LL sample (*Green*) and of the HL sample (*Blue*) and their smoothed distribution from which random samples were drawn (*Black and Red Solid line, respectively*).

samples generated according to redshift distributions of sources in the two groups. The two projected cross-correlation function are plotted in Fig. 5.7

Though the HL sample contains more sources than the LL sample, the cross-correlation function cluster-HL AGN is more noisy than that of the LL sample. This can be explained with the fact that most of the HL AGN have redshifts greater than z=1 while, the all the LL AGN have z<1. As one can notice in Fig. 5.4, only one cluster has redshift z> 1, therefore the actual number of pairs playing a role in the clustering, is lower in the HL than in the LL sample. We fitted $w_p(r_p)$ obtained with both the samples by using eq. 5.6 as model in the projected separation range 0.3 Mpc $h^{-1} < r_p <$35 Mpc h^{-1}. In the LL sample we obtained r_0=7.27$\pm$0.70 Mpc h^{-1} and γ=1.67$\pm$0.04 while, in the HL sample, we obtained r_0=8.34$\pm$3.24 Mpc h^{-1} and γ=1.43$\pm$0.23. Though the lower statistics in measuring $w_p(r_p)$ in the HL sample, the results do not show any significant difference. The sample was also divided into two subsamples according to the redshift of the AGN: one including sources with z<0.4 and the other with sources with z>0.4. This comparison did not return any reliable result since below z=0.4 there are 3 strong spikes in the redshift distribution which artificially increases the signal at low z. Thus, despite the good statistics in both the samples, the production of the smoothed redshift distribution

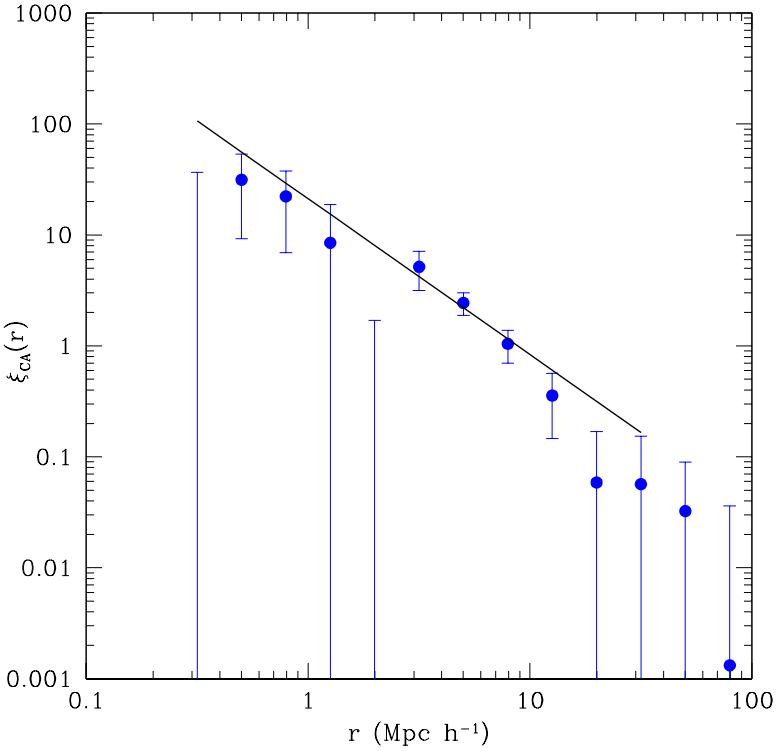

Figure 5.8: The real space cluster-AGN cross-correlation function $\xi_{CA}(r)$ in the COSMOS-field (*Blue circles*), the best fit power is plotted as a *black dashed line*.

on such a small redshift interval is unreliable because of the presence of the spikes.

It is possible to directly estimate $\xi_{CA}(r)$ by inverting $w_p(r_p)$ without the assumption that it behaves like a power-law. By inverting eq. 5.5 we get

$$\xi_{CA}(r) = -\frac{1}{\pi}\int_r^\infty \frac{dw_p(r_p)/dr_p}{\sqrt{(r_p^2 - r^2)}} dr_p. \tag{5.7}$$

The solution of this integral can be obtained assuming a step function for $w_p(r_p)$=w_i in bins centered in r_{p_i} and interpolating between values. Thus the solution is

$$\xi_{CA}(r_{p_i}) = \frac{1}{\pi}\sum_{j>i} \frac{w_{j+1} - w_j}{r_{p_{j+1}} - r_{p_j}} \log\left(\frac{r_{p_{j+1}} + \sqrt{r_{p_{j+1}}^2 - r_{p_i}^2}}{r_{p_j} + \sqrt{r_{p_j}^2 - r_{p_i}^2}}\right) \tag{5.8}$$

for r=r_{p_i}.

The real space cluster-AGN cross-correlation function $\xi_{CA}(r)$ in the COSMOS-field is presented in Fig. 5.8. The clustering signal is significantly greater

than zero on scales smaller than 20 Mpc h^{-1}. At a separation of $\sim$1.5 Mpc h^{-1} the signal shows a dip. We consider this feature, which is also detectable in the angular auto correlation function of AGN and in the real space autocorrelation function of AGN (Gilli et al. 2008, private communication), as an artefact of the XMM-COSMOS survey. As for the projected cross-correlation function $\xi_{CA}(r)$ has been fitted assuming a powerlaw of the form $\xi_{CA}(r)=(r/r_0)^{-\gamma}$. In the separation range 0.3 Mpc $h^{-1} < r_p <$35 Mpc h^{-1}, the best fit parameters are r_0=7.11$\pm$0.85 Mpc h^{-1} and γ=1.48$\pm$0.09. It is worth to note that the results obtained with two different technique give similar and consistent results.

5.4 Biasing of ξ_{CA}

As explained in the previous chapter, the linear theory of biasing predicts that

$$\xi_{CA}(r) = b_C b_A \xi_{\rho\rho}(r), \tag{5.9}$$

where b_C, b_A and $\xi_{\rho\rho}$ are the bias factor of AGN, the bias factor of clusters and the autocorrelation function of matter, respectively. For practical reasons, we will refer to the product $b_C b_A$ as b_{CA}. In the context of the linear biasing theory b_{CA} is scale independent. Starting from eq. 4.1 and using the formalism of Peebles (1980) we can relate ξ_{CA} with the differential radial number density of AGN around clusters with

$$n(r) =< n_A > [1+\xi_{CA}(r)], \tag{5.10}$$

that combined with eq. 5.9 yields

$$n(r) =< n_A > [1+b_{CA}\xi_{\rho\rho}(r)]. \tag{5.11}$$

If the number density of AGN at certain distance r from a cluster simply depends on the underlying mass overdensity, then b_{CA} is constant at all scales. On the other hand, if the efficiency of producing AGN is a function of the distance from the cluster and as a consequence of the mass overdensity, the b_{CA} should show a scale dependence and eq. 5.9 must be rewritten as

$$\xi_{CA}(r) = b_{CA}(r)\xi_{\rho\rho}(r). \tag{5.12}$$

Sánchez et al. (2005) showed that the analogous of b_{CA}, b_{CG}, in the cross-correlation clusters-galaxies shows a decline at small separations. This decline has been detected around clusters of different mass, including clusters

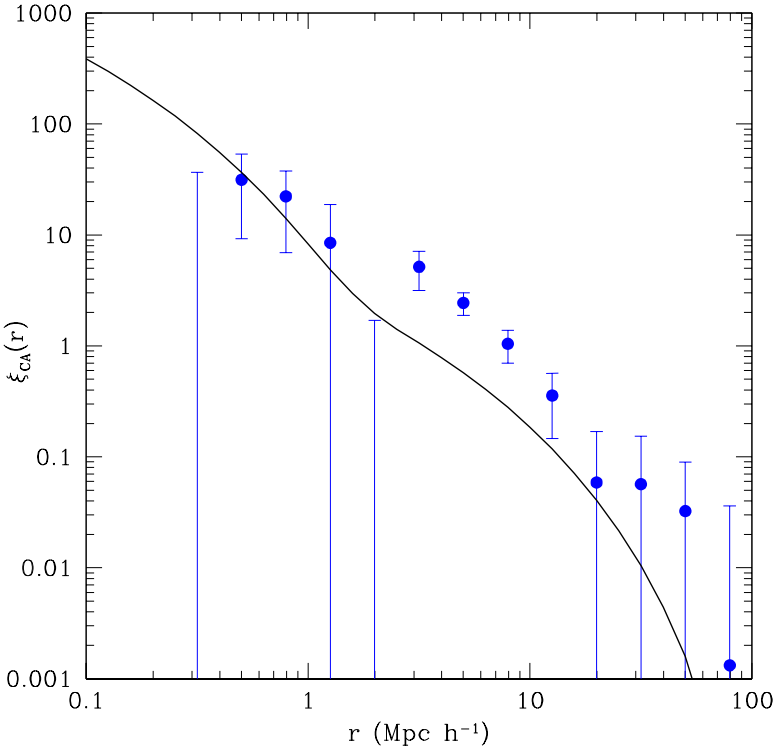

Figure 5.9: The real space cluster-AGN cross-correlation function $\xi_{CA}(r)$ in the COSMOS-field (*Blue circles*), compared to the dark matter halo auto-correlation function of Hamana et al. (2002) (*Black solid line*).

of mass similar to those used in this work. This decline has been interpreted as a loss of efficiency for galaxy formation in high density environment.

In order to test if also b_{CA} is scale dependent in cluster/groups with mass of the order of 10^{13}M/M$_{\odot}$, we compared the autocorrelation of matter $\xi_{\rho\rho}(r)$ with $\xi_{CA}(r)$. Making use of the Fortan library available on line[2] , we computed the halo model prediction for the dark matter two-point correlation function according to Hamana, Yoshida & Suto (2002). The autocorrelation function of matter has been computed assuming a flat, Λ dominated universe, H_0=70 km/s/Mpc, σ_8=0.75 and a median effective redshift $< z_{eff} >$=0.38.

In Fig. 5.9 is plotted $\xi_{CA}(r)$ together with $\xi_{\rho\rho}(r)$. As one can notice $\xi_{CA}(r)$ shows a systematically higher amplitude. The ratio of the two curves is basically the bias factor b_{CA}. In order to estimate its value, in the linear theory framework, we adopted the method described in Chapter 3. If $\xi_{CA}(r)$ is a power-law in the form of $\xi_{CA}(r)=(r/r_0)^{-\gamma}$, then the amplitude of density fluctu-

[2]http://th.nao.ac.jp/ hamanatk/OPENPRO/index.html

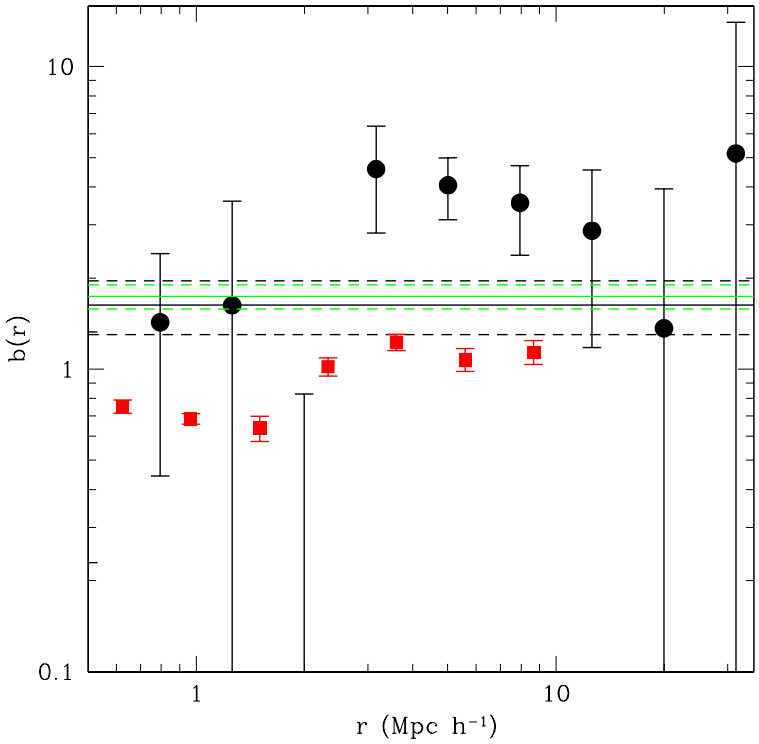

Figure 5.10: The bias factor $b_{CA}(r)$ (*Black Dots*) compared with $b_{CG}(r)$ (*Red Circles*) rescaled to the median redshift of our sample. The *green solid line* is the best fit of b_{CA} obtained using $\sigma_{8,CA}(z)$ estimated with the best fit power-law, the 1σ confidence interval of the fit is plotted with a *green dashed line*. The *black solid line* is the best fit of $b_{CA}(r)$ obtained with the ratio $\xi_{CA}(r)/\xi_{\rho\rho}(r)$, the 1σ confidence interval of the fit is plotted with a *black dashed line*.

ation in a sphere of 8 Mpc h^{-1} $\sigma_{8,CA}(z)$ is given by eq. 3.14. If$\sigma_{8,CA}$ is known the bias factor is given by

$$b_{CA}(z) = \frac{\sigma_{8,CA}(z)}{\sigma_8 D(z)}. \tag{5.13}$$

Where σ_8=0.75 at z=0 and D(z) is the linear growth factor. Using the best fit values obtained by fitting ξ_{CA} in the real space, we obtain $\sigma_{8,CA}$=1.11$\pm$0.10 and consequently $b_{CA}(z=0.38)$=1.74$\pm$0.16. By knowing the autocorrelation of matter it is possible to derive $b_{CA}(r)$ and compare it with the value of $b_{CG}(r)$ obtained by Sánchez et al. (2005). $b_{CA}(r)$ is then given by the ratio $\xi_{CA}(r)/\xi_{\rho\rho}(r)$. $b_{CA}(r)$ and $b_{CG}(r)$ are plotted in Fig. 5.10.

In order to test the robustness of our measurement of $b_{CA}(z=0.38)$ we performed a χ^2 fit of the points using a constant as model. In the separation range

0.5 Mpc h^{-1} -30 Mpc h^{-1} we obtained as a best fit $b_{CA}(r)$=1.63$\pm$0.33. This value is in excellent agreement with that estimated above using $\sigma_{8,CA}(z)$ estimated from the best fit power-law. However Fig. 5.10 shows at a 2σ level a tendency of $b_{CA}(r)$ to decrease, by about a factor 3, at small separations. Though this effect could be introduced below $\sim$1-1.5 Mpc h^{-1} by the fact that the detection of X-ray emitting AGN is less likely inside the extended clusters emission, a similar behaviour is observed with galaxy around clusters. Allthough we can not confirm such an effect we can consider this decline of $b_{CA}(r)$ as an interesting hint. In order to compare the behaviour of the bias factor of AGN with that of galaxies near clusters, we plotted $b_{CG}(r)$ obtained by Sánchez et al. (2005) by cross-correlating 2dF groups and APM galaxies. These groups have masses comparable to those of our sample. Since the median redshift of the sample of Sánchez et al. (2005) is $< z_{eff} > \sim$0.15 we rescale their measurement by multiplying their estimate of $b_{CG}(r)$ by the ratio of the growth factor of their and our survey (i.e. $\sigma_8 D(0.15)/\sigma_8 D(0.38)$). In the separation range 0.5-10 Mpc h^{-1} the best fit to their values are $b_{CG}(r, z = 0.38)$=0.78$\pm$0.01. $b_{CG}(r)$ shows a decrease of a factor 2 below 3Mpc h^{-1} . The possible decline indicated for AGNs is consistent with that observed with galaxies, however can not yet be determined significantly. This would be possible, for example, with the final photo-z catalog of AGN in COSMOS for which, the decline of sensitivity on AGN in the inner part of clusters could be reproduced and corrected.

From studies of the autocorrelation function, on large scales, galaxies are known to be unbiased tracers of large scale structure, therefore both theory and measurement of their autocorrelation function (i.e. determined independently from the environment) constrain their bias factor to 1 with a very small uncertainty. Since $b_{CG} = b_C b_G$ and as stated above $b_{CA} = b_C b_A$, by assuming b_g=1 and the ratio of the bias factor between XMM-COSMOS cluster to 2dF clusters close to 1, the ratio b_{CA}/b_{CG} should give an estimate of $b_A(z)$. Making use of the fits to $b_{CG}(r)$ derived from the Sánchez et al. (2005) work and that to $b_{CA}(r)$, we obtain $b_A(z = 0.38)$=2.08$\pm$0.42.

5.5 Density profile of AGN

The cross-correlation function is related to the differential density profile via eq. 5.10.

Ruderman & Ebeling (2005) estimated the space density profile of AGN around a sample of 56 clusters in the Chandra archive. They coverted the angular

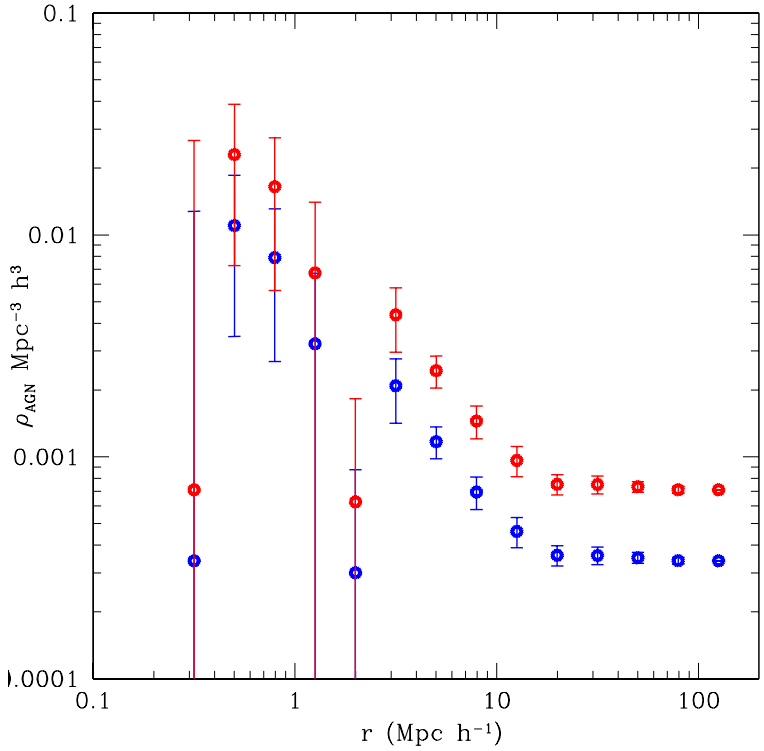

Figure 5.11: The AGN density profile around a XMM-COSMOS cluster/group. The *Blue dots* represents the direct estimate at z=0.38, while, *Red dots* represent the expected density profile at z=1.4 under the assumption of no evolution of the clustering length.

AGN profile into a 3D profile and measured significant overdensities within 4 Mpc from the cluster centres. Eastman et al. (2007) measured the average number density of AGN in galaxy cluster as a function of the cluster redshift. They detect an increase of the AGN number density by factor 20 in the redshift range $0.2< z <0.6$.
In order to compare our results with estimates performed by Ruderman & Ebeling (2005); Eastman et al. (2007) we plot the AGN density profile around a typical cluster of our sample (i.e, with mass of the order of $10^{13} M_{\odot}$). As a normalization of the profile we use the source density of AGN at z=0.38 derived from the luminosity function of Hasinger et al. (2005) for AGN with $L_{x_{0.5-2}} > 10^{41.5}$ erg/s. The density profile is plotted in Fig. 5.11. Since it was not possible to test for a possible evolution of the cross-correlation function, by assuming that the co-moving clustering length r_0 was constant with redshift, we plot the expected density profile of AGN around a cluster at z=1.4. In this redshift

interval we measure an increase of the AGN activity of about of factor 2. This is much less than what observed by Eastman et al. (2007) who detected an increase of the AGN density in clusters of about a factor 20. Such a result requires, however, further measurements since an eventual validation could have an important impact on the study of the AGN-environment relation, as well as the measurement of the X-ray observables of galaxy clusters.

5.6 Discussion

We presented the first determination of the cluster/group-AGN real space X-ray cross-correlation function in the 1.5 Msec XMM-COSMOS survey. This result confirms, and improves the precision, of previous findings in ROSAT/NEP survey. The redshift space cross-correlation function in the XMM-COSMOS shows a 2.5σ higher amplitude than that measured in the ROSAT-NEP survey. This difference, which in part can be attributed to the poor statistics in the NEP survey, can also arise in the nature of the two sample. NEP clusters are intrinsically more massive, more luminous and they are at lower redshift. At the same time, AGN in the NEP survey are, on average, 2 order of magnitude brighter than those in the COSMOS survey. In this Chapter we measured a real space cross-correlation length of $\sim$7.7 Mpc h^{-1} . The correlation length does not show a significant luminosity dependence though, given the small number of HL AGN in the redshift interval of the clusters, it cannot be ruled out. The best fit parameter allowed us to estimate a bias factor b_{CA} $\sim$1.74. By dividing $\xi_{CA}(r)$ by the autocorrelation function of matter we detect, at 2σ, an indication of a factor$\sim$3 decrease of $b_{CA}(r)$ at small separations. This result, potentially intriguing, requires a better estimate of the selection function of AGN inside the cluster diffuse emission. A decrease $b_{CA}(r)$, if real, is analogous but with a bigger amplitude to the decrease of $b_{CG}(r)$. This decrease has been explained with a lower efficiency of galaxy formation in the inner part of clusters. On the other hand, a smaller $b_{CA}(r)$, is not completely unexpected. An important ingredient for AGN activity is the presence of a cold gas reservoir in galaxies to fuel the accretion. When falling in the galaxy clusters potential, a galaxy loses cold gas via ram pressure stripping, inhibiting AGN activity and star formation. This effect could give a reduction of the number density of AGN about at the virial radius of the cluster, similar to what is indicated here. The combination of this effect, with a low probability of producing a galaxy in clusters, should give a significant reduction

of $b_{CA}(r)$. The calculation, of $\xi_{CA}(r)$, with the full sample of X-ray sources of XMM-COSMOS, with full photo-z determination, will provide an increase of the significance of a factor $\sim$3, allowing us to detected the possile scale dependence of $b_{CA}(r)$ around clusters.

We estimated, in a way independent from that in Chap. 3, $b_A(z=0.38) \sim 2.1$. This value, according to the prediction of Sheth et al. (2001), suggests that AGN preferentially reside in dark matter halos with mass of the order of $10^{13} M/M_{\odot}$. The results obtained in this work show that X-ray selected AGN are hosted in more massive DM halos than what is estimated from luminous optically selected QSO (i.e. $10^{12} M/M_{\odot}$). This could have interesting implication for the study of the evolution of AGN. In fact luminous QSO are known to peak in space density around z=2-3 and to reside in relatively low mass halos: one can therefore speculate that they were triggered by low mass DM merging when the universe was smaller and with a higher probability of merging. On the other hand medium-low luminosity AGN, like the X-ray selected in XMM-COSMOS, peak at z$\sim$0.5-1, and are hosted in higher density environments. In this case one can speculate that low-luminosity AGN are triggered by mergers of DM halos in groups. Since these structure form later than the hosts of optical QSO, COSMOS-like AGN peak later in the history of the Universe since they are waiting for the formation of massive structures.

To conclude we measured and predicted the number density of AGN in clusters according to the observed evolution of AGN. We predict, between redshift 0.5 and 1.4, an increase of the source density in clusters of about a factor 2. This increase is much lower than what claimed by Eastman et al. (2007) (i.e. factor 10 lower), but in good agreement with the evolution of the source surface density of AGN around and in galaxy clusters observed by Cappelluti et al. (2005); Branchesi et al. (2006).

Together with a complete photo-z catalog, the recently completed 1.8 Msec *Chandra* observations of COSMOS will increase the statistics on the number of AGN in clusters. Future surveys like the eROSITA all-sky and deep survey, will provide an unprecedented number of clusters and AGN. This should allow to obtain importan new informations on the physics of AGN in clusters, and will allow to improve cosmological parameters making use of the galaxy clusters mass function. It is therefore important to well constrain the impact of AGN to the total cluster emission, especially at high-z, to correctly estimate the X-ray parameters of galaxy clusters.

Summary

In this thesis I investigated the relation between active galactic nuclei (AGN) and environment. The work has been performed making use of wide-field, moderately deep X-ray surveys most importantly the XMM-COSMOS survey data.

An elaborated technique for the data analysis and the source detection was developed. This technique was also successfully exported and adapted to *Chandra* observations.

It has been demonstrated that the XMM-COSMOS survey is weakly affected by the so called cosmic variance, and derived source number counts relations which will serve as a future reference. These relations are in excellent agreement with the predictions of XRB population synthesis models.

The angular auto-correlation function of AGN was derived in the XMM-COSMOS survey. This relation has been deprojected to the 3D correlation function, yielding, at $z \sim 1$ a correlation length of $\sim$9 Mpc h^{-1} . We derived the linear biasing factor of AGN at the same redshift. This value is consistent with the picture that X-ray selected AGN reside in DM halos of order $10^{13-14} M/M_{\odot}$.

Using the 99.5% spectroscopic complete data of the ROSAT-NEP survey, the redshift space soft-X-ray cross-correlation function was derived. We measured, for the first time, a significant clustering of AGN in cluster environments.

This result has been extended in the 1.5 Msec XMM-COSMOS survey. In this case the cross-correlation function was derived in real space. The correlation length we estimated around 7.7 Mpc h^{-1} . This correlation length does

not show any luminosity dependence, though the statistics does not allow to rule out this possibility.

We also have a marginal (2σ) indication that the probability of activating an AGN drops in the inner 3 Mpc h^{-1} of the cluster in a similar way to what observed for galaxies. The bias factor of AGN b_A has been estimated in the real space also at z=0.38. $b_A(z = 0.38)$ is $\sim$2, which is consistent with the finding that AGN preferentially reside in DM halos with mass of $\sim 10^{13} M/M_{\odot}$.

Once compared with optical studies of high luminosity QSO, this work points out different clustering properties between high-z luminous-QSO and lower-z, lower luminosity X-ray selected AGN. One interpretation of this observing evidence is that high-z luminous-QSO are triggered by low mass (i.e. $10^{12} M/M_{\odot}$) mergings. On the other hand lower luminosity X-ray selected AGN are triggered by merging in dark matter halos with masses of the order of galaxy groups. This picture is in good agreement with the observed evolution of AGN. In fact low-luminosity AGN are rare in the early universe and the reason of that could be that massive structures have still to form. Therefore, this work gives additional evidences that the evolution of AGN is correlated with that of large scale structure in our Universe.

To conclude, it was not possible to confirm recent claims of an evolution of AGN in clusters of more than a factor 20 between z=0.2 and z=0.6. This work has however potential implication for future X-ray survey missions like eROSITA.

Acknowledgements

The fact that I am writing this chapter means that my PhD thesis is completed and that it is time to say some thank you. More important is that this chapter means that I finished to go to school right after 22 years.

The first thank you goes to my family that encouraged me in all my adventures and that helped me for many other things.

The other important person that I met during my PhD period is my girlfriend Elena who sustained me all over these years. Thank you so much, but words in this case can not express all my gratefulness.

A big thank you to my supervisors Günther Hasinger and Hans Böhringer. They both encouraged and believed to me from the first time we met. I will be indebted with them forever for their patient supervision, their teachings and their friendship. I hope to continue this friendly and stimulating scientific and personal collaboration. Thank you very much.

This work would not be possible without the help of Marcella Brusa whit whom I positively collaborated in these years.

All the members of the XMM-COSMOS group at MPE and all over the world gave substantial contribution to my scientific skills and especially to this work.

A special thanks goes to Peter Schuecker that was an excellent and patient teacher. Thank you also to all the scientists of the MPE who were always available for illuminating discussions.

I would like to thank all the friends that I met during these years at the IMPRS. In particular Filiberto, Elisabetta, Marco, Paola, Rene, Italo, Eugenio and Mattia and all the other students. Thanks a lot to Michael Bauer and Hermann Brunner for helping me in translating the summary into German.

My friend Andrea Antonelli, his girlfriend Silvia and their wolf Ela who came to meet me and with whom I try to sail some regatta with also some discrete results. I hope to sail more in the next years.

And thanks also to whom I forgot to mention.

Thank you very much.

Garching 26/08/2007

Bibliography

Akylas, A., Georgantopoulos, I., & Plionis, M. 2000, MNRAS, 318, 1036

Baldi, A., Molendi, S., Comastri, A., Fiore, F., Matt, G., & Vignali, C. 2002, ApJ , 564, 190

Barger, A. J., Cowie, L. L., Mushotzky, R. F., Yang, Y., Wang, W.-H., Steffen, A. T., & Capak, P. 2005, AJ , 129, 578

Basilakos, S., Georgakakis, A., Plionis, M., & Georgantopoulos, I. 2004, ApJL , 607, L79

Basilakos, S., Plionis, M., Georgakakis, A., & Georgantopoulos, I. 2005,MNRAS, 356, 183

Blake, C., Pope, A., Scott, D., & Mobasher, B. 2006, MNRAS, 368, 732

Bauer, F. E., Alexander, D. M., Brandt, W. N., Schneider, D. P., Treister, E., Hornschemeier, A. E., & Garmire, G. P. 2004, AJ , 128, 2048

Böhringer, H., et al. 2002, ApJ , 566, 93

Branchesi, M., Gioia, I. M., Fanti, C., Fanti, R., & Perley, R. 2006, A&A , 446, 97

Branchesi, M., Gioia, I. M., Fanti, C., Fanti, R., Cappelluti, N., 2007, Accepted for publication on A&A

Brandt, W. N., & Hasinger, G. 2005, ARA%A , 43, 827

Brunner, A. et al. 2006, A&A , submitted

Brusa, M. et al. 2007, ApJS , in press

Budavári, T., et al. 2003, ApJ , 595, 59

Butcher, H., & Oemler, A., Jr. 1984, ApJ , 285, 426

Cappelluti, N., Cappi, M., Dadina, M., Malaguti, G., Branchesi, M., D'Elia, V., & Palumbo, G. G. C. 2005, A&A , 430, 39

Cappelluti, N., Boehringer, H., Schuecker, P., Pierpaoli, E., Mullis, C. R., Gioia, I. M., & Henry, J. P. 2006, accepted by A&A , ArXiv Astrophysics e-prints, arXiv:astro-ph/0611553

Cappelluti, N., et al. (2007) ApJS , in press

Cappi, M., et al. 2001, ApJ , 548, 624

Carrera, F. J., Barcons, X., Fabian, A. C., Hasinger, G., Mason, K. O., McMahon, R. G., Mittaz, J. P. D., & Page, M. J. 1998, MNRAS, 299, 229

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

Cavaliere, A., & Fusco-Femiano, R. 1976, A&A , 49, 137

Comastri, A., Setti, G., Zamorani, G., & Hasinger, G. 1995, A&A , 296, 1

Cowie, L. L., Barger, A. J., & Kneib, J.-P. 2002, AJ , 123, 2197

Croft, R. A. C., Dalton, G. B., Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1997, MNRAS, 291, 305

Croft, R. A. C., Dalton, G. B., & Efstathiou, G. 1999, MNRAS, 305, 547

Croom, S. M., et al. 2005, MNRAS, 356, 415

D'Elia, V., Fiore, F., Elvis, M., Cappi, M., Mathur, S., Mazzotta, P., Falco, E., & Cocchia, F. 2004, A&A , 422, 11

Dickey, J. M., & Lockman, F. J. 1990, ARA%A , 28, 215

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

Della Ceca, R., et al. 2004, A&A , 428, 383

De Luca, A., & Molendi, S. 2004, A&A , 419, 837

Dressler, A., Thompson, I. B., & Shectman, S. A. 1985, ApJ , 288, 481

Eddington, A. S., Sir 1940, MNRAS, 100, 354

Eastman, J., Martini, P., Sivakoff, G., Kelson, D. D., Mulchaey, J. S., & Tran, K.-V. 2007, ApJL , 664, L9

Evrard, A. E. 1991, MNRAS, 248, 8P

Finoguenov, A. et al. 2007, ApJS , this volume

Fiore, F., et al. 2001, MNRAS, 327, 771

Fiore, F., et al. 2003, A&A , 409, 79

Georgantopoulos, I., Stewart, G. C., Shanks, T., Griffiths, R. E., & Boyle, B. J. 1993, MNRAS, 262, 619

Giacconi, R., et al. 1979, ApJL , 234, L1

Giacconi, R., & Zamorani, G. 1987, ApJ , 313, 20

Giacconi, R., et al. 2001, ApJ , 551, 624

Giavalisco, M., Steidel, C. C., Adelberger, K. L., Dickinson, M. E., Pettini, M., & Kellogg, M. 1998, ApJ , 503, 543

Gilli, R., Salvati, M., & Hasinger, G. 2001, A&A , 366, 407

Gilli, R., et al. 2003, ApJ , 592, 721

Gilli, R., et al. 2005, A&A , 430, 811

Gilli, R., Comastri, A. & Hasinger, G. 2006, A&A, in press, astro-ph/0610939

Gioia, I. M., Henry, J. P., Mullis, C. R., Böhringer, H., Briel, U. G., Voges, W., & Huchra, J. P. 2003, ApJS , 149, 29

Giommi, P., Perri, M., & Fiore, F. 2000, A&A , 362, 799

Grazian, A., Negrello, M., Moscardini, L., Cristiani, S., Haehnelt, M. G., Matarrese, S., Omizzolo, A., & Vanzella, E. 2004, AJ , 127, 592

Green, P. J., et al. 2004, ApJS , 150, 43

Hamana, T., Yoshida, N., & Suto, Y. 2002, ApJ , 568, 455

Hamilton, A. J. S. 2001, MNRAS, 322, 419

Hasinger, G., Burg, R., Giacconi, R., Hartner, G., Schmidt, M., Trumper, J., & Zamorani, G. 1993, A&A , 275, 1

Hasinger, G., Burg, R., Giacconi, R., Schmidt, M., Trumper, J., & Zamorani, G. 1998, A&A , 329, 482

Hasinger, G., et al. 2001, A&A , 365, L45

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

Hasinger, G. et al. 2007, ApJS , this volume

Hawkins, E., et al. 2003, MNRAS, 346, 78

Henry, J. P., & Briel, U. G. 1991, A&A , 246, L14

Henry, J. P., Mullis, C. R., Voges, W., Böhringer, H., Briel, U. G., Gioia, I. M., & Huchra, J. P. 2006, ApJS , 162, 304

Jeltema, T. E., Mulchaey, J. S., Lubin, L. M., Rosati, P., Böhringer, H. 2006, ApJ , 649, 649

Johnson, O., Best, P. N., & Almaini, O. 2003, MNRAS, 343, 924

Kaiser, N. 1984, ApJL , 284, L9

Kaiser, N. 1987, MNRAS, 227, 1

Kenter, A., et al. 2005, ApJS , 161, 9

Kim, D.-W., et al. 2004, ApJ , 600, 59

La Franca, F., et al. 2005, ApJ , 635, 864

Landy, S. D., & Szalay, A. S. 1993, ApJ , 412, 64

Ledlow, M. J., & Owen, F. N. 1996, AJ , 112, 9

Lehmer, B. D., et al. 2005, ApJS , 161, 21

Lilje, P. B., & Efstathiou, G. 1988, MNRAS, 231, 635

Lilly, S. J. et al. 2007, ApJS , in press

Ling, E. N., Barrow, J. D., & Frenk, C. S. 1986, MNRAS, 223, 21P

Loaring, N. S., et al. 2005, MNRAS, 362, 1371

Lumb, D. H., Warwick, R. S., Page, M., & De Luca, A. 2002, A&A , 389, 93

Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, ApJ , 326, 680

Madau, P., Ghisellini, G., & Fabian, A. C. 1994, MNRAS, 270, L17

Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 43P

Mainieri, V., et al. 2007, ApJS , this volume

Manners, J. C., et al. 2003, MNRAS, 343, 293

Marinoni, C., et al. 2005, A&A , 442, 801

Martini, P., Kelson, D. D., Kim, E., Mulchaey, J. S., & Athey, A. A. 2006, ApJ , 644, 116

Martini, P., Mulchaey, J. S., & Kelson, D. D. 2007, ArXiv e-prints, 704, arXiv:0704.3455

Mather, J. C., et al. 1990, ApJL , 354, L37

Mc Cracken, et al. 2007, ApJS , this volume

Menci, N., Fiore, F., Perola, G. C., & Cavaliere, A. 2004, ApJ , 606, 58

Miyaji, T., Hasinger, G., & Schmidt, M. 2000, A&A , 353, 25

Miyaji, T., et al. 2007, ApJS , this volume

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347

Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, Nature, 379, 613

Moretti, A., Campana, S., Lazzati, D., & Tagliaferri, G. 2003, ApJ , 588, 696

Mullis, C. R., Henry, J. P., Gioia, I. M., Böhringer, H., Briel, U. G., Voges, W., & Huchra, J. P. 2004, ApJ , 617, 192

Mullis, C. R., et al. 2004, ApJ , 607, 175

Murray, S. S., et al. 2005, ApJS , 161, 1

Mushotzky, R. F., Cowie, L. L., Barger, A. J., & Arnaud, K. A. 2000, Nature, 404, 459

Nandra, K., et al. 2005, MNRAS, 356, 568

Peebles, P. J. E. 1980, The large-scale structure of the universe (Princeton, N.J., Princeton University Press)

Popesso, P., & Biviano, A. 2006, A&A , 460, L23

Porciani, C., Magliocchetti, M., & Norberg, P. 2004, MNRAS, 355, 1010

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1986 ,Numerical Recipes in Fortran, Cambrige University Press

Puccetti et al. 2006, A&A , 457, 501

Roche, N., Shanks, T., Metcalfe, N., & Fong, R. 1993, MNRAS, 263, 360

Roche, N., & Eales, S. A. 1999, MNRAS, 307, 703

Rosati, P., et al. 2002, ApJ , 566, 667

Ruderman, J. T., & Ebeling, H. 2005, ApJL , 623, L81

Sánchez, A. G., Lambas, D. G., Böhringer, H., & Schuecker, P. 2005, MNRAS, 362, 1225

Scharf, C. A., Jahoda, K., Treyer, M., Lahav, O., Boldt, E., & Piran, T. 2000, ApJ , 544, 49

Shen, Y., et al. 2007, AJ , 133, 2222

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Schmitt, J. H. M. M., & Maccacaro, T. 1986, ApJ , 310, 334

Schuecker, P., & Boehringer, H. 1998, A&A , 339, 315

Schuecker, P., et al. 2001, A&A , 368, 86

Scoville, N.Z. et al., 2007, ApJS , in press

Setti, G., & Woltjer, L. 1973, IAU Symp. 55: X- and Gamma-Ray Astronomy, 55, 208

Spergel, D. N., et al. 2003, ApJS , 148, 175

Tozzi, P., et al. 2006, A&A , 451, 457

Treyer, M., Scharf, C., Lahav, O., Jahoda, K., Boldt, E., & Piran, T. 1998, ApJ , 509, 531

Truemper, J. 1982, Advances in Space Research, 2, 241

Trump, R.J. et al. 2007, ApJS , in press

Ueda, Y., et al. 1999, ApJ , 518, 656

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

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

Vikhlinin, A., & Forman, W. 1995, ApJL , 455, L109

Virani, S. N., Treister, E., Urry, C. M., & Gawiser, E. 2006, AJ , 131, 2373

Voges, W., Henry, J. P., Briel, U. G., Böhringer, H., Mullis, C. R., Gioia, I. M., & Huchra, J. P. 2001, ApJL , 553, L119

Yang, Y., Mushotzky, R. F., Barger, A. J., Cowie, L. L., Sanders, D. B., & Steffen, A. T. 2003, ApJL , 585, L85

Yang, Y., Mushotzky, R. F., Steffen, A. T., Barger, A. J., & Cowie, L. L. 2004, AJ , 128, 1501

Yang, Y., Mushotzky, R. F., Barger, A. J., & Cowie, L. L. 2006, ApJ , 645, 68

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

Worsley, M. A., et al. 2005, MNRAS, 357, 1281

Zehavi, I., et al. 2004, ApJ , 608, 16