High energy studies of active galactic nuclei with *XMM-Newton*

Rebecca Jane Smith

Supervisor:
Gordon Stewart

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester.

X-ray & Observational Astronomy Group

Department of Physics and Astronomy

University of Leicester

September 2008

Declaration

I hereby declare that no part of this thesis has been previously submitted to this or any other University as part of the requirement for a higher degree. I carried out the work described herein except for contributions from colleagues as acknowledged in the text.

Rebecca Smith.

September 2008.

Publications

Smith, R. and Vaughan, S. (2007). X-ray and optical variability of Seyfert 1 galaxies as observed with *XMM-Newton*. *MNRAS*, **375**, 1479-1487

Markowitz, A., Reeves, J. N., George, I. M., Braito, V., Smith, R., Vaughan, S., Arévalo, P. and Tombesi, F. (2008). A comprehensive X-ray spectral analysis of the Seyfert 1.5 NGC 3227. *arXiv:0810.1249*, Accepted for publication in ApJ

Acknowledgements

My first and biggest thank you must go to my long suffering supervisor, Gordon. Thank you for sticking with me and never tiring of answering my many questions, for being there when I've needed you but letting me find my own way at the same time. You were the perfect supervisor for me and I've really enjoyed being your student.

Other big academic thank yous must go to Mark Simpson, Simon Vaughan and Silvia Mateos for their very useful help and assistance with parts of this thesis. Also thank you to Jonny Tedds and Nicolas Lodieu for supplying some of the data used. Thanks to Kim Page for much miscellaneous help over the years and for doing the final big print for me.

To my many office mates over the years (even Levan!), thanks for brightening my days and making coming to work more fun than it should be. Thanks also to the poker players, dressing-up buddies and Lost watchers - good times! Huge thanks in particular to Ann-Marie for being a fabulous friend for the past 4 years.

Massive thanks to the Lab Ladies - to Cat for her never failing honesty, to Gail for never letting life get boring and to Rhian for sharing my brain and my love of trashy TV and for staying up with me for 'the longest night' - you're a true friend. It's safe to say that I got through the last few years with a smile on my face thanks to you three. Thanks for being great friends. I mustn't forget to say thank you to Papillon as well of course.

Thanks to Helen, one of my oldest friends, for always seeming happy to listen to me moan and for many yummy evenings out back in Teesside.

A massive thank you to my loving family - Mum, Dad and sister Charlotte - for always being proud of me, supporting me and never questioning my choices. I love you all so much.

Finally to Ben. I love you like you wouldn't believe. You mean the world to me and I can't wait to marry you. Thanks for pushing me to succeed, for always believing in me and for always being just a little bit crazy.

CONTENTS

Chapter 1

Introduction

It is currently a golden age for X-ray astronomy with both NASA and ESA having dedicated X-ray telescopes in orbit (*Chandra* and *XMM-Newton* respectively). In 2005, the Japanese space agency (JAXA) launched their latest X-ray observatory, *Suzaku*, and there is also an X-ray telescope aboard the gamma-ray observing satellite, *Swift*. These telescopes are currently providing the X-ray observing community with huge amounts of data but this will not be the case forever. *XMM-Newton* has already well surpassed its intended mission length but, along with *Chandra*, will continue to operate through to approximately 2012. While complementary to *Chandra* and *XMM-Newton*, *Suzaku* does not have the high quality spatial or spectral resolution required to replace these missions when they do finally retire and the X-ray telescope on *Swift* is primarily for use in gamma-ray burst observations. The next large scale X-ray mission planned, the *International X-ray Observatory* (a joint mission between NASA, ESA and JAXA), is currently only in the early planning phase and will not launch until 2018 at the earliest.

This could mean a long period of time when no new X-ray observations are possible. Fortunately, thanks to missions such as *XMM-Newton* and *Chandra*, there are now large archives of X-ray observations from which data can be extracted to study interesting and important science goals.

In this thesis, archival data is used to explore X-ray properties of active galactic nuclei (AGN) through the use of surveys.

1.1 X-ray surveys

The first X-ray all-sky survey was made in the 1970s by *Uhuru*, the first dedicated X-ray satellite mission, and detected 339 sources (Forman *et al.* 1978). *Uhuru* showed that the X-ray Background (XRB) emission, which had been detected in 1962 by Giacconi *et al.* (1962), was extragalactic in origin. Only with the use of imaging telescopes was the discrete nature of the XRB revealed.

In 1978, *Einstein*, the first fully imaging X-ray telescope, was launched. *Einstein* performed the first medium and deep X-ray surveys. Deep *Einstein* observations in the $1-3\,\mathrm{keV}$ band (flux limit $\approx 3\times10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) resolved around 25% of the XRB into discrete sources (Giacconi *et al.* 1979).

The next major X-ray satellite to focus on survey work was *ROSAT*, launched in 1990. *ROSAT* observed soft band $(0.1-2.5\,\mathrm{keV})$ X-rays and spent the first six months of its mission conducting the first X-ray imaging all sky survey. Over 100000 discrete

X-ray sources were detected, demonstrating the improvement in X-ray telescope technology since the days of *Uhuru*. *ROSAT* surveys also helped resolve much of the soft band ($1 - 2\,\mathrm{keV}$) XRB. The *ROSAT* Deep Survey (RDS) with limiting flux of $5.5 \times 10^{-15}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ resolved 70 to 80% of the $0.5 - 2\,\mathrm{keV}$ XRB (Hasinger *et al.* 1998). Optical follow up studies showed that $\approx 80\%$ of the sources detected in the RDS were AGN (Schmidt *et al.* 1998). The RDS was followed by the *ROSAT* Ultra Deep Survey (UDS) which detected 94 sources with flux greater than $1.2 \times 10^{-15}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the $0.5 - 2\,\mathrm{keV}$ band and again a large proportion (75%) of the sources were shown to be AGN (Lehmann *et al.* 2001).

The Japanese X-ray satellite, *ASCA*, was launched in 1993 and surveyed harder wavebands than those probed by *ROSAT*. A survey in the $2 - 10\,\mathrm{keV}$ band down to limiting flux of $2 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ resolved $\approx 35\%$ of the hard band XRB (Gendreau *et al.* 1998).

The launches of *XMM-Newton* and *Chandra* in 1999 with their unprecedented angular resolution and fields of view have led to almost complete resolution of the soft band XRB. For example, the Chandra Deep Fields (CDF) achieved limiting flux of $\approx 2.5 \times 10^{-17}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the $0.5 - 2\,\mathrm{keV}$ band and $2 \times 10^{-16}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the $2 - 8\,\mathrm{keV}$ band and resolved $\approx 90\%$ of the XRB (Bauer *et al.* 2004).

1.2 X-ray emission mechanisms

It is important to understand the main X-ray emission mechanisms that are relevant to astrophysics: bremsstrahlung, synchrotron and inverse Compton scattering.

Bremsstrahlung radiation occurs when a charged particle is decelerated by an atom. Thermal bremsstrahlung is the term given to emission from a plasma of thermal particles (as opposed to a single charged particle). The radiation of each separate encounter is highly polarised but there is no preferred direction in general so the resulting radiation from the ensemble will not be polarized. Bremsstrahlung is observed from regions of hot gas e.g. intracluster gas in galaxy clusters. The soft X-ray component of AGN may also be due to bremsstrahlung radiation.

Synchrotron radiation is emitted by relativistic electrons moving in a magnetic field. In this case, the radiation is highly polarized in the direction of the electron's motion. Non-thermal synchrotron radiation caused by electrons with a power law distribution has a power law spectrum (resulting from the superposition of the individual electron power laws).

Inverse Compton scattering radiation occurs when low energy photons interact with high energy electrons and gain energy. This is thought to be an important mechanism in the production of X-rays in AGN - ultraviolet (UV)/optical photons from the accretion disk are thought to be scattered by relativistic electrons in the corona producing a power law X-ray spectrum. The hard X-ray tail of AGN spectra is believed to be caused by

Compton upscattering of soft photons in a hot plasma above the accretion disk.

A process known as Synchrotron Self Compton (SSC) is also thought to be important in AGN - when the density is high enough, the photons emitted due to synchrotron radiation are inverse Compton scattered up to X-ray energies off the electrons that caused the synchrotron radiation in the first place.

1.3 Active galactic nuclei

Active galactic nuclei (AGN) are the brightest non-transient objects in the Universe with bolometric luminosity of between 10^{40} and $10^{48}\,\mathrm{erg\,s^{-1}}$. This large luminosity means they are observable out to very high redshift (current record is $z = 6.43$, Fan *et al.* 2003) and therefore make excellent probes of the early Universe.

1.3.1 Basic characteristics

AGN are typically characterised by high luminosity, activity across the whole electromagnetic (EM) spectrum, rapid variability and, in many cases, broad emission lines. Variability across the whole EM spectrum is a key characteristic of AGN and is most rapid in the X-ray regime. This indicates that the X-ray emitting region is close to the source of the AGN's power. Light travel time arguments from X-ray variability suggest that the size of the power source must be, astronomically speaking, very small (less than the size of the Solar System). This initially created a problem for astronomers - if the large distances to quasars were correct then what could be so small and yet produce enough power to allow such large luminosity? Nuclear reactions, the power source of stars, can not produce the energy required but accretion is a more efficient source of energy. It is now widely accepted that the source of AGN energy is from accretion onto a central supermassive black hole (SMBH). As material is pulled into the SMBH, its potential energy is converted to kinetic energy. An accretion disk forms around the black hole to disperse angular momentum by viscous drag (Lynden-Bell 1969). It is the accretion disk from which optical and UV emission is thought to originate. Drag dissipates heat, increasing the temperature of the disk and giving rise to thermal photons.

1.3.2 Spectral features

AGN are observed in all wavebands, from radio to gamma-rays. The overall shape of the spectral energy distribution can be roughly described by a power law of the form:

$$F_\nu \propto \nu^{-\alpha} \tag{1.1}$$

where F_ν is the flux at frequency ν and α is typically observed to be between zero and one.

The optical spectrum of an AGN shows, in general, a thermal continuum with what is referred to as the ‘Big Blue Bump’ (BBB) - an excess of energy at short wavelengths thought to be due to thermal emission from the hot accretion disk. Strong broad and narrow emission lines are also an important feature. In the infra-red (IR), the emission is almost all thermal and thought to be due to absorption and re-emission by dust in the central regions of the AGN. The IR to optical spectrum forms a roughly continuous power law, peaking at around 100μm.

Around 10% of AGN are strong radio sources with a power law spectrum in the radio regime formed through synchrotron radiation. For radio quiet sources, the energy detected rapidly decreases into the radio regime from the IR power law. Radio loud AGN have either single or double sided jets of energetic particles emerging from the central region and extending beyond the optical extent of the galaxy. The jets are believed to emit synchrotron and Compton scattered radiation from radio to gamma-ray wavelengths.

All AGN are luminous X-ray sources and this thesis focuses on the X-ray regime. The X-ray part of the spectrum can be characterised by a power law. Other features in the X-ray region are a soft excess below around $2\,$keV, a broad hump around $20-30\,$keV and a $6.4\,$keV Fe Kα fluorescence line. The soft excess is thought to be thermal emission from the accretion disk and may be an extension of the BBB. The broad hump and iron line are signatures of reflection suggesting that some of the hard X-rays are reprocessed by reflection from the accretion disk (Reynolds 1998). Very rapid variability at X-ray wavelengths suggests that the X-rays originate close to the SMBH in the innermost regions of the accretion disk. Because of the proximity to the central engine, X-ray observations are very important in understanding how AGN work. Gamma-rays, thought to arise from inverse Compton scattering of photons in the jet of radio loud

AGN, also produce a power law spectrum.

There are gaps in the spectral coverage of AGN in the extreme UV (EUV) and at mm wavelengths. The EUV emission is blocked due to the opacity of the ISM (interstellar medium) in the Milky Way and the Earth's atmosphere is opaque to much of the mm band radiation.

As well as emission lines, high redshift quasars have many narrow absorption lines in their spectra. These lines are from material between the quasar and Earth and provide a way of mapping the distribution of material throughout the Universe. An example is the Lyman α forest - a series of absorption lines all due to the Lyman α transition but redshifted by different amounts because they arise from material at different distances between Earth and the quasar.

Figure 1.1 shows a very basic approximation of a typical AGN spectral energy distribution (SED).

1.3.3 Classification

AGN are split into various categories based mainly on their optical properties (as these were the first to be observed) and luminosity. The main types are Seyfert galaxies, quasars, blazars and radio galaxies.

Seyfert galaxies, named after Carl Seyfert who identified them as a new class of object (Seyfert 1943), are generally (comparatively) low luminosity and therefore low redshift

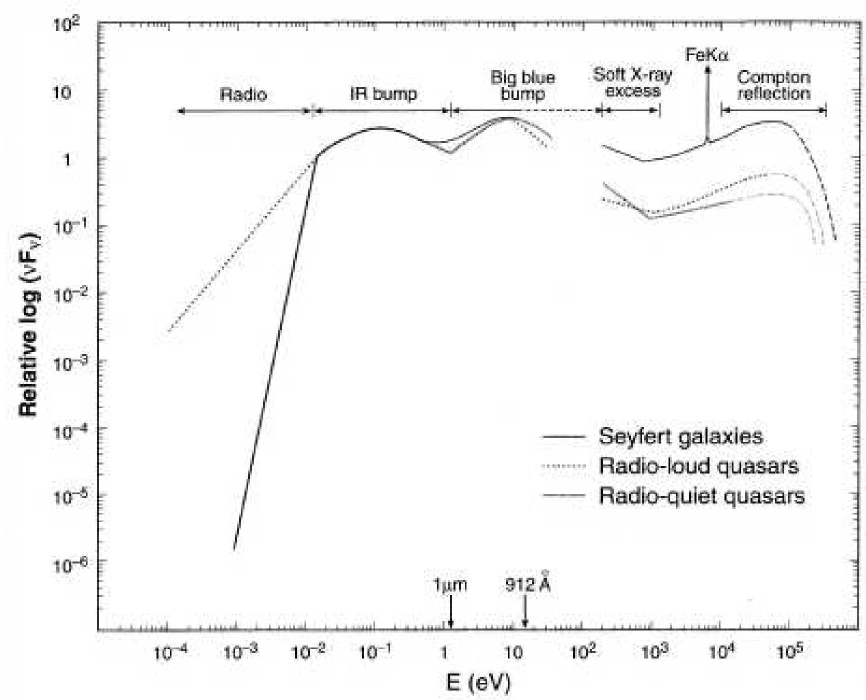


Figure 1.1: Basic representation of an AGN SED. Image taken from Koratkar and Blaes (1999).

AGN. The host galaxy is usually visible and there is a high central surface brightness. The luminosity of Seyfert galaxies is typically in the range $10^{41} - 10^{44}\,\mathrm{erg\,s^{-1}}$. Seyferts are split into two sub-categories - Seyfert 1 and Seyfert 2. Seyfert 1 sources show both broad permitted lines (Full Width Half Maximum (FWHM) $= 1 - 10000\,\mathrm{km\,s^{-1}}$) and narrow permitted and forbidden lines (FWHM $< 1000\,\mathrm{km\,s^{-1}}$) in their spectra but Seyfert 2 galaxies show only the narrow lines. The term 'narrow' in this context is in comparison to the broad lines - the narrow lines are still broader than those in normal galaxies. Some Seyfert galaxies are given a classification between 1 and 2 (1.5, 1.8 or 1.9) depending on the relative strengths of the broad and narrow lines (Osterbrock 1981). Due to variability, a Seyfert's subclass can change over time. The host galaxies of Seyferts are mainly spirals and often have disturbed morphology or are close to other galaxies (e.g. Simkin *et al.* 1980, MacKenty 1990). This discovery has led to the theory that galaxy mergers and interactions may be responsible for triggering activity in galaxies (e.g. Bekki and Noguchi 1994).

Around two decades after the discovery of Seyferts, another class of AGN - the quasar or QSO (quasi-stellar object)[1] - was identified (Schmidt 1963). Quasars are more luminous than Seyferts ($10^{44} - 10^{46}\,\mathrm{erg\,s^{-1}}$) and are at greater distances. The host galaxy of quasars is usually undetected because it is swamped by the high luminosity of the nucleus. Despite the difference in luminosity and distance, the spectra of quasars are similar to that of Seyfert type 1s and, in fact, the properties of Seyferts and quasars overlap, with the highest luminosity Seyferts being indistinguishable from low luminosity quasars. An arbitrary cut off in magnitude has been adopted to distinguish between a Seyfert and a quasar - objects with V band magnitude $M_v < -23$ are classed as quasars and fainter objects are Seyferts. The distinction in X-ray luminosity between quasars and Seyferts is at $L_x = 10^{44}\,\mathrm{erg\,s^{-1}}$. As with Seyferts, there are type 1 and type 2 quasars with type 2 quasars being detected only relatively recently (e.g. Stern *et al.* 2002, Norman *et al.* 2002). Like Seyfert 1s, type 1 QSOs show both broad and narrow emission lines and type 2 QSOs show only narrow lines. Despite the fact that the first quasars were detected due to their radio emission, only around $5 - 10\%$ of quasars are radio loud. Unlike the host galaxies of Seyferts, both radio loud and radio quiet QSOs

[1] The term quasar technically refers to radio loud objects and QSO to radio quiet but they are often used interchangeably

appear to live predominantly in elliptical galaxies and show no evidence for enhanced signs of interaction compared to quiescent elliptical galaxies (Dunlop *et al.* 2003).

Blazar is the term given to two other classes of AGN - BL Lacs and OVVs (Optically Violent Variable quasars). All blazars are radio loud sources. Blazars show the non-thermal spectra characteristic of AGN but other than that show a strong featureless continuum with weak (OVV) or no (BL Lac) emission lines. Blazars show rapid variability at all wavelengths and are strongly polarized in optical light. Blazars are referred to as type 0 AGN.

Radio galaxies are also AGN. These galaxies have radio loud jets and lobes extending to large distances from the central galaxy. Like Seyferts and quasars, there are two types of radio galaxy - broad line (BLRG) and narrow line (NLRG). The NLRGs are split into the two Fanaroff-Riley types - FR1 and FR2 (Fanaroff and Riley 1974). FR1 galaxies are lower luminosity and have bright nuclei but weak radio lobes and double sided jets. In contrast, FR2 galaxies are high luminosity and often have only single sided jets (although some show a faint counter jet). They have weak nuclei but bright radio lobes. Most radio galaxies are hosted by elliptical galaxies. Only around 10% of AGN are radio loud.

Figure 1.2 shows example Seyfert 1, Seyfert 2, quasar and BL Lac spectra.

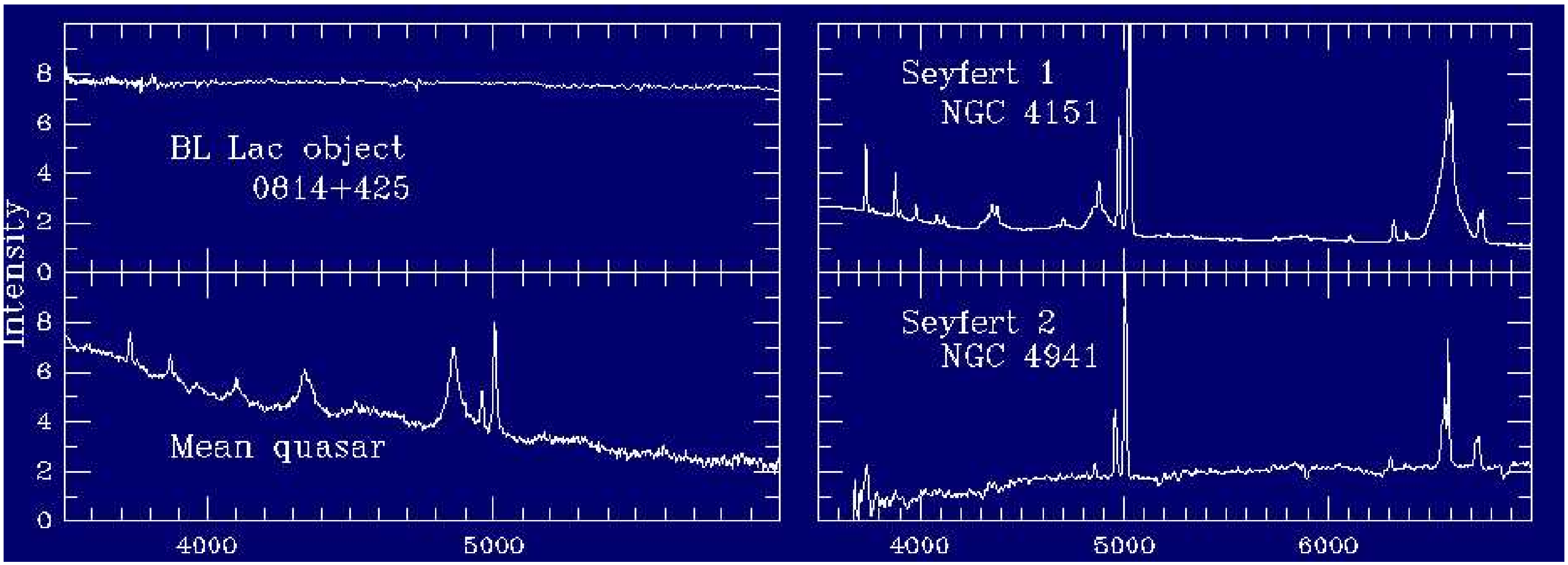


Figure 1.2: Examples of different AGN spectra. All spectra are shown in the emission rest frame for ease of comparison. Image is taken from www.astr.ua.edu/keel/agn/spectra.html.

1.3.4 Unification

It is popularly believed, with increasing evidence, that the different classes of AGN are all in fact a single type of object viewed from different angles. The currently accepted picture is of a SMBH ($M_{BH} > 10^6\,\mathrm{M}_\odot$) surrounded by an accretion disk. There is also a hot, optically thin plasma forming a corona around the black hole from which X-rays are thought to be emitted. Close to the accretion disk (around $1\,\mathrm{pc}$ from the central engine) is the BLR (Broad Line Region), a dense region of clumpy, rapidly moving gas clouds from which the broad spectral lines originate. Further out ($10\,\mathrm{pc}$ - $1\,\mathrm{kpc}$) is the NLR (Narrow Line Region), a less dense, slower moving gas region producing the narrow lines. The key component is a molecular torus which surrounds the central region (SMBH, accretion disk and BLR). When observed from ‘above’ (i.e. looking into the torus) both the broad and narrow line regions are exposed and a type 1 object is seen but, when observing through the torus, the BLR is obscured and a type 2 object is seen. A similar scenario is proposed for radio loud AGN with the addition of jets, perpendicular to the torus, emerging from the central region which extend for 10s of kiloparsecs. The apparent one sidedness of jets in some sources can also be explained by the unification scheme - all radio sources have both jets but in some cases the orientation causes one jet to be brightened by relativistic beaming and the other one appears faint or ‘invisible’ in comparison. Blazars are thought to be seen if the line of sight of the observer is along the jet, where the emission is relativistically beamed. The beaming causes the jet to appear brighter and also exaggerates variability, explaining the rapid variability seen in blazars. Figure 1.3 shows the radio loud and radio quiet unification schemes.

One of the key pieces of evidence supporting the unification model is the detection of broad lines visible in polarized light from type 2 objects (e.g. in observations of the Seyfert 2 galaxy, NGC 1068, by Antonucci and Miller 1985). The polarization results from reflection of light from the hidden Seyfert 1 nucleus. Unification can not currently account for the difference in radio loudness between sources or for why QSOs are more luminous than Seyfert galaxies. The fundamental difference between Seyferts and QSOs could be the accretion rate and/or the mass or spin of the central black hole. The unified models of radio quiet and radio loud objects are reviewed in Antonucci (1993) and Urry and Padovani (1995) respectively.

1.3.5 Variability

Variability on timescales of hours up to years is a key property of AGN and can be used in a variety of ways to determine details of AGN. The rapid X-ray variability detected in quasars led to the discovery that the central engine of AGN must be relatively small and that the X-ray emission region must be close to the central engine by light travel time arguments.

The variability of the different wavelength regimes can be used to examine connections between their emission mechanisms. For example, the optical and UV respond approximately simultaneously to changes suggesting they come from the same region. The IR variability is similar to the optical/UV but with smaller amplitude and a slight time

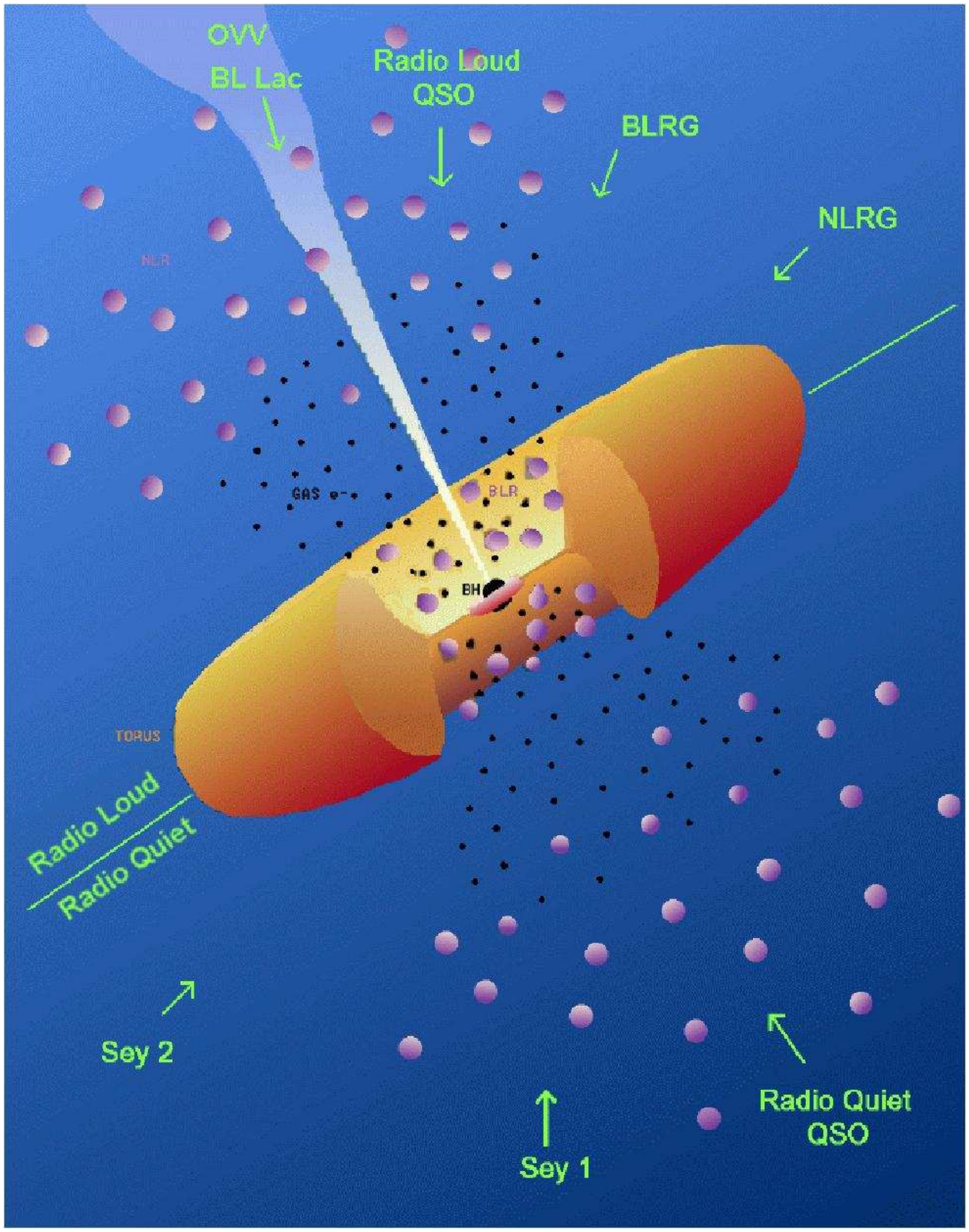


Figure 1.3: Schematic of the AGN unification scheme for radio loud and radio quiet objects. Adapted from Urry and Padovani (1995). Taken from www.star.le.ac.uk/˜kpa/work/research.html.

delay (e.g. Clavel *et al.* 1989). This is consistent with dust around the central source being the source of IR emission. The delay between UV and IR variations is due to the distance between the emission regions. In Chapter 5 the connection between X-ray and optical/UV emission is investigated by examining the variability of light curves in the two bands. The two regions are expected to be connected by either reprocessing of X-rays absorbed by the accretion disk to optical/UV emission (e.g. Guilbert and Rees 1988) or by optical/UV photons entering the corona above the accretion disk and being Compton upscattered into X-rays (e.g. Haardt and Maraschi 1993). Studying the variability of the two regions with respect to each other should reveal if either of these mechanisms is occurring in AGN.

As well as continuum variations, AGN also show variations in the emission line flux which is correlated with the continuum emission. This property can be used to determine the mass of the central black hole and also the geometry and kinematics of the broad line emitting region. This is known as reverberation mapping. Reverberation mapping studies have now determined the central black hole mass of many AGN and have also shown evidence that the broad line region has radial ionization stratification (e.g. Peterson 1993).

1.3.6 AGN surveys

QSOs are visible to high redshifts because of their high luminosity and this makes them excellent probes of the early Universe. Large surveys are essential to determine the evolution and properties of AGN with redshift.

Optical surveys often use the UV excess method for identifying QSO candidates. Sources are selected because of their ‘blueness’, arising from the BBB. Most quasars are brightest in the rest frame near-UV near the Lyman α emission. The UV excess method works well up to around $z = 2$. At higher redshifts, Lyman α emission is redshifted beyond the U band reducing the effectiveness of the selection. Multi-colour selection can find higher redshift AGN because of the difference between stellar and AGN SEDs. Optical surveys are biased against the type 2 population of AGN because of obscuration of emission by the torus. Optical surveys of AGN out to $z \approx 5$ found that the QSO number count peaks at $z \approx 2$ and that QSO evolution could be described by Pure Luminosity Evolution (PLE), i.e. the shape of the luminosity function is constant but shifts to greater luminosity at higher redshifts (e.g. Richards *et al.* 2006).

X-ray surveys are vital to get a more complete sample of AGN. X-ray surveys are more efficient for detecting AGN because they are less affected by absorption than optical surveys and they also detect less of the host galaxy starlight which can dilute the AGN signal. X-ray surveys are also important to resolve the X-ray Background radiation (XRB), as discussed in section 1.1. The advent of *Chandra* and *XMM-Newton* with their superior imaging and spatial resolution have revolutionised QSO surveys. Previous X-ray telescopes had very poor resolution and so matching X-ray sources to optical sources for spectroscopic identification was very difficult. This problem is greatly reduced for *Chandra* and *XMM-Newton* observations and so large X-ray surveys with multi-wavelength follow up are now possible. Multi-wavelength identification studies reveal that most X-ray sources in deep *Chandra* and *XMM-Newton* surveys are AGN (more than 75%, Barcons *et al.* 2007). A low redshift, hard X-ray survey by Barger *et al.* (2005) found that the X-ray luminosity function of AGN agrees well with the bright end of the optical luminosity function in keeping with the PLE scenario. However, X-ray surveys covering larger redshift ranges suggest that low luminosity ($L_x \approx 10^{43}\,\mathrm{erg\,s^{-1}}$) AGN peak around redshift of 0.7 but, in agreement with optical surveys, that high luminosity AGN ($L_x \geq 10^{45}\,\mathrm{erg\,s^{-1}}$) peak at $z \approx 2$ (e.g. Hasinger *et al.* 2005, La Franca *et al.* 2005). These results suggest a Luminosity Dependent Density Evolution (LDDE) where the shape of the X-ray luminosity function changes with redshift.

While almost 100% of the XRB is now resolved below $5\,\mathrm{keV}$, the unresolved fraction increases with energy. The spectrum of the XRB is very hard, peaking at around $40\,\mathrm{keV}$. While the summation of type 1 AGN spectra can recreate the soft end of the XRB it cannot explain the hard end. This led to the realization that there must be a large population of objects with harder spectra to account for the rest of the XRB. Absorption of soft X-rays causes hard spectra and so many AGN are expected to be obscured. Spectral synthesis models have predicted that, at high redshift, the number of obscured sources should be around 4-10 times the number of unobscured sources (Gilli *et al.* 2001). For a time this population of heavily obscured sources remained unobserved but, in recent years, detections of type 2 quasars have been made (e.g. Stern *et al.* 2002, Barger *et al.* 2003). The number of spectroscopically confirmed type 2 quasars is low because these objects are found in hard X-ray and infra-red surveys but are often very faint at optical wavelengths.

There are still many questions to be answered with surveys of AGN. Further details of the quasar luminosity and mass functions will help to determine the accretion history of the Universe. The formation mechanisms of AGN are still unclear. One theory is that interactions and mergers in the early Universe help to fuel AGN - surveys of the environment of AGN out to high redshift would help to determine if this is the case. Surveys could also help to determine the typical opening angle of the proposed molecular torus in AGN by determining the number of objects in different AGN classes. Such a study can not be performed by targeted observations alone.

Both deep and large area surveys will be important in future AGN studies. Deep surveys are required to find the very faintest QSOs and detect the missing population of obscured AGN. Shallower, large area surveys are important for e.g. clustering studies (as will be demonstrated in Chapter 6).

1.3.7 Clustering and large scale structure

Studies of the large scale structure of the Universe are important as they provide information on how the Universe evolved. On small scales, stars are grouped into galaxies which in turn form clusters of galaxies. On larger scales the clusters form superclusters which are separated by large voids. Analysing the clustering of sources will help understanding of how the structure seen today grew from the initial inhomogeneities observed in the Cosmic Microwave Background.

While studies of galaxy clustering have been possible for years, quasar clustering surveys have only recently become possible now that large scale surveys are being conducted. Until recently, quasar clustering studies were conducted using optically selected samples but large X-ray surveys are now becoming possible thanks to *Chandra* and *XMM-Newton*. Clustering studies usually make use of either the spatial or angular two point correlation function to characterise clustering.

An important aim of clustering studies is to measure the bias parameter. Bias is a measure of how the observable matter (e.g. galaxies, quasars) traces the underlying dark matter. Bias can then be used to estimate the mass of the dark matter haloes in which quasars reside by assuming that the halo mass is the main source of the bias. Understanding the properties of dark matter haloes is important for models of galaxy formation and large scale structure. Models connecting the bias and dark matter halo mass have been proposed by e.g. Sheth and Tormen (1999) and Sheth *et al.* (2001).

Recent results (e.g. Croom *et al.* 2005, Myers *et al.* 2007) suggest that quasars reside in massive dark matter haloes ($10^{12} - 10^{14}\,\mathrm{M}_{\odot}$) and are biased tracers of the underlying dark matter with some studies finding an evolution in the bias with redshift. The *XMM-Newton* archive is used here in Chapter 6 to investigate quasar clustering using a new technique.

1.4 Thesis aims

This thesis aims to show that archival X-ray data can be used to create large samples of AGN and produce valid scientific results. Various techniques and topics are covered to show the potential broad uses of archive data.

In Chapter 2 a broad overview of the instruments and catalogues used in this thesis is given. A photometric redshift technique for use with X-ray selected quasars is developed in Chapter 3 and used to determine photometric redshifts for a large sample of objects from the 2XMM catalogue. In Chapter 4, techniques developed in Chapter 3 are used to study the X-ray light curves of serendipitous objects in the 3C 273 field. The simultaneous optical and X-ray observing capability of *XMM-Newton* is used in Chapter 5 to investigate the connection between optical and X-ray variability in a sample of Seyfert 1 galaxies. In Chapter 6, the 2XMMp catalogue is used to investigate cosmic variance and AGN bias. The main conclusions of the thesis are summarised in Chapter 7.

Chapter 2

Instrumentation and catalogues

2.1 Introduction

The majority of the data presented in this thesis was extracted from observations made by *XMM-Newton*, the European Space Agency's (ESA) X-ray observatory. X-ray and optical data were downloaded from the *XMM-Newton* Science Archive (XSA) and both the 2XMMp and 2XMM Serendipitous Source Catalogues were used to compile the samples discussed here. To complement the *XMM-Newton* data, optical data from the Sloan Digital Sky Survey (SDSS) was used. These instruments and catalogues are described in more detail in this chapter.

2.2 *XMM-Newton* instrumentation

XMM-Newton (Jansen *et al.* 2001) was launched by ESA on December 10th 1999 from French Guiana on an Ariane 5 rocket. Originally planned as a two year mission, the lifetime of *XMM-Newton* has now been extended to at least the end of 2012. *XMM-Newton* carries three co-aligned X-ray telescopes and an optical/UV telescope, all four of which can be operated simultaneously.

X-rays are more energetic than optical photons so they require a different telescope system. Using a normal reflecting system does not work as the X-rays simply pass through the mirror. The solution is to use grazing angles to bring the X-rays to a focus. The X-ray telescopes in *XMM-Newton* are Wolter type 1 and are each made up of 58 nested mirror shells. The Wolter type 1 X-ray imaging system comprises a paraboloid and hyperboloid mirror to focus X-rays to a point (see figure 2.1). Using the paraboloid and hyperboloid mirror system reduces off axis blurring of sources. Nesting the mirrors allows for a large effective area (around twice the size of *Chandra*) and therefore better sensitivity.

At the primary focus of each of the X-ray telescopes is a CCD detector. Two of the detectors are MOS CCDs (referred to as MOS1 and MOS2, Turner *et al.* 2001) and at the third is a pn CCD detector (Strüder *et al.* 2001). These detectors together are referred to as the European Photon Imaging Camera (EPIC). In the light path of the two telescopes with MOS detectors are grating assemblies which divert around 40% of the light received towards the Reflection Grating Spectrometer (RGS) detectors at the secondary

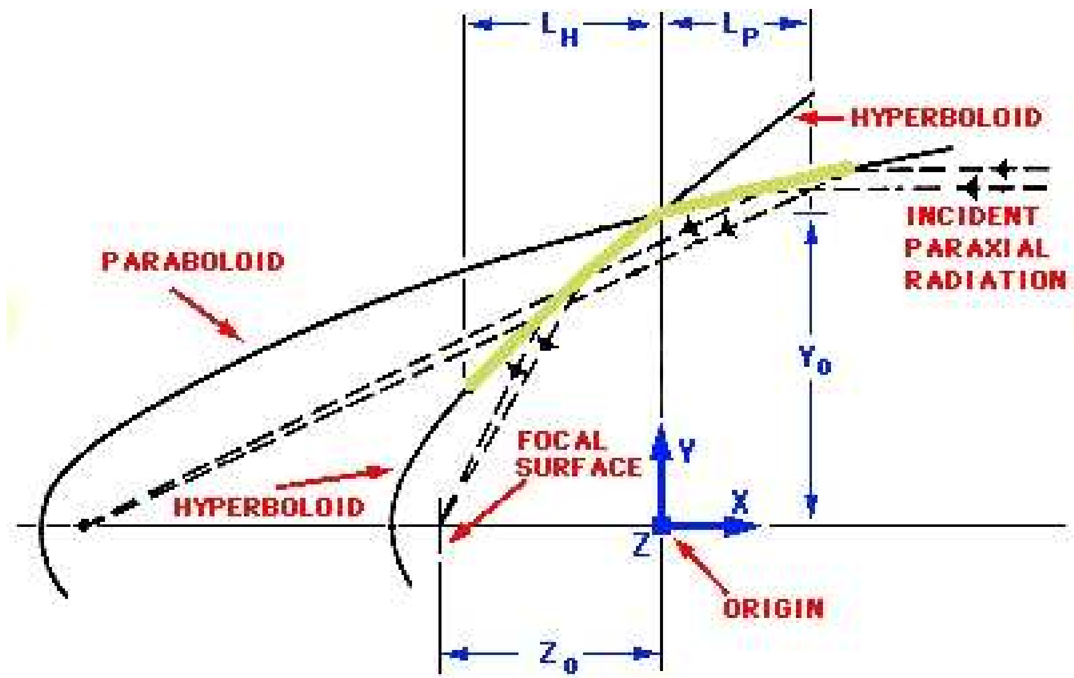


Figure 2.1: The mirrors in *XMM-Newton* use the Wolter type 1 system illustrated here. Image from xmm.esac.esa.int/external/xmm_user_support/documentation/build/index.shtml

focus (den Herder *et al.* 2001). Around 44% of the remaining light reaches the MOS detectors with the rest being absorbed by structures around the detector and reflection grating. The RGS detectors provide high resolution spectroscopy. The optical/UV telescope is called the Optical Monitor (OM, Mason *et al.* 2001). Figure 2.2 shows the relative positions of each of the instrument. Data from EPIC and OM were used in this thesis so more detail on these instruments is given in sections 2.2.1 and 2.2.2.

One of the advantages of *XMM-Newton* is that all six instruments (the three EPIC cameras, two RGS detectors and OM) can be operated simultaneously, allowing multi-

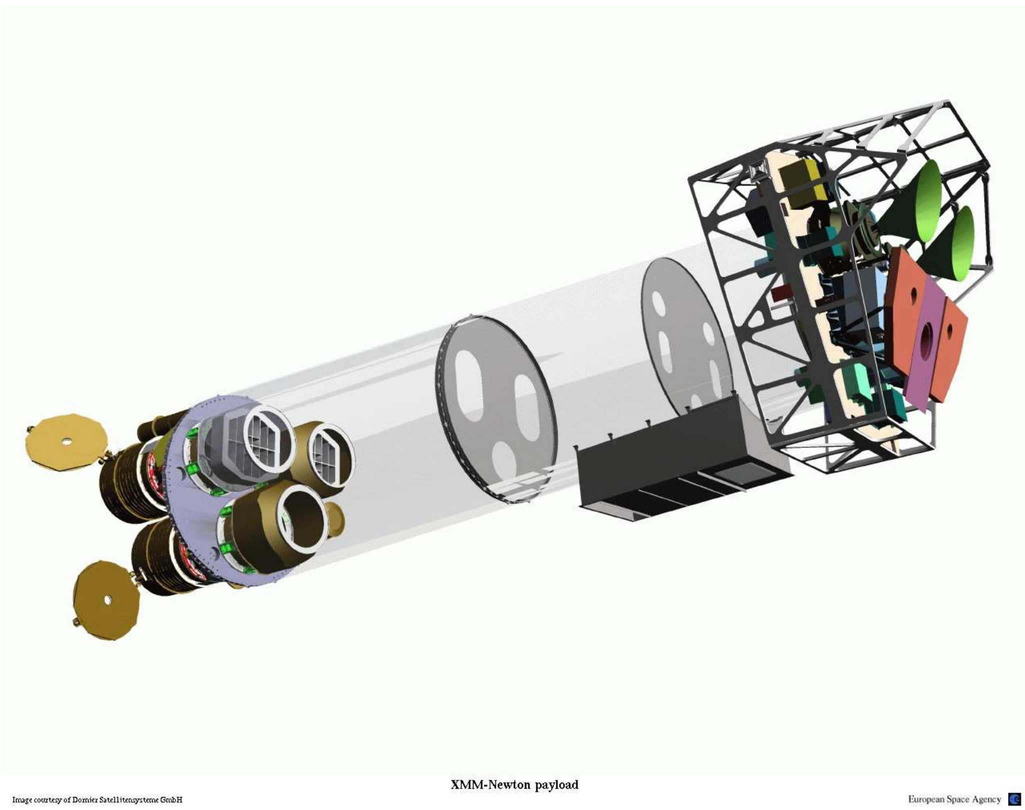


Figure 2.2: The left-hand side of the image shows the three X-ray telescopes and (just visible behind the lower X-ray telescope) the Optical Monitor. The diffraction gratings diverting light to the RGS detectors are visible on the top two X-ray telescopes. At the right hand side are the detectors: the MOS (green), pn (purple) and RGS detectors (orange). Image courtesy of Dornier Satellitensysteme GmbH and ESA.

waveband observations to be made. This capability is exploited in Chapter 5. Other advantages of the satellite include its large effective area and good angular resolution (point spread function (PSF) of around $6''$ FWHM). *XMM-Newton* has the largest ever effective area for an X-ray focusing telescope with $1550\,\mathrm{cm}^2$ for each telescope at $1.5\,$keV giving a total effective area of $4650\,\mathrm{cm}^2$. *XMM-Newton* is in a highly elliptical orbit (apogee $\approx 115000\,$km, perigee $\approx 6000\,$km) providing $\approx 132\,$ks (37 hours) of scientific observation time per orbit. This gives the potential for long continuous target visibility which is useful for variability studies of sources. In contrast to *XMM-Newton*, NASA's X-ray satellite, *Chandra*, has a smaller area ($555\,\mathrm{cm}^2$ for the imaging camera, ACIS-S) but higher spatial resolution ($0.2''$ FWHM). Although also capable of X-ray imaging and spectroscopy, *Chandra* cannot perform both types of observation simultaneously.

2.2.1 European Photon Imaging Camera

As mentioned above, EPIC consists of two types of detector - MOS and pn. MOS1 and MOS2 are each made up of seven Metal Oxide Semiconductor (MOS) front illuminated CCD chips. They cover the bandpass $0.15-12\,$keV and have a $30'$ diameter field of view. The central CCD includes the focal point and the total CCD array has a $2.5 \times 2.5\,$cm imaging area, covering $28.4'$ diameter. On axis spatial resolution is $\approx 5''$ FWHM.

The pn camera is a single silicon wafer with twelve integrated rear illuminated pn CCD chips. The pn has a slightly wider bandpass than the MOS cameras, covering

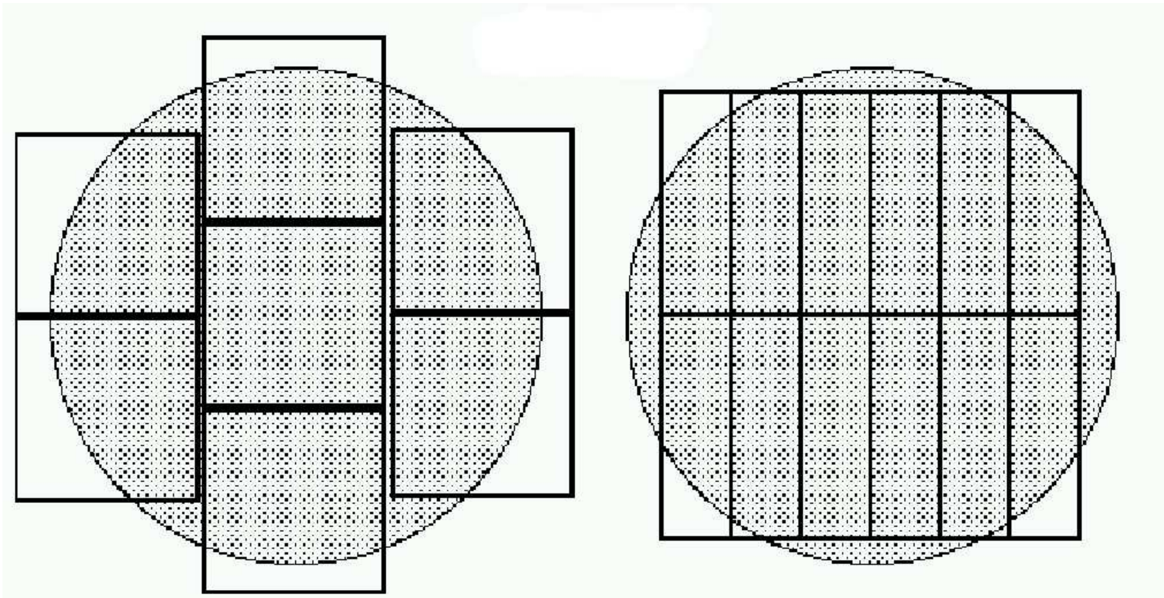


Figure 2.3: The image shows the layout of the MOS CCDs (left) and the pn CCDs (right). MOS2 is rotated 90 degrees with respect to MOS1 to cover the chip gaps. The grey circle in each case is a $30'$ diameter area. Taken from the *XMM-Newton* Users Handbook (xmm.esac.esa.int/external/xmm_user_support/documentation/uhb), section 3.3.

$0.15 - 15$ keV. The wafer is split into four quadrants, each with three CCDs. As the centre of the pn detector does not coincide with a chip, the focal point is offset to chip 0 in quadrant 1. Readout time for the pn is quicker than for the MOS cameras (73 ms compared to 2.6 s).

The positions of the chips in the two camera types are shown in figure 2.3. The MOS1 and MOS2 cameras are oriented orthogonally to one another so that the chip gaps in one are covered by chips in the other.

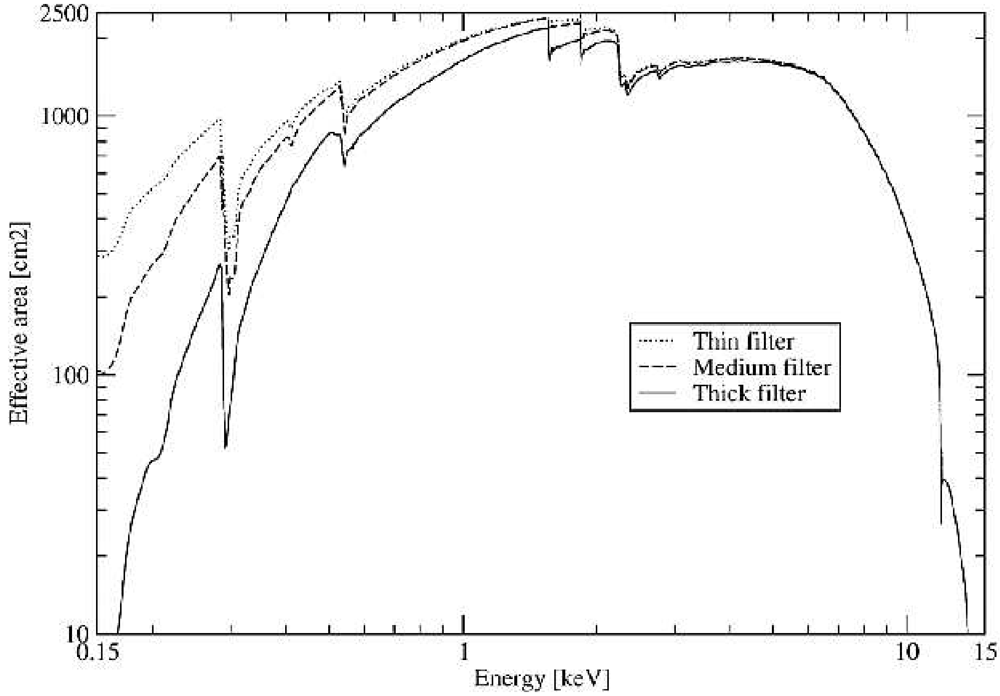


Figure 2.4: The combined EPIC effective area dependent on filter (assuming all three cameras use the same filter). Taken from the *XMM-Newton* Users Handbook (xmm.esac.esa.int/external/xmm_user_support/documentation/uhb), section 3.3.6.

Although designed to detect X-ray photons, the EPIC CCDs are also sensitive to optical photons. To prevent contamination of observations by optical light, three optical blocking filters are available - thin, medium and thick. The combined effective area of EPIC with the different filters is shown in figure 2.4.

The angular resolution of EPIC is determined by its PSF. At $1.5\,$keV, the on-axis PSF of the cameras is $6.6''$, $6.0''$ and $4.5''$ for pn, MOS1 and MOS2 respectively. The PSF varies very little with energy between 0.1 and $6\,$keV but does have a dependence on off-axis angle.

EPIC CCDs can be operated in several different science modes. Full frame mode (or extended full frame for pn) uses the full field of view - all pixels of all CCDs are read out. In partial window mode for MOS cameras, the central CCD can be operated in a different mode to the outer chips. For pn there are two partial window options - large and small window. In large window mode, half the area of all the CCDs is read out. In small window mode only part of CCD4 is used. EPIC CCDs can also be operated in timing mode. For more information on observing modes, see the *XMM-Newton* Users Handbook, section 3.3.2.

EPIC CCDs operate in photon counting mode and the output of the EPIC detectors are event lists - lists of photon events in each detector including time, position and energy of event. These lists can be processed using the *XMM-Newton* Science Analysis System (SAS) to produce images, spectra and lightcurves.

Further details on the EPIC cameras can be found in the *XMM-Newton* Users Handbook, section 3.3.

2.2.2 Optical Monitor

The OM gives *XMM-Newton* the unique capability of being able to observe at X-ray and optical/UV wavelengths simultaneously. The OM is a $30\,$cm Ritchey-Chrétien type telescope which means it has hyperbolic primary and secondary mirrors, designed to combat spherical aberration. The waveband covered is $180 - 600\,$nm with a sensitivity limit of 20.7mag. Photons are detected by a micro-channel plate (MCP) intensified CCD

- two MCPs amplify the photon signal before it reaches the CCD. There are 256×256 pixels available for science, giving a field of view of $17' \times 17'$. The PSF of the OM is $1.4''$ to $2''$ dependent on filter. High time resolution photometry and low resolution grism spectra are obtainable depending on the choice of filter but, unlike the X-ray telescopes, the OM can not be used for imaging and spectroscopy simultaneously. The OM filter wheel contains six broad band filters, two grisms, a white filter and a blocked filter. The broad band filters are U, B and V in the visible wavelength range and UVW1, UVM2 and UVW2 in the UV. One grism is for optical light ($300 - 600\,\mathrm{nm}$) and the other for UV ($180 - 360\,\mathrm{nm}$). The blocked filter is used to protect the OM from overly bright sources - a brightness limit is imposed of $m_v = 7.4$ for an A0 type star in order to avoid damage to the CCD. Filter throughput curves are shown in figure 2.5.

The OM can be operated in two modes - image mode and fast mode. Image mode gives better spatial coverage at the expense of timing information and vice versa for fast mode. Fast mode is more like an X-ray observation in that it produces an event list style output. The trade off between spatial coverage and timing information is necessary due to the limitations of on-board memory. Both modes can be used simultaneously providing the total memory capacity is not exceeded.

Output from the OM is in the form of accumulated images if used in imaging mode and X-ray style event lists if used in fast mode. There are set processing chains in the SAS that can be used to process OM data to produce images, source lists and spectra. The OM image processing chain, `omichain`, performs source detection and photometry. The output from `omichain` is a source list with parameters including positions, count

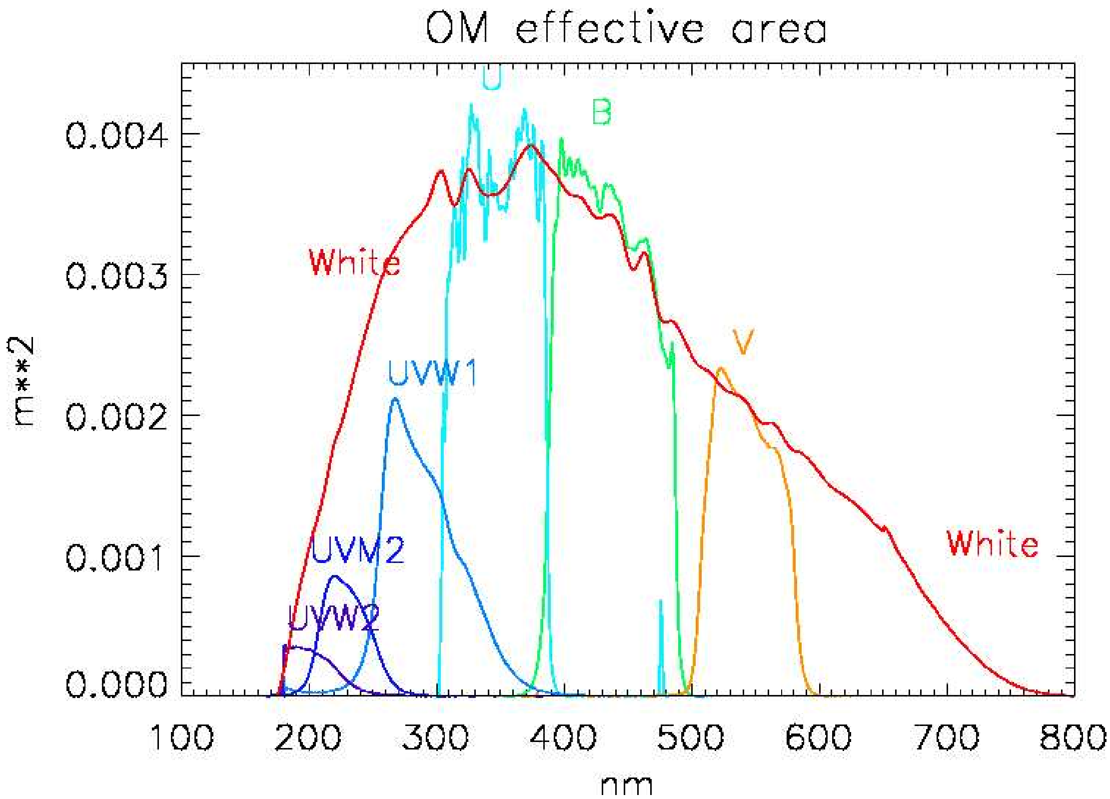


Figure 2.5: Throughput curves for the OM filters folded with the detector sensitivity. Image taken from the *XMM-Newton* Users Handbook, section 3.5.5.1.

rates and instrumental magnitudes.

More details on the OM can be found in the *XMM-Newton* Users Handbook, section 3.5.

2.3 *XMM-Newton* data access

2.3.1 *XMM-Newton* Science Archive

All observations made with *XMM-Newton* are made publicly available through the *XMM-Newton* Science Archive (XSA). Proprietary observations can be searched but are not available to download by anyone, except the principal investigator (PI), until the proprietary period (one year) is over. Observations can be searched using many different parameters e.g. target name, coordinates, observation date etc. Both raw ODF (Observation Data Files) and processed PPS (Pipeline Processing Subsystem) files are available to download and can be filtered before download to reduce file size. EPIC spectra and light curves are not immediately available but can be generated through the XSA interface. Once the required observations have been selected they can be retrieved either by direct download or ftp. The XSA website is xmm.esac.esa.int/xsa

2.3.2 *XMM-Newton* Serendipitous Source Catalogues

XMM-Newton has a large field of view meaning that during a pointed observation (i.e. when the telescope is observing a specific object), many other sources are visible in the field of view. These sources are referred to as serendipitous and represent a significant resource in X-ray astronomy. The 1st *XMM-Newton* Serendipitous Source Catalogue (1XMM) was released in April 2003 and contained 28279 unique sources taken from 585 observations. The 2nd *XMM-Newton* Serendipitous Source Catalogue (2XMM), released in August 2007, is the successor to this. The visual screening required to ensure good quality data was extremely time consuming so, while extensive screening was being carried out, a subset of the data was released, the 2XMM Pre Release (2XMMp).

2XMMp Serendipitous Source Catalogue

2XMMp was released on the July 24th 2006 and contains source detections from 2400 *XMM-Newton* observations taken between February 4th 2000 and April 20th 2006 and covering $400\mathrm{deg}^2$ ($285\mathrm{deg}^2$ when overlapping fields are accounted for). The catalogue includes 153105 detections of 123170 unique sources. 6271 detections are extended of which 5717 are unique extended sources. Sources are included in the catalogue if their EPIC detection likelihood (det_ml) in the $0.2 - 12\,\mathrm{keV}$ band is greater than or equal to six[1]. The median flux of the catalogue in the $0.2 - 12\,\mathrm{keV}$ band is $\approx 2.4 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, more sensitive than 1XMM. Positional accuracy is typically better than $5''$.

2XMMp used the latest processing pipeline, developed to reprocess all existing *XMM-Newton* observations. All fields in the date range were then visually screened and those with deficiencies in the automatic processing which required more extensive screening were excluded from the 2XMMp catalogue (approximately 17% of fields). The catalogue contains source parameters for nine energy bands (a combination of broad and narrow, see table 2.1) including fluxes in each band. Time series and spectra are auto-

[1] Detection likelihood values calculated using the SAS application `emldetect` which is based on the algorithm of Cruddace *et al.* (1988)

Table 2.1: 2XMMp/2XMM energy bands

Band number	Energy range (keV)
(1)	(2)
1	0.2-0.5
2	0.5-1.0
3	1.0-2.0
4	2.0-4.5
5	4.5-12.0
6	0.2-2.0
7	2.0-12.0
8	0.2-12.0
9	0.5-4.5

matically extracted in the $0.2 - 12$ keV range for sources with more than 500 counts in the combined EPIC camera.

More information about the 2XMMp catalogue can be found in the 2XMMp User Guide[2].

[2]Available from http://xmmssc-www.star.le.ac.uk/Catalogue/2XMMp/UserGuide_2xmmp.html

2XMM Serendipitous Source Catalogue

The full 2XMM Serendipitous Source Catalogue was released on August 22nd 2007. The catalogue, which is the largest X-ray source catalogue ever produced, was constructed by the XMM Survey Science Centre (XMM-SSC) based in Leicester. It contains data from 3491 observations taken between February 3rd 2000 and March 31st 2007. All datasets included were publicly available by May 1st 2007 although not all publicly available observations are included - some fields were excluded due to pipeline processing difficulties. 246897 detections (with $0.2 - 12\,\mathrm{keV}$ band EPIC det_ml ≥ 6) of 191870 unique sources are included in the catalogue making it around six times larger than 1XMM and nearly twice as big as 2XMMp. 20837 detections are classed as extended (18804 unique sources). The range in flux of the catalogue is 10^{-9} to $10^{-16}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$. The median flux level of the catalogue is 2.5×10^{-14}, 5.8×10^{-15} and $1.4 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ for the $0.2-12$, $0.2-2$ and $2-12\,\mathrm{keV}$ bands respectively. The catalogue covers a total area of 560deg^2 which equates to 360deg^2 when overlapping fields are accounted for. Source positions are, in general, accurate to better than $5''$. For each source, parameters such as count rate, flux and hardness ratio[3] are available for various energy bands and for each camera as well as a combined EPIC value. Thumbnail images are also available for each source. For sources with more than 500 total band EPIC counts, time series and spectra are automatically generated and included in the catalogue. 38320 sources have these automatically generated products.

[3]Hardness ratio is a camera specific X-ray colour, $HR_n = (R_{n+1} - R_n)/(R_{n+1} + R_n)$ where R is the count rate and n is the energy band ($1-5$ from table 2.1)

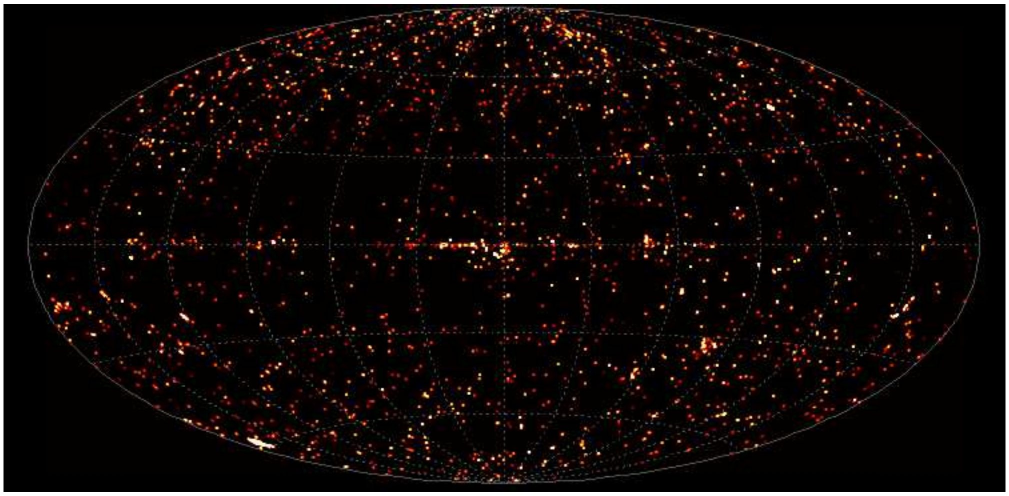


Figure 2.6: Map of the source density of 2XMM. Kindly provided by Professor Mike Watson.

Apart from being much bigger, the 2XMM catalogue differs from 2XMMp principally in its visual screening. In order to facilitate a quick release without compromising on quality, the visual screening of 2XMMp removed any fields with spurious detection regions from the catalogue. The visual screening of 2XMM is more thorough and utilises a manual flagging system to flag regions (not full fields) if detections are spurious making 2XMM a much more inclusive catalogue.

Each detection in the catalogue is assigned a unique ID (detid). Multiple detections identified as being the same source are then assigned a source ID (srcid). A slimline version of the catalogue containing one line for each unique source is available in addition to the full catalogue which contains one line for each detection.

A map of the source density of the 2XMM catalogue is shown in figure 2.6.

More detailed information about how the catalogue was constructed is available in the 2XMM User Guide[4] and the catalogue paper (Watson *et al.* 2009).

Both the 2XMMp and 2XMM catalogues are available in FITS and CSV formats from the user guide websites (see footnotes) and also through the XSA and LEDAS (LEicester Database and Archive Service). The 2XMM catalogue will be increased incrementally on an annual basis to incorporate the latest publicly available observations. The first of these incremental updates was released on August 20th 2008 and includes nearly 30000 new unique sources.

2.4 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) is an optical survey aiming to image more than a quarter of the sky. The majority of fields are in the Northern hemisphere with a small set in the South.

A dedicated 2.5 m reflecting telescope at Apache Point, New Mexico is used to make the observations. The telescope has a roll away enclosure instead of a traditional dome and is protected from light and wind by internal and external baffles (Gunn *et al.* 2006). The detector is a 120 megapixel camera made up of 30 CCDs arranged in five rows. Each CCD is 2048×2048 pixels with one pixel equivalent to $0.4''$ giving a total field of view of 1.5deg^2. Each row of the detector corresponds to a filter - u, g, r, i and

[4]Available at http://xmmssc-www.star.le.ac.uk/Catalogue/2XMM/UserGuide_xmmcat.html

z with central wavelengths $3500, 4800, 6250, 7700$ and 9100 Å respectively (Fukugita *et al.* 1996). During an observation, a given point on the sky passes over each of the five filters in turn. The bands are wide to ensure high efficiency detection of faint objects. The magnitude limits of the filters are $22, 22.2, 22.2, 21.3$ and 20.5 for u, g, r, i and z respectively. Filter curves are shown in figure 2.7. There are also two fibre-fed spectrographs which can each measure spectra for over 600 objects in a single observation. The spectroscopic data will be used to determine accurate distances to objects. This will be particularly useful in the case of distant quasars for cosmological work.

Each source detected is assigned a morphological classification. The two main classifications are star (i.e. point source) or galaxy (i.e. extended source). The classification is assigned based on the magnitude of the source as determined by three different models. The models used are PSF (for a star), de Vaucouleurs (for a typical elliptical galaxy) and exponential (for a disk galaxy). The magnitude assigned by the PSF model is compared to that of the best fit galaxy profile to determine the classification. This classification method is correct at the 95% confidence level to r magnitude of 21.

Over five years, SDSS has imaged over 8000deg^2 on the sky and detected nearly 200 million objects as well as recording spectra for almost a million sources. The first phase of SDSS was completed with Data Release 5 (DR5). A second phase with new goals, SDSS-II, has now begun.

All data from the SDSS is available to the public. The data are released regularly in large chunks - the current release is Data Release 6 (DR6) - with FITS images and spectra

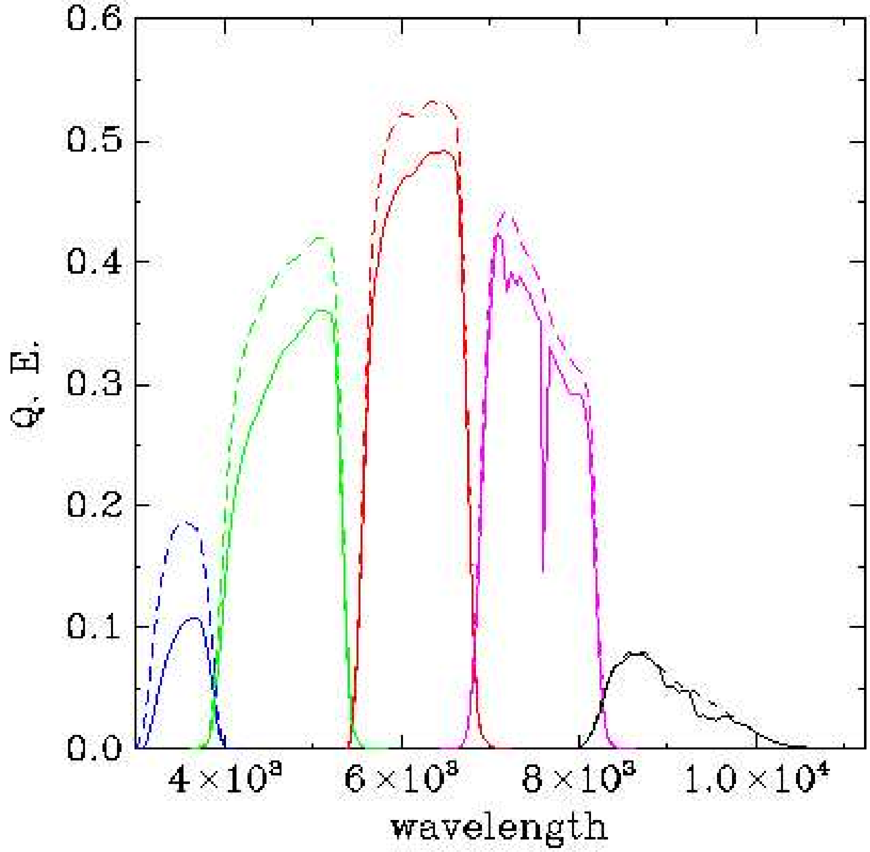


Figure 2.7: The throughput of the SDSS filters. Wavelength is in angstroms. Image taken from www.sdss.org/dr3/instruments/imager/index.html.

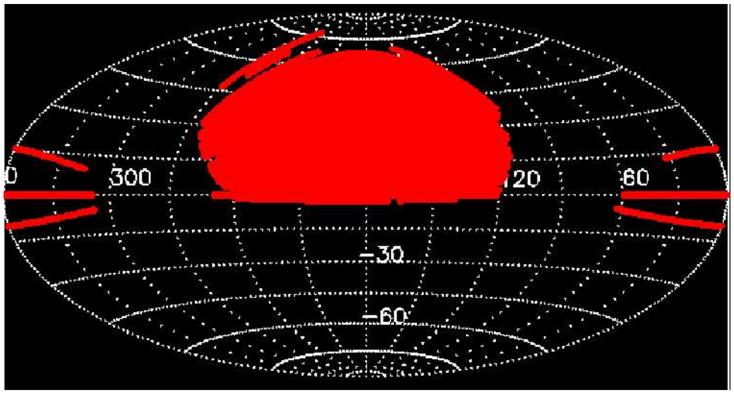


Figure 2.8: The sky coverage of the SDSS imaging program as of Data Release 6. Figure from http://www.sdss.org/dr6/.

available to download. DR6 contains imaging data for 287 million unique objects over 9583deg^2 and spectra of 1271680 sources (Adelman-McCarthy *et al.* 2008). Figure 2.8 shows the area of the sky that has been imaged so far by SDSS.

Part of the SDSS project is to construct a quasar catalogue from the objects detected. The most recent version is the DR5 quasar catalogue (using data from DR5, Schneider *et al.* 2007). This version was used here to extract objects for the photometric redshift method (see Chapter 3) so the following details refer to this version. Sources are automatically selected as quasar candidates based on their position in SDSS colour space. Candidates without spectra are then rejected and the remaining candidates are visually inspected to ensure they are actually quasars. Finally, luminosity and line width criteria are applied to give a final catalogue of 77429 objects covering a redshift range of 0.08 to 5.41. The quasar catalogue contains the usual SDSS data for each object plus details of any radio, X-ray or infra-red objects (from FIRST, ROSAT and 2MASS catalogues respectively) within a set radius from the source. The selection method for the catalogue is designed to specifically target type 1 quasars and not other types of AGN (e.g. type 2 quasars, Seyferts, blazars). The DR5 catalogue is available to download in ASCII or FITS formats from http://www.sdss.org/dr5/products/value_added/qsocat_dr5.html.

SDSS data is used in this thesis to find optical counterparts to *XMM-Newton* sources. The overlap between SDSS and 2XMM can be seen by comparing figures 2.6 and 2.8. Around 27% of 2XMM sources are in the SDSS DR6 image region.

Chapter 3

Optimising photometric redshift techniques for use with serendipitous X-ray selected AGN

3.1 Introduction

The redshift of an astronomical object is a crucial parameter as it enables the distance of the object to be determined. Without a distance estimate any other values obtained, e.g. flux, have less physical significance. With redshift, flux can be converted to intrinsic luminosity, a more useful parameter of a source.

Redshift is the term used for a change in the observed wavelength of light from an object due to the movement of the object relative to the observer. If an object is moving away from the observer, the wavelength of the light is lengthened and therefore reddened (as, in the optical band, longer wavelengths are redder). Wavelengths are shortened (blueshifted) if the object moves towards the observer. As the majority of objects are moving away from us the term redshift is most commonly used. The equation for redshift, z, is:

$$z = \frac{\lambda_{obs} - \lambda_{emit}}{\lambda_{emit}} \tag{3.1}$$

where λ_{obs} is the observed wavelength and λ_{emit} is the wavelength of the light when it was emitted. A negative redshift is therefore a blueshift.

Most astronomical objects are moving away from each other due to the expansion of the Universe (although it is the expansion of space rather than the movement of the galaxies that causes this). Objects that are further away have a larger redshift so redshift can be used as a measure of the distance of an object. For objects up to $z \approx 0.2$, distance, d, is determined by:

$$d = \frac{cz}{H_0} \tag{3.2}$$

where c is the speed of light and H_0 is the value of the Hubble constant. For objects at larger redshifts this equation becomes less accurate and cosmological parameters must be considered.

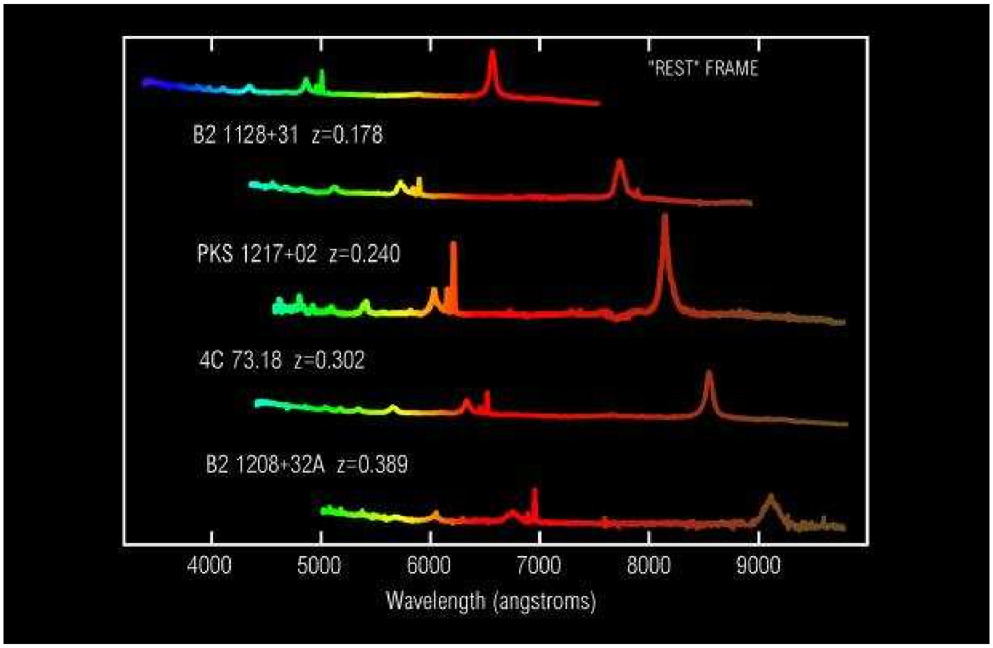


Figure 3.1: Example of quasar spectra at increasing redshifts (from top to bottom) showing the movement of spectral lines away from the expected position (top line) due to redshift. C. Pilachowski, M. Corbin/NOAO/AURA/NSF.

The most accurate method of determining the redshift of an object is spectroscopically. The spectrum of an object shows, in general, an underlying continuum with emission lines and absorption lines and edges superimposed on it. The emission lines indicate which elements are present in the object. Each element has a specific signature of line positions allowing it to be identified in a spectrum.

If an object is moving away from or towards the observer the lines in the spectrum will move according to the redshift (see figure 3.1 for an example). The elements present in the spectrum of an object can be determined from the pattern of the lines and the shift from the expected position (i.e. that measured in the laboratory) can be calculated. This is then used to calculate the redshift of the object using equation 3.1. Although accurate, this method is also time consuming. High quality spectra require a long integration time to achieve a high signal to noise level and so large amounts of telescope time are required for accurate spectroscopic redshifts.

A more efficient, although less accurate, method is to use photometry to determine redshifts. Photometric redshifts use the spectral energy distribution (SED) of an object constructed from photometric filters. Broad features in the spectrum of the object are visible in the passbands and can be used to determine an approximate redshift. As the features are redshifted through the different filters, the observed colours change indicating the redshift of the source. The main features which enable this method to work are the 4000 Å break in galaxy spectra, the 912 Å Lyman break and, for high redshift quasars, the Lyman α forest. Figure 3.2 shows a range of galaxy SED templates. The 4000 Å break can be seen in each of these spectra. This break is caused by absorption of high energy radiation by metals and a lack of young stars and is therefore strongest in old (i.e. elliptical) galaxy spectra. In younger galaxies (spirals and irregulars) an increase of flux in the UV is also apparent due to the large number of young, blue stars in the galaxy. The 912 Å break is visible in normal galaxies at high redshift. A lack of features in the optical band can cause problems for photometric redshifts of quasars between $z = 0.8$ and $z = 2.5$. There are several common methods of implementing photometric redshifts which will be discussed in section 3.2.

Photometric redshift estimations have recently become popular due to the increase in the number of large surveys being conducted at different wavelengths. Surveys such as

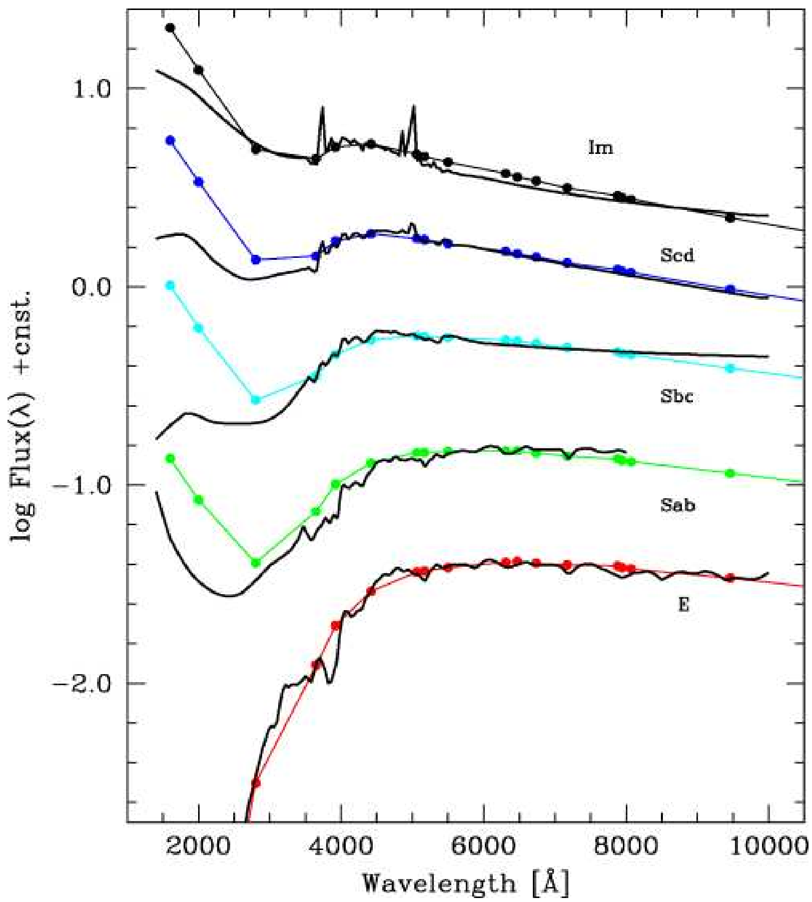


Figure 3.2: Template spectra for several galaxy types showing the broad features which photometric redshift methods use e.g. 4000 Å break. The solid black lines are mean empirical spectra and coloured lines are template spectra. Image taken from Buzzoni (2005).

SDSS and the serendipitous catalogues from *XMM-Newton* contain hundreds of thousands of objects. The large number of objects makes obtaining spectroscopic redshifts prohibitive and even currently impossible for very faint sources. Photometric redshifts on the other hand can be used to determine the approximate redshift of many objects simultaneously, making them a popular option for users of large surveys.

3.2 Photometric redshift methods

There are two main methods of photometric redshift estimation - the template fitting method and the training set method.

For the template fitting method, a range of sample SEDs of various source types is required. These can be either composite spectra from real observations or artificial SEDs from population synthesis models. The templates are shifted to represent a range of redshifts resulting in a set of templates covering different source types over a wide redshift range. The SED of the object which requires redshift determination is then compared to each of the templates until the best match is found, usually by χ^2 minimisation. Template fitting is a relatively simple method and there are several publicly available codes which make use of it, e.g. HyperZ (Bolzonella *et al.* 2000). There are, however, some disadvantages. Composite templates tend to come from nearby objects which may not be a good representation of more distant objects. Real spectra may depend on the age of the source and may therefore be different at different redshifts. Simply adjusting spectra of nearby objects to different redshifts by shifting the features will not provide accurate high redshift spectra if this is the case and so template fitting to high redshifts is difficult.

The other popular method is to use a training set of objects with known spectroscopic redshifts to develop a colour-redshift relationship which can then be applied to a set of objects requiring photometric redshifts. The training set must be representative of the objects requiring photometric redshifts and must use the same photometric filter system so each different survey requires its own training set. The training set method uses real data so does not have the same problem with evolution as the template fitting method. With this method however, it is not easy to extrapolate to redshifts beyond the range of the training set, so a large redshift range must be used in the training set if it is thought to be required for the photometric sample. The main disadvantage of this method is that it is impossible without a representative set of objects with spectroscopic redshifts.

3.3 Previous photometric redshift studies

The first person to suggest the use of photometry to make redshift estimates was Baum (1962) who determined the redshift of eight galaxy clusters by comparing their mean SED to the SED of a cluster of known redshift. In 1985 the idea of photometric redshifts was revisited by Koo (1985) who investigated the accuracy possible for photometric redshifts and concluded, using an early template fitting method, that accuracies of $0.03 - 0.05$ would be possible for redshift from $z = 0$ to above $z = 0.5$. Connolly *et al.* (1995) were the first to propose the training set method. Using four filters and a sample of 254 galaxies with spectroscopic redshift to determine a photometric redshift relation, an accuracy of $\Delta z \leq 0.05$ was achieved where Δz is the difference between spectroscopic and photometric redshifts.

The use of photometric redshifts gained popularity in the last decade due to the increase in the number of large photometric surveys. Various methods and codes for the determination of photometric redshifts of galaxies are now available. For example, Bolzonella *et al.* (2000) developed the publicly available code HyperZ[1], a template fitting code using eight template types and between four and eight filters. Using Hubble Deep Field North (HDF-N) sources, Bolzonella *et al.* (2000) found $\delta_z = 0.09 - 0.26$ depending on the template set and redshift range where

$$\delta_z = \frac{\sigma}{1+\overline{z_s}} = \sqrt{\frac{\Sigma(z_p - z_s)^2}{N-1}}\frac{1}{1+\overline{z_s}} \tag{3.3}$$

where z_s is spectroscopic redshift, z_p is photometric redshift, N is the number of sources, $\overline{z_s}$ is the average spectroscopic redshift and σ is the rms scatter between photometric and spectroscopic redshifts. Benítez (2000) used Bayesian techniques to determine galaxy redshifts for the HDF-N. Bayesian methods have been shown to be potentially more accurate than template fitting techniques when using fewer filters if appropriate prior knowledge is available. Benítez (2000) found the rms of $\Delta z/(1 + z_s)$ to be 0.06 with an average difference between photometric and spectroscopic redshifts of $\overline{\Delta z} = 0.01$. Other methods include artificial neural networks (ANNz, Collister and Lahav 2004) and kernel regression (Wang *et al.* 2007), training set codes which use novel methods to create the colour-redshift relationship. Both claim to produce results with accuracy comparable to more established template fitting routines.

Initially, quasars were thought to be unsuitable for photometric redshift methods because of their mainly featureless spectra. However, strong emission lines at low redshift and the arrival in the optical spectra of the Lyman α forest and Lyman break above

[1]http://webast.ast.obs-mip.fr/hyperz

$z \approx 2.5$ mean that this is not the case. In 2001, Budavári *et al.* (2001) and Richards *et al.* (2001) investigated methods for obtaining photometric redshifts for quasars using the SDSS. Budavári *et al.* (2001) developed a hybrid training and template method resulting in $\sigma_{rms} = 0.77$ where:

$$\sigma_{rms} = \sqrt{\frac{\Sigma(z_s - z_p)^2}{N}} \tag{3.4}$$

Richards *et al.* (2001) used a training set method (which was the basis of the method implemented here) with the SDSS QSO catalogue from the commissioning phase of the survey and found $\Delta z \leq 0.2$ for around 70% of the sample. Another photometric redshift study using quasars (Wu *et al.* 2004) used the SDSS Early Data Release (EDR) QSOs to create a composite spectrum for use as a template. Applying this to a template fitting code resulted in a similar accuracy to Richards *et al.* (2001) with the benefit that the composite spectrum can be applied to other surveys by use of filter transmission functions.

Much less work has been done on the use of photometric redshifts with X-ray selected sources. Gonzalez and Maccarone (2002) used template fitting to derive photometric redshifts for X-ray selected galaxies and concluded that the accuracy achievable was comparable to that of X-ray quiet galaxies ($\sigma_{rms} = 0.15$). Any QSOs in the sample were selected and removed before photometric redshift estimates were made. Gandhi *et al.* (2004) used HyperZ (Bolzonella *et al.* 2000) to determine photometric redshifts of X-ray hard, optically dim sources selected from *Chandra* fields. For the eleven sources for which spectroscopic redshifts were available they found that sources dominated by the host galaxy had good photometric redshifts ($\overline{\Delta z} = 0.14$) but poor results were found for the AGN dominated sources ($\overline{\Delta z} = 0.58$). Kitsionas *et al.* (2005) used galaxy templates for extended sources and QSO templates for point sources for a sample of bright sources from the *XMM-Newton*/2dF survey with optical photometry from SDSS. They found $|\Delta z| \leq 0.3$ for 68% of point sources. For extended sources, results were dependent on g-r colour. Red galaxies ($g - r > 0.5$) were well fit ($\Delta z \leq 0.15$ for 73%) but blue sources ($g - r < 0.5$) were much less accurate with only 27% with fractional error $\Delta z/(1 + z_s) \leq 0.2$. HyperZ and Bayesian analysis were used with sources from the CDF-S (Zheng *et al.* 2004) to find photometric redshift estimates of X-ray selected AGN (types 1 and 2) and galaxies. They found an average dispersion of $\Delta z/(1 + z_s) = 0.08$.

3.4 The photometric redshift code

The 2XMM catalogue contains data on approximately 192000 unique sources. To maximise the information available from this data, redshift information will be essential. As the majority of X-ray sources are AGN, a redshift method that works for X-ray selected quasars as well as galaxies is needed. Weinstein *et al.* (2004) created a photometric redshift code for optically selected quasars specifically for use with the SDSS based on the work of Richards *et al.* (2001). As the SDSS has a large degree of overlap with 2XMM ($\approx 27\%$ between SDSS DR6 and 2XMM) this code is ideal for adaptation to determine photometric redshifts of 2XMM objects.

The code works in three stages.

1. A training catalogue of quasars with spectroscopic redshifts is required. The SDSS DR5 Quasar Catalogue (Schneider *et al.* 2007, see section 2.4) is ideal for this purpose and was used here. The quasars are separated into N redshift bins. When Weinstein *et al.* (2004) plotted the colours of the quasars (u-g etc.) against their spectroscopic redshift it became apparent that some objects were significantly redder than the majority of sources in that redshift bin because of internal dust extinction. The code therefore removes these 'reddened' quasars from the catalogue, on the basis of their u-g and g-r colours, as these objects skew the colour-redshift relation. If a source has u-g colour redder than 97.5% of sources in the same redshift bin and its g-r colour is redder than the median g-r for the redshift bin then the source is classed as 'reddened' and removed from the catalogue.

2. The remaining catalogue objects are used to create a colour-redshift relationship (CZR) by calculating the mean colour vector and the colour covariance matrix. The mean colour vector, $\mathbf{M}_i$ has four components:

$$M_i^j = \frac{1}{Q_i} \sum_{q=1}^{Q_i} x_{j,q} \tag{3.5}$$

where i refers to the redshift bin, Q_i is the number of non-reddened quasars in the i^{th} redshift bin, $x_{j,q}$ is the colour of the q^{th} quasar where j is $1, 2, 3$ or 4 referring to each of the four colours (u-g, g-r, r-i and i-z respectively). The colour covariance matrix, $\mathbf{V}_i$, has sixteen components:

$$V_i^{j,k} = \frac{1}{Q_i - 1} \sum_{q=1}^{Q_i} (x_{j,q} - M_i^j)(x_{k,q} - M_i^k) \tag{3.6}$$

where k represents the colours in the same way as j.

3. The CZR is then used to determine redshift estimates for the set of objects requiring photometric redshifts. The colours of each object are compared to the CZR for each redshift bin and a χ^2 value for each bin is calculated:

$$\chi_i^2 = (\mathbf{X_0} - \mathbf{M_i})^T (\mathbf{V_0} + \mathbf{V_i})^{-1} (\mathbf{X_0} - \mathbf{M_i}) \tag{3.7}$$

for $i = 1 - N$ where $\mathbf{X}_0$ is the vector of colours of the object in question and $\mathbf{V}_0$ is the covariance matrix of the colours:

$$\begin{pmatrix} \sigma_u^2 + \sigma_g^2 & -\sigma_g^2 & 0 & 0 \\ -\sigma_g^2 & \sigma_g^2 + \sigma_r^2 & -\sigma_r^2 & 0 \\ 0 & -\sigma_r^2 & \sigma_r^2 + \sigma_i^2 & -\sigma_i^2 \\ 0 & 0 & -\sigma_i^2 & \sigma_i^2 + \sigma_z^2 \end{pmatrix}$$

where σ_u is the uncertainty in the u band magnitude and similarly for the other bands. The probability that the true value of the redshift is in that redshift bin is calculated based on the χ^2 value and covariance matrices. Regions of the probability-redshift distribution above a given threshold (1/number of redshift bins) are possible photometric redshift ranges. An upper and lower value for each potential redshift range is given based on the first and last bin above the threshold (see figure 3.3). For each range, the assigned photometric redshift value is the one with the highest probability. The overall probability of the range is the sum of probabilities of the bins in the range. The photometric redshift range assigned to the source is the one with the highest overall probability.

More detail on the equations used to calculate χ^2 and probability of redshift bins can be found in Weinstein *et al.* (2004).

3.5 Determining optical matches

The objects in the XMM catalogues are X-ray sources so, to utilise the photometric redshift method described above, matches to optical counterparts in the SDSS are required. To ensure confidence in the optical matches, the reliability method of Rutledge *et al.* (2000) was implemented. This method uses the magnitude of potential counterparts as well as their proximity to the X-ray source to determine the 'true' counterpart. It is assumed that brighter, lower density optical objects are more likely matches to the X-ray sources.

The method is as follows. A $10''$ radius circle around each X-ray position is searched in the SDSS to gather candidate optical counterparts. $10''$ is a generous search region considering the positional accuracy of both SDSS and *XMM-Newton* but no candidates with distance greater than $6''$ are selected by the following method as counterparts for

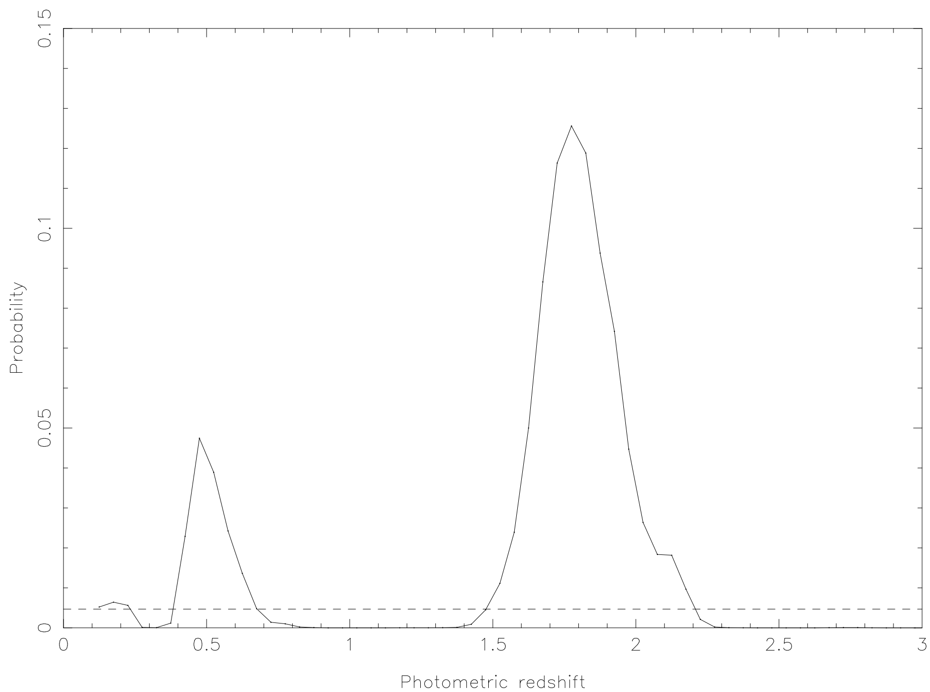


Figure 3.3: For a single example source, the probability that the true redshift lies in a variety of photometric redshift bins. The area under the curve for each of the regions above the threshold (dashed line) is calculated and the peak redshift in the region with the largest value is selected as the photometric redshift estimate (1.8 in this case). The lower and upper bounds for the photometric redshift are the points at which the probability falls back below the threshold value (1.5 and 2.2 in this case).

the final analysis. A large number of $10''$ fields around the outside of the source field are also searched to create a list of background objects with which to compare the candidate objects. Each object in the background and candidate lists has a value, r, which is the distance from the object to the centre of the field. For each background and candidate object, the likelihood ratio, LR, is calculated based on 'Gaussian distributed positional coincidence for sources and an expectation that the counterpart be bright' (Rutledge *et al.* 2000):

$$LR = \frac{e^{-(\frac{r^2}{2\sigma^2})}}{\sigma \frac{N(<m)}{F_{bg}}} \tag{3.8}$$

where σ is the uncertainty in the X-ray position (taken to be $2''$), F_{bg} is the number of background fields used (590) and m is the apparent magnitude of the candidate in question. $N(< m)$ is the number of background objects with magnitude brighter than m. The uncertainty in the optical position is not included as it is negligible compared to the X-ray position uncertainty. The likelihood ratio is the likelihood that the optical candidate is associated with the X-ray source (or, in the case of the background sources, that it is associated with an object at the centre of the field). It should be high for 'true' counterparts but does not represent the probability that an object is the true counterpart.

Candidate objects are then assigned a reliability value, R.

$$R(LR) = \frac{\frac{N_{src}}{F_{src}} - \frac{N_{bg}}{F_{bg}}}{\frac{N_{src}}{F_{src}}} \tag{3.9}$$

where F_{src} and F_{bg} are the number of source and background fields respectively. N_{src} and N_{bg} are the number of candidate and background objects respectively that have a likelihood ratio that falls in the bin $10^{\log(LR_i)-0.6}$ to $10^{\log(LR_i)+0.6}$ where LR_i is the likelihood ratio of the candidate in question. Reliability is the probability that the candidate is associated with the X-ray source but does not consider other potential candidates so more than one candidate in the source field may be assigned the same reliability.

Finally, the probability that a candidate is the unique counterpart of the X-ray source,

P_{id}, is calculated. For each X-ray source the candidate objects are assigned a number, 1 to N. The probability that a candidate is the unique counterpart is:

$$P_{id,i} = \frac{\frac{R_i}{1-R_i}\prod_{j=1}^{N}(1-R_j)}{S} \tag{3.10}$$

where i is the index of the candidate in question and j does not include i. S is the normalisation factor (see below). The probability that none of the candidate objects is the true counterpart is also calculated.

$$P_{no-id} = \frac{\prod_{j=1}^{N}(1-R_j)}{S} \tag{3.11}$$

The normalisation factor S ensures that:

$$\sum_{i=1}^{N} P_{id,i} + P_{no-id} = 1 \tag{3.12}$$

and is calculated as:

$$S = \sum_{i=1}^{N}\left(\frac{R_i}{1-R_i}\prod_{j=1}^{N}(1-R_j)\right) + \prod_{j=1}^{N}(1-R_j) \tag{3.13}$$

where again, in the summation, j does not include i. For each X-ray source, the optical candidate with the highest P_{id} value was assigned as the counterpart. In cases where the P_{no-id} value is greater than any of the P_{id} values, the X-ray source was assigned no counterpart and was not used in the photometric redshift sample.

3.6 Optimizing the method

To determine whether the code described in section 3.4 would be accurate enough for future use with large samples of 2XMM sources, it was applied to X-ray sources in the region of space known as the Lockman Hole (RA=10h52m, dec=$57°21'36''$). The Lockman Hole is a well observed field because of the low Galactic absorption in its direction ($N_H = 5.7 \times 10^{19}\,\mathrm{cm}^{-2}$). Source positions were taken from the paper Mateos *et al.* (2005) with kind permission of Dr Silvia Mateos. The data were from seventeen combined *XMM-Newton* observations obtained between April 2000 and December 2002. The total exposure time was approximately $850\,$ks (MOS cameras) and $650\,$ks (pn camera) after removal of high background periods.

Many of the sources have spectroscopic redshifts which can be compared to the photometric redshifts to determine the accuracy of the code and optimise its usage. Optical matches to the X-ray sources were obtained as described in section 3.5. Of 123 X-ray sources, 105 had candidate counterparts and 95 (77%) were assigned optical counterparts. 70 (57%) counterparts had ID probability (P_{id}) greater than 90%. This compares favourably with the X-ray/optical cross correlation work of Kitsionas *et al.* (2005) who found 60% of their *XMM-Newton* sources to have SDSS optical counterparts. The final sample used contained only the 65 sources which had both optical counterparts (in-

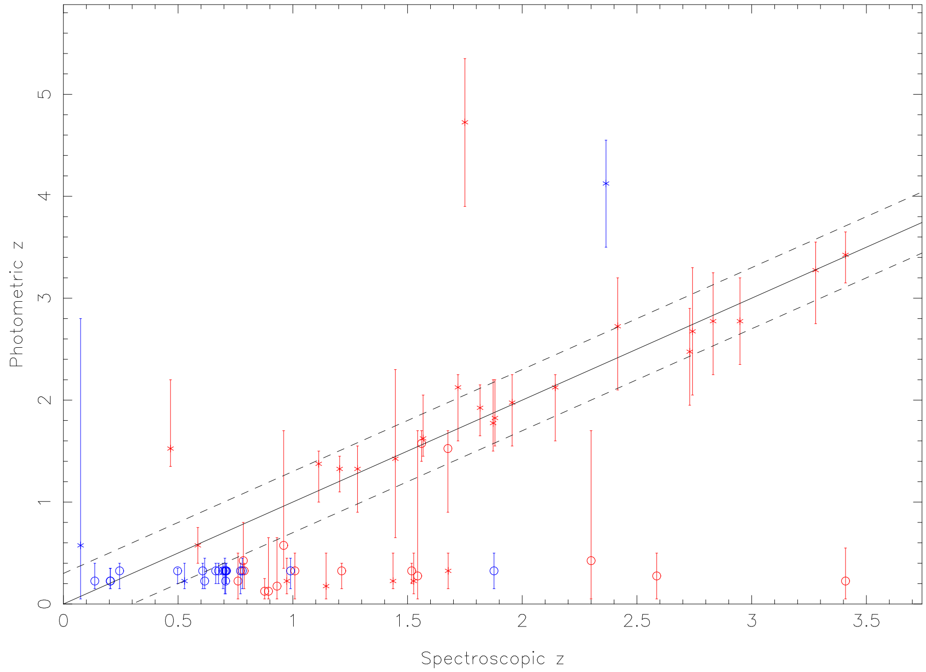


Figure 3.4: The spectroscopic and photometric redshift of each source as assigned by the photometric redshift code using the SDSS DR5 QSO catalogue as training set and default parameters. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

cluding those with $P_{id} < 0.9$) and spectroscopic redshifts. For clarity, X-ray objects are referred to as sources and optical matches to these sources are termed counterparts. Classification as type 1 or type 2 AGN is determined by Mateos *et al.* (2005) and therefore a property of the X-ray sources but morphological classification (point or extended) is a property of the optical counterparts taken from SDSS.

To get an initial determination of the accuracy of the code for a sample of X-ray selected quasars the photometric redshift code was run with default parameters. The DR5 QSO

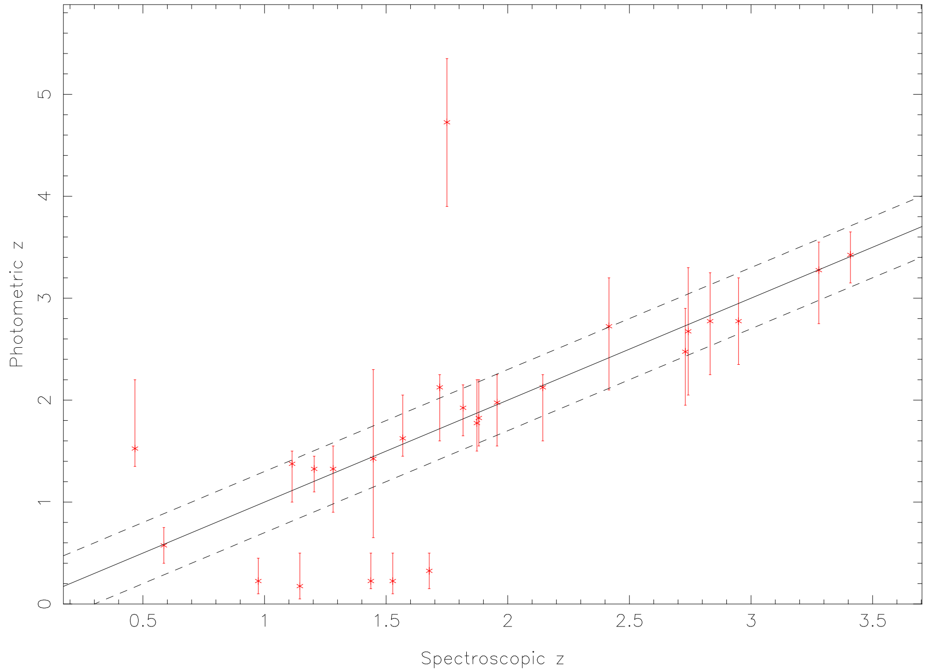


Figure 3.5: The spectroscopic and photometric redshift of each source classified as type 1 AGN and with point counterpart. The photometric redshift is as assigned by the photometric redshift code using the QSO catalogue as the training set. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

catalogue from SDSS (Schneider *et al.* 2007), was used as the training set for the sample giving a result of 44% of sources with the 'correct' photometric redshift. A source is described as having the 'correct' photometric redshift if its spectroscopic redshift lies within the upper and lower photometric redshift limits output by the code (z_{high} and z_{low} respectively). Figure 3.4 shows the results of this initial run.

As the training set is composed of type 1 AGN and mainly objects with point like morphology while the test sample contains both type 1 and type 2 AGN and point and extended counterparts the poor accuracy encountered was anticipated. Running just the sources classified by Mateos *et al.* (2005) as type 1 AGN which have point counterparts with the QSO training set yielded better results (73% 'correct', see figure 3.5).

Objects classified as extended by SDSS are likely to be nearby galaxies. The light received from these counterparts may therefore be dominated by the host galaxy. As extended objects are nearby (in order to appear extended) the code makes it possible to restrict the photometric redshift assigned to them to less than a specified value. Type 2 AGN are also likely to be dominated by the host galaxy light as the central AGN is typically obscured. To determine accurate photometric redshifts for these objects, a catalogue of galaxies was created using the SDSS DR6 release and the DEEP2 catalogue (Deep Extragalactic Evolutionary Probe[2], Davis *et al.* 2003, Davis *et al.* 2007). 61899 sources were randomly selected from the 185706 sources in DR6 that are spectroscopically classified as galaxies and have redshift confidence greater than 90%. The redshift range of these sources is $0.0 - 0.6$ which is not wide enough for the use required here. DEEP2 has collected spectroscopic redshifts of nearly 50000 galaxies to redshifts beyond 1.4 and has some spatial coincidence with SDSS making it a useful catalogue with which to enhance the galaxy catalogue. 31123 sources in the DEEP2 DR3 catalogue have redshift quality ≥ 3 (90% or better confidence). These objects were cross correlated with the SDSS using a $1''$ search radius to find 7522 matches. Using the SDSS photometry for these sources with the DEEP2 spectroscopic redshift values increased the galaxy catalogue to 69422 sources with redshifts out to around two (see table 3.1). The CZR created from this was useful to $z = 1.65$ (above which values could not be

[2]http://deep.berkeley.edu/DR3/dr3.primer.html

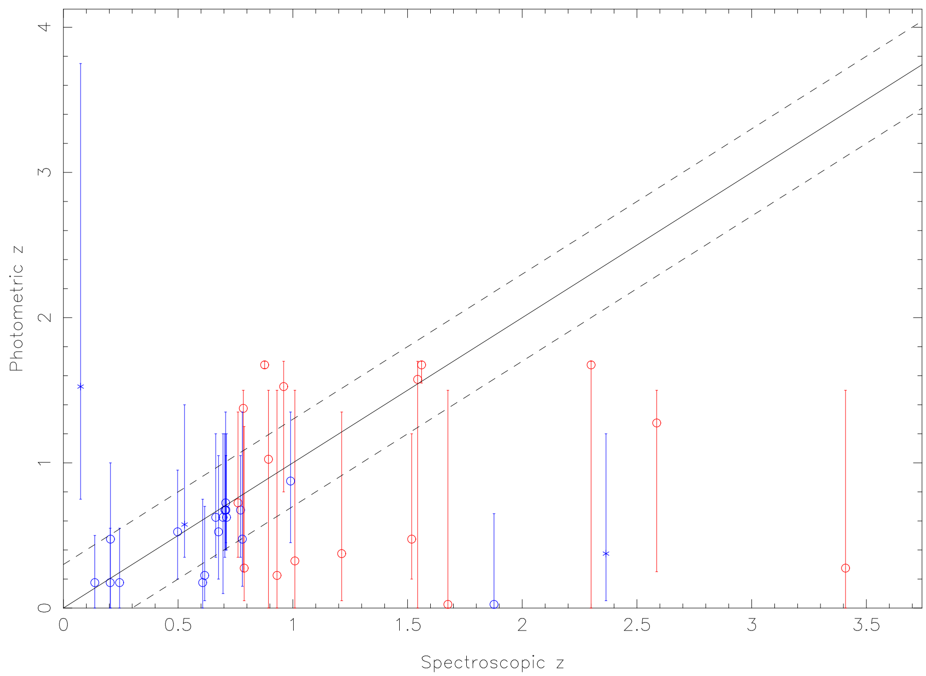


Figure 3.6: The spectroscopic and photometric redshift of each source classified as type 2 or with extended counterpart. The photometric redshift is as assigned by the photometric redshift code using the galaxy catalogue as the training set. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

constrained due to a lack of sources at higher redshifts) and so the extended counterparts were constrained to photometric redshifts less than 1.65. This catalogue was used as the training set for the counterparts classed as extended and/or type 2 AGN sources. 76% of such objects were assigned the 'correct' redshift with this catalogue (see figure 3.6).

To further investigate the effect of source properties on the accuracy of the photometric redshift determination, the X-ray sources were cross correlated with the UKIDSS (UKIRT Infra-red Deep Sky Survey, Lawrence *et al.* 2007) database to look for infra-red matches (based on a search radius of $2''$). For the 36 objects that had both optical and IR counterparts, the z-J colour was calculated. It was determined empirically that photometric redshifts for objects with $z - J > 1.72$ (i.e. redder than average) were all wrongly estimated by the QSO catalogue and this was taken as the cutoff point between 'red' and 'normal' sources for the sample. Given that step one of the photometric redshift code is to remove objects from the training set that are 'reddened', it is unsurprising that red objects in the sample are not assigned the 'correct' photometric redshifts. To rectify this, a third training set was compiled, containing the 'reddened' QSOs that had been removed from the original QSO catalogue (see table 3.1). Although using the red catalogue led to an improvement on the results that were found with the QSO catalogue, all red sources also had extended counterparts or were type 2 AGN and were assigned better estimates of photometric redshift by the galaxy catalogue. For the case of a type 1 AGN source with a point counterpart and $z - J > 1.72$, the red catalogue is expected to perform better than the QSO catalogue. As there are no such sources in the sample used here, use of the red catalogue was not pursued further. UKIDSS is currently in the early stages of its mission. As more data becomes available the opportunity to exploit the data in the way described above will become more important.

The optimal choice of catalogue to use as training set for each source was determined as follows:

- If counterpart is extended and/or source is type 2 AGN: galaxy
- If counterpart is point like and source is type 1 AGN: QSO

Table 3.1: Training set details

Catalogue	Num. sources (original)	Num. sources (after step 1)	z range
(1)	(2)	(3)	(4)
QSO	77292	73910	0-5.5
Galaxy	69422	67140	0-4
Red	3382	2979	0-5.2

A photometric redshift for each object was then determined using the assigned catalogue as training set. This improved results to 75% of sources with the 'correct' photometric redshift (see figure 3.7).

To further improve accuracy, cuts on the counterpart identification probability ($P_{id} \geq 0.9$) and the photometric redshift probability ($P_{phot} \geq 0.8$) were made improving the result to 86% 'correct' (see figure 3.8).

The use of $1/N$ (where N is the number of redshift bins) as the threshold for determination of the photometric redshift range is arbitrary and was adopted following the example of Weinstein *et al.* (2004). While use of this threshold leads to a high percentage of sources assigned the 'correct' photometric redshift (where 'correct' means the spectroscopic redshift is within the range of the photometric redshift), figures 3.8 and 3.9 show that a significant proportion of these 'correct' sources actually have large $|\Delta z|$ values while sources with well estimated photometric redshifts are assigned a large error range which doesn't reflect the actual accuracy of the method. An alternative to using

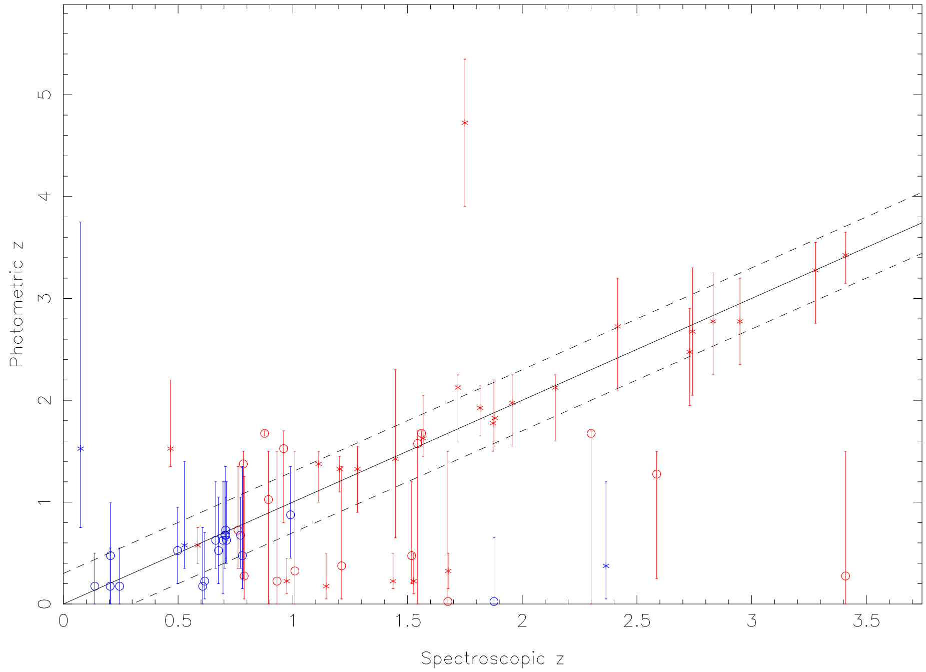


Figure 3.7: The spectroscopic and photometric redshift of each source. The photometric redshift is as assigned by the photometric redshift code using the designated catalogue for each object. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

the $1/N$ threshold is to use the χ^2 values determined for each source to assign standard 1σ confidence regions to each photometric redshift estimate. Using this method will reduce the range of the photometric redshift estimates (i.e. $z_{high} - z_{low}$). The effect of this will be that good estimates of photometric redshift, those with low Δz values, will have more useful error ranges while those sources with large Δz values will no longer be described as 'correct'. This change to the code is implemented as follows: The peak probability used in the original method corresponds to a minimum in the χ^2 distribution,

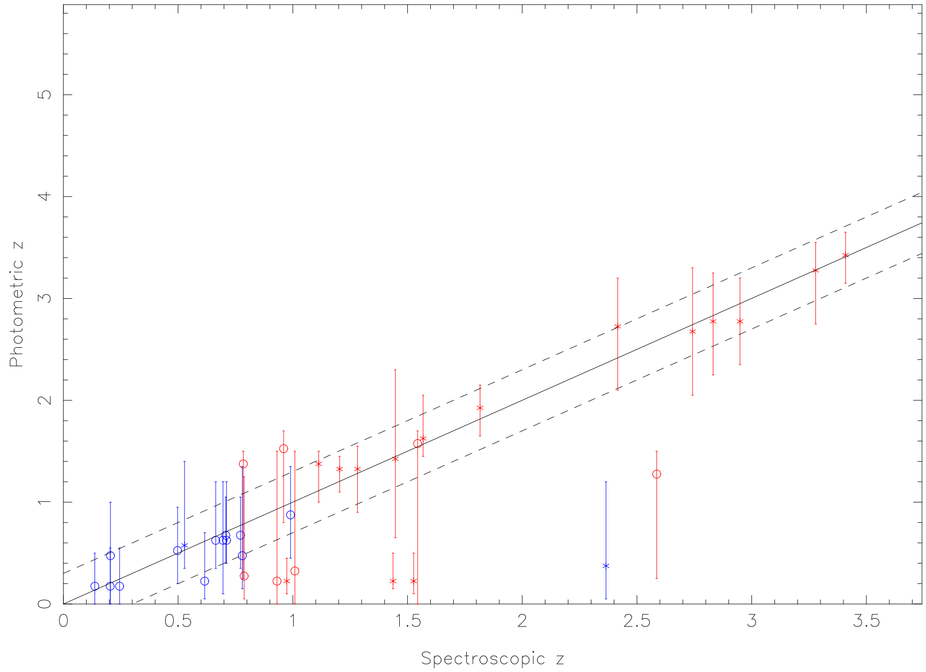


Figure 3.8: The spectroscopic and photometric redshift of each counterpart with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$. The photometric redshift is as assigned by the photometric redshift code using the designated catalogue for each object. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

χ^2_{min}. Finding χ^2_{min} provides the photometric redshift estimate. The threshold for error bars is then χ^2_{min}+1, corresponding to the 68.3% (1σ) confidence region. The redshift bin at which the χ^2 value passes above this threshold on either side of the minimum bin corresponds to the upper and lower redshift limits. This method is illustrated in figure 3.10.

The result of using the modified method is shown in figure 3.11. As the peak probability corresponds to the minimum χ^2 value, the photometric redshift estimate itself does not

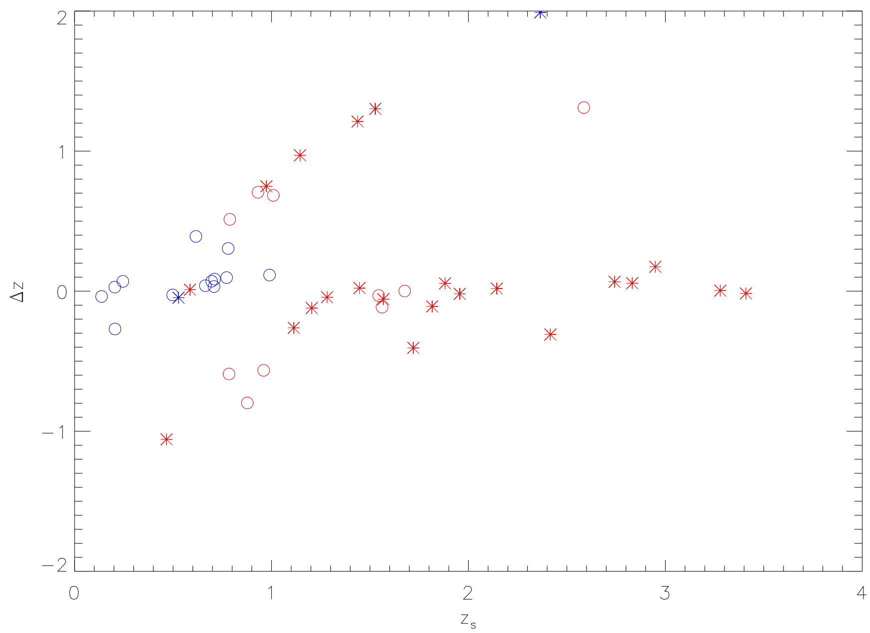


Figure 3.9: The difference between spectroscopic and photometric redshift (Δz) of each counterpart with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$. The photometric redshift is as assigned by the photometric redshift code using the designated catalogue for each object. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts.

change so the percentage of sources with $|\Delta z| \leq 0.3$ for example should not change[3]. The change is evident in the average range ($\overline{z_{high} - z_{low}}$) of the assigned redshifts: 0.29 with the modified code compared to 0.88 with the original (see figure 3.12). This range now represents the 1σ error on the photometric redshift estimates, not just an arbitrary range encompassing a large set of possible values.

[3]In practice there are slight differences in these values between the original and modified method because the cut in P_{phot} is not used with the modified method and so more sources are assigned a photometric redshift.

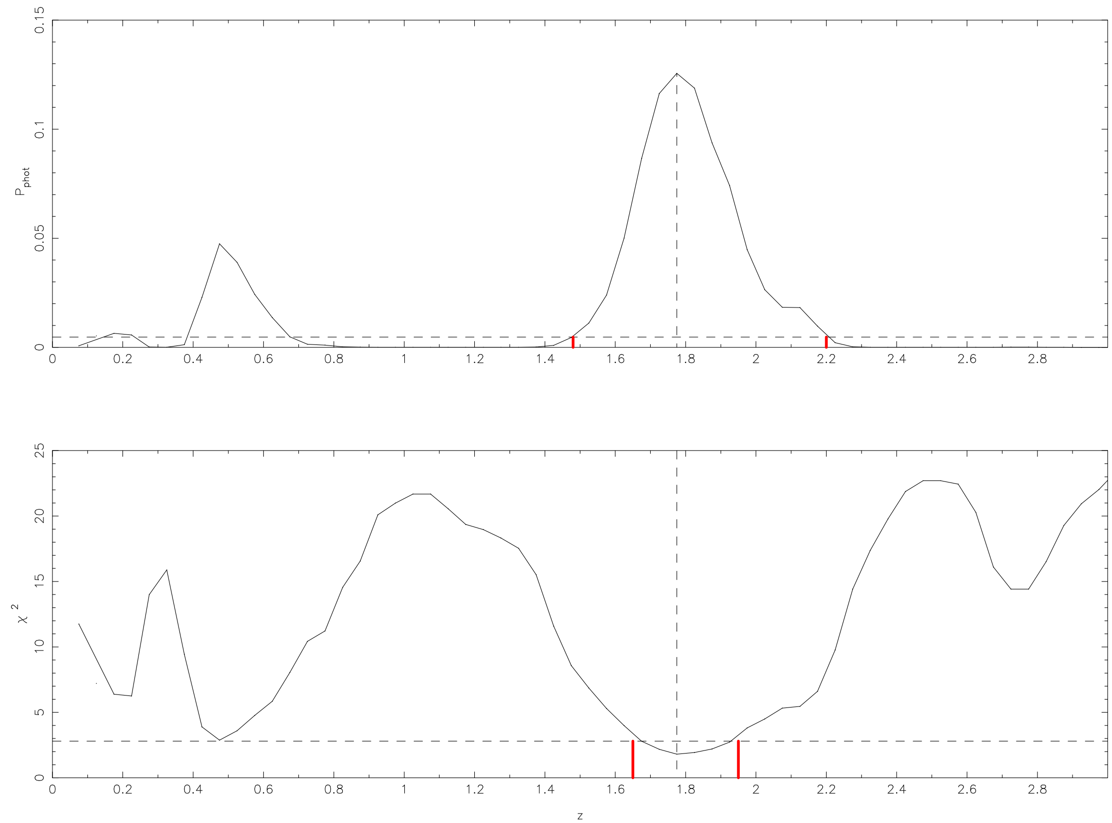


Figure 3.10: Demonstration of how upper and lower photometric redshift estimates are made with the modified code. The peak of each range above the threshold (horizontal dashed line in top plot) corresponds to a minimum χ^2 value in the bottom diagram (χ^2_{min}). The upper and lower redshift limits are the points at which $\chi^2 > \chi^2_{min} + 1$ (horizontal dashed line in bottom plot). The thick red lines in the upper plot demonstrate the upper and lower redshifts that would have been assigned with the original method and in the bottom plot the ones selected with the new method.

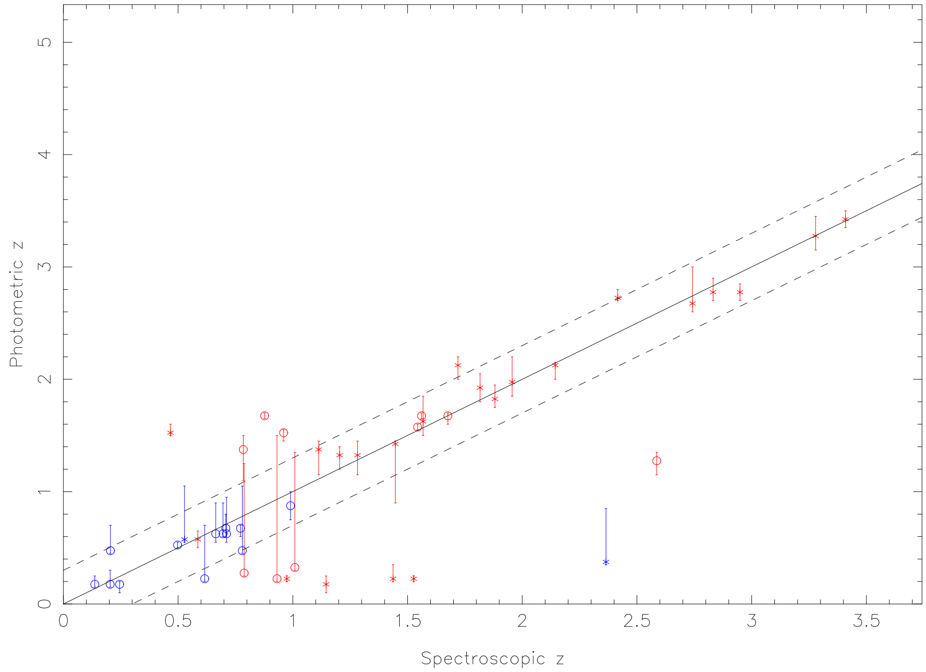
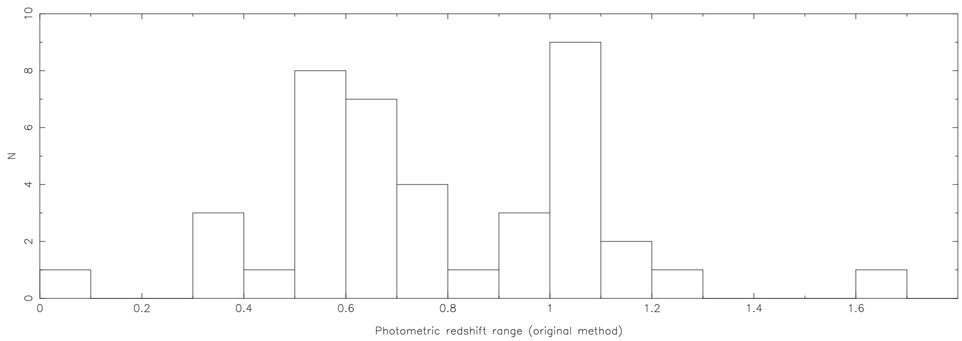


Figure 3.11: The spectroscopic and photometric redshift of each counterpart with $P_{id} \geq 0.9$. The photometric redshift is as assigned by the modified photometric redshift code using the designated catalogue for each source. Type 1 AGN sources are in red and type 2 in blue. Point counterparts are indicated by asterisks and open circles represent extended counterparts. The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

The assignment of type 1 and type 2 AGN used here was taken from Mateos *et al.* (2005) and was determined spectroscopically. This information may not always be available (e.g. for 2XMM) so the method must be able to rely on only the morphological information available. To test the effect this would have on the results the code was run again with point counterparts using the QSO catalogue as training set and extended counterparts with the galaxy catalogue. The results were encouraging with a drop of less than 3% in the number of sources assigned 'correct' photometric redshifts. Assuming



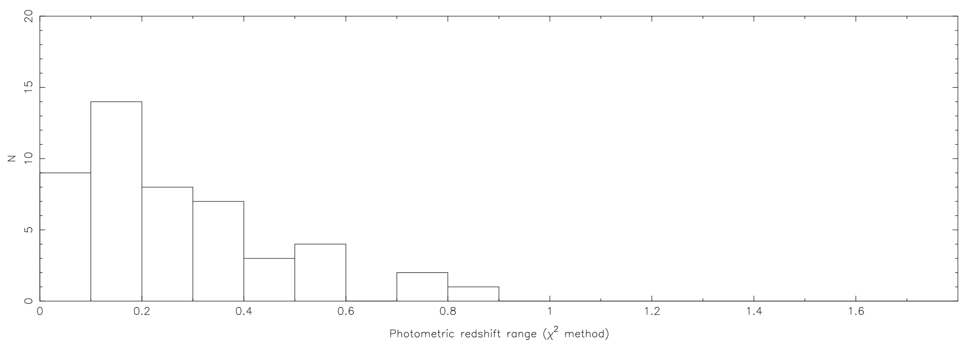


Figure 3.12: Photometric redshift range distributions assigned with the original method ($P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$, top figure) and with the χ^2 method ($P_{id} \geq 0.9$, bottom figure).

approximately 30% overlap between 2XMM and SDSS with an optical match rate of $\approx 60\%$, a sample of ≈ 20000 sources with accurate photometric redshifts ($|\Delta z| \leq 0.3$) should be possible for statistical studies such as luminosity evolution.

Table 3.2 shows details of objects in the Lockman Hole test sample including photometric redshift estimates obtained with a) the original version of the code with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$ and b) the modified version of the code with $P_{id} \geq 0.9$.

Table 3.2: Photometric redshift results for sources with $P_{id} \geq 0.9$

					Original method					χ^2 method				
ID	Morph.	AGN type	P$_{id}$	Spec. z	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
5	0	1	0.991624	2.144	-	-	-	-	-	2.0	2.125	2.15	0.019	0.15
6	0	2	0.987266	0.528	0.35	0.575	1.4	0.047	1.05	0.55	0.575	1.05	0.047	0.50
21	1	2	0.934349	0.498	0.2	0.525	0.95	0.027	0.75	0.5	0.525	0.55	0.027	0.05
39	1	2	0.978145	0.711	0.4	0.625	1.2	0.086	0.8	0.55	0.625	0.95	0.086	0.40
85	0	1	0.963114	1.145	-	-	-	-	-	0.1	0.175	0.25	0.970	0.15
88	0	1	0.95793	3.279	2.75	3.275	3.55	0.004	0.80	3.15	3.275	3.45	0.004	0.30
90	0	1	0.972944	0.467	-	-	-	-	-	1.5	1.525	1.6	1.058	0.10
96	0	1	0.98425	2.832	2.25	2.775	3.25	0.057	1.00	2.7	2.775	2.9	0.057	0.20
121	1	2	0.97616	0.772	0.35	0.675	1.05	0.097	0.70	0.6	0.675	0.7	0.097	0.10
124	1	1	0.905685	1.544	0.0	1.575	1.7	0.031	1.7	1.55	1.575	1.6	0.031	0.05

Table 3.2: continued

					Original method					χ^2 method				
ID	Morph.	AGN type	P$_{id}$	Spec. z	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
148	0	1	0.994837	1.113	1.0	1.375	1.5	0.262	0.50	1.15	1.375	1.45	0.262	0.30
156	0	2	0.943182	2.365	0.05	0.375	1.2	1.99	1.15	0.35	0.375	0.85	1.990	0.50
163	0	1	0.943911	0.974	0.1	0.225	0.45	0.749	0.35	0.2	0.225	0.25	0.749	0.05
168	0	1	0.990338	1.956	-	-	-	-	-	1.85	1.975	2.2	0.019	0.35
171	1	2	0.992806	0.205	0.0	0.475	1.0	0.270	1.00	0.45	0.475	0.7	0.270	0.25
176	0	1	0.982591	1.527	0.1	0.225	0.5	1.302	0.40	0.2	0.225	0.25	1.302	0.05
183	1	1	0.944796	0.96	0.8	1.525	1.7	0.565	0.9	1.45	1.525	1.55	0.565	0.10
191	1	1	0.985658	0.784	0.25	1.375	1.5	0.591	1.25	1.1	1.375	1.5	0.591	0.4
270	0	1	0.986931	1.568	1.45	1.625	2.05	0.057	0.60	1.5	1.625	1.85	0.057	0.35
272	1	2	0.974705	0.616	0.05	0.225	0.7	0.391	0.65	0.2	0.225	0.7	0.391	0.50

Table 3.2: continued

ID	Morph.	AGN type	P$_{id}$	Spec. z	Original method: Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range	χ^2 method: Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
277	0	1	0.98562	1.816	1.65	1.925	2.15	0.109	0.50	1.8	1.925	2.05	0.109	0.25
290	1	2	0.989787	0.204	0.0	0.175	0.55	0.029	0.55	0.15	0.175	0.3	0.029	0.15
300	1	1	0.967469	0.788	0.05	0.275	1.25	0.513	1.20	0.25	0.275	1.25	0.513	1.00
306	1	2	0.982591	0.708	0.4	0.675	1.05	0.033	0.65	0.6	0.675	0.8	0.033	0.20
321	1	1	0.93774	1.009	0.0	0.325	1.5	0.684	1.5	0.3	0.325	1.35	0.684	1.00
326	1	2	0.977426	0.78	0.15	0.475	1.35	0.305	1.2	0.45	0.475	1.05	0.305	0.60
332	1	1	0.966538	1.676	-	-	-	-	-	1.6	1.675	1.7	0.001	0.10
342	0	1	0.987135	0.586	-	-	-	-	-	0.5	0.575	0.65	0.011	0.15
354	0	1	0.989441	3.409	3.15	3.425	3.65	0.016	0.5	3.35	3.425	3.5	0.016	0.15
358	0	1	0.966227	2.742	2.05	2.675	3.3	0.067	1.25	2.6	2.675	3.0	0.067	0.40

Table 3.2: continued

					Original method					χ^2 method				
ID	Morph.	AGN type	P$_{id}$	Spec. z	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
364	1	1	0.919624	0.931	0.0	0.225	1.5	0.706	1.5	0.2	0.225	1.5	0.706	1.3
387	0	1	0.977426	1.447	0.65	1.425	2.3	0.022	1.65	0.9	1.425	1.45	0.022	0.55
394	1	2	0.985482	0.664	0.35	0.625	1.2	0.039	0.85	0.55	0.625	0.9	0.039	0.35
406	0	1	0.911221	1.282	0.9	1.325	1.55	0.043	0.65	1.15	1.325	1.45	0.043	0.30
407	1	2	0.930205	0.99	0.45	0.875	1.35	0.115	0.90	0.75	0.875	1.0	0.115	0.25
411	1	2	0.979863	0.245	0.0	0.175	0.55	0.070	0.55	0.1	0.175	0.2	0.070	0.10
427	1	2	0.917409	0.696	0.1	0.625	1.25	0.071	1.15	0.6	0.625	0.9	0.071	0.30
442	1	1	0.97616	2.586	0.25	1.275	1.5	1.311	1.25	1.15	1.275	1.35	1.311	0.2
450	0	1	0.988016	2.949	2.35	2.775	3.2	0.174	0.85	2.7	2.775	2.85	0.174	0.15
456	1	1	0.988835	0.877	-	-	-	-	-	1.65	1.675	1.7	0.798	0.05

Table 3.2: continued

					Original method					χ^2 method				
ID	Morph.	AGN type	P$_{id}$	Spec. z	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range	Phot. z low	Phot. z	Phot. z high	$\lvert\Delta z\rvert$	Phot. z range
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
475	0	1	0.986492	1.204	1.1	1.325	1.45	0.121	0.35	1.2	1.325	1.4	0.121	0.20
505	1	2	0.988016	0.137	0.0	0.175	0.5	0.038	0.50	0.15	0.175	0.25	0.038	0.10
527	0	1	0.92655	1.881	-	-	-	-	-	1.75	1.825	1.95	0.056	0.20
553	0	1	0.94921	1.437	0.15	0.225	0.5	1.212	0.35	0.2	0.225	0.35	1.212	0.15
599	0	1	0.950646	2.416	2.1	2.725	3.2	0.309	1.10	2.7	2.725	2.8	0.309	0.10
902	1	1	0.915341	1.561	-	-	-	-	-	1.65	1.675	1.7	0.114	0.05
2020	0	1	0.958999	1.72	-	-	-	-	-	2.0	2.125	2.2	0.405	0.20

NOTES: ID is the *XMM-Newton* identification number taken from Mateos *et al.* (2005). In column (2), 0 is a point counterpart and 1 is an extended counterpart, as classified by SDSS. AGN type is a property of the X-ray source as determined by Mateos *et al.* (2005). Spec. z is the spectroscopic redshift. Phot. z low is the lower limit on the photometric redshift assigned and phot. z high is the upper limit. Phot. z is the photometric redshift value. $|\Delta z|$ is the absolute difference between the spectroscopic and photometric redshifts. Phot. z range is the difference between the upper and lower photometric redshift values.

3.7 2XMM photometric redshift results

2XMM was released on August 22nd 2007 (see section 2.3.2 for more details). The code developed here was used to determine photometric redshifts for 2XMM point sources with SDSS optical counterparts. Optical counterparts were determined for X-ray sources in each X-ray field in the manner described in section 3.5. Of the 128023 best quality unique point sources in 2XMM (i.e. those with no warning flags), 33016 had potential counterparts in the SDSS catalogue and 21083 were assigned optical counterparts with $P_{id} \geq 0.9$. After removing potential stars (objects with log(F_x/F_{opt}) < -1), 18537 sources remained in the sample. These high P_{id} counterparts were split into point and extended samples (based on the SDSS classification) and the photometric redshift code was run with the QSO catalogue as training set for point counterparts and the galaxy catalogue for extended counterparts. Figures 3.13 and 3.14 show the results for the subset of sources with SDSS spectroscopic redshifts using the original code (with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$, 625 sources) and the modified code (with $P_{id} \geq 0.9$, 762 sources) respectively. The statistics from these plots are impressive (see table 3.3) with 93% and 87% of sources with $|\Delta z| \leq 0.3$ for original and modified methods respectively. The number of outliers in the samples (sources with $|\Delta z| > 1$) is just 4% (9%) for the original (modified) method. Figure 3.15 shows how accurate the majority of photometric redshift estimates are and also that the main 'problem area' in determining photometric redshifts is $0.5 \leq z \leq 2.5$ where there is a lack of features in the optical band of QSO spectra as discussed in section 3.1. A histogram of the photometric redshift distribution for the full sample of 18537 objects as determined by the modified

Table 3.3: Photometric redshift statistics for 2XMM

	Original method	χ^2 method
	$P_{id} \geq 0.9$	$P_{id} \geq 0.9$
	$P_{phot} \geq 0.8$	
(1)	(2)	(3)
% 'correct'	94.08	66.80
% with $\|\Delta z\| \leq 0.1$	74.23	70.34
% with $\|\Delta z\| \leq 0.2$	87.84	82.68
% with $\|\Delta z\| \leq 0.3$	93.12	87.40
$\overline{\|\Delta z\|}$	0.14	0.20
$\overline{z_{range}}$	0.54	0.21
% outliers	4.32	8.66

code is shown in figure 3.16.

The results surpass the accuracy expectations from the Lockman Hole test sample. A possible explanation for this is the relationship between $|\Delta z|$ and magnitude. Figure 3.17 and figure 3.18 show the fractional difference between photometric and spectroscopic redshift plotted against the SDSS r band magnitude for the Lockman Hole and 2XMM samples respectively using the modified version of the code. The fractional error clearly increases as the sources get fainter. The Lockman Hole sample is taken from deep observations of a region of space with very low Galactic absorption and should therefore be able to sample very faint objects. Conversely, 2XMM is a serendipitous,

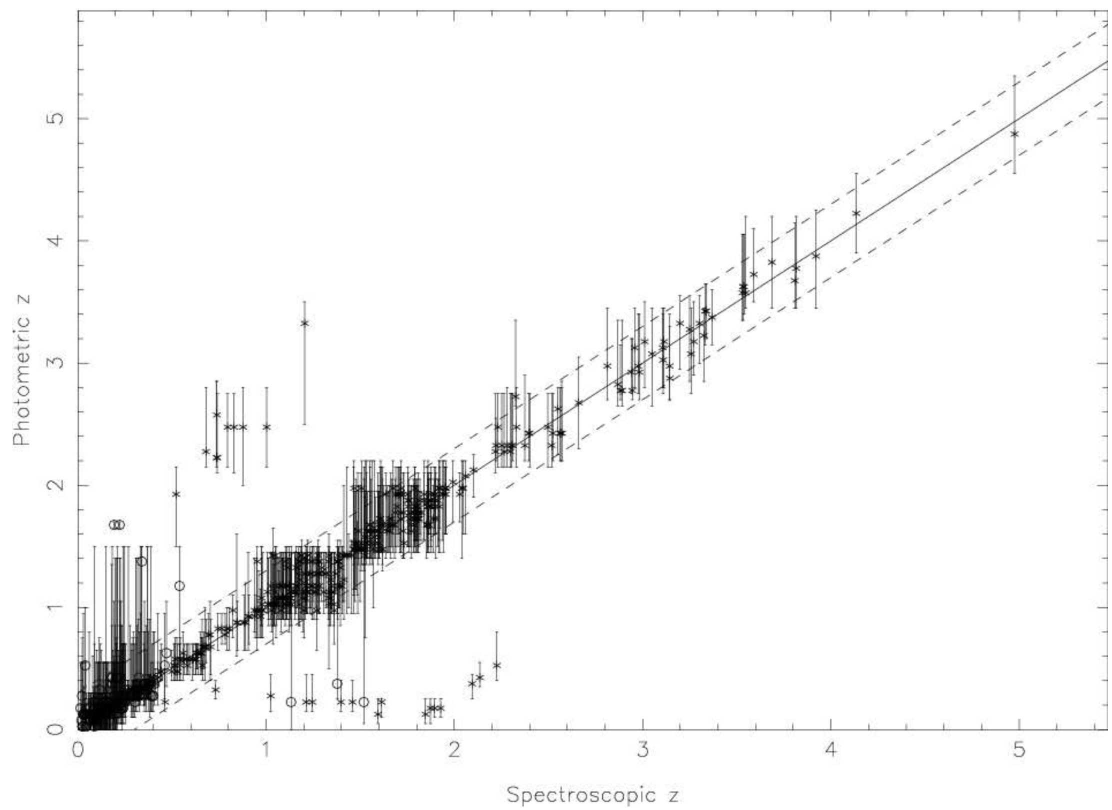


Figure 3.13: The spectroscopic and photometric redshift of each counterpart in the 2XMM sample with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$. The photometric redshift is as assigned by the photometric redshift code using the QSO catalogue for point counterparts (asterisks) and galaxy catalogue for extended counterparts (open circles). The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

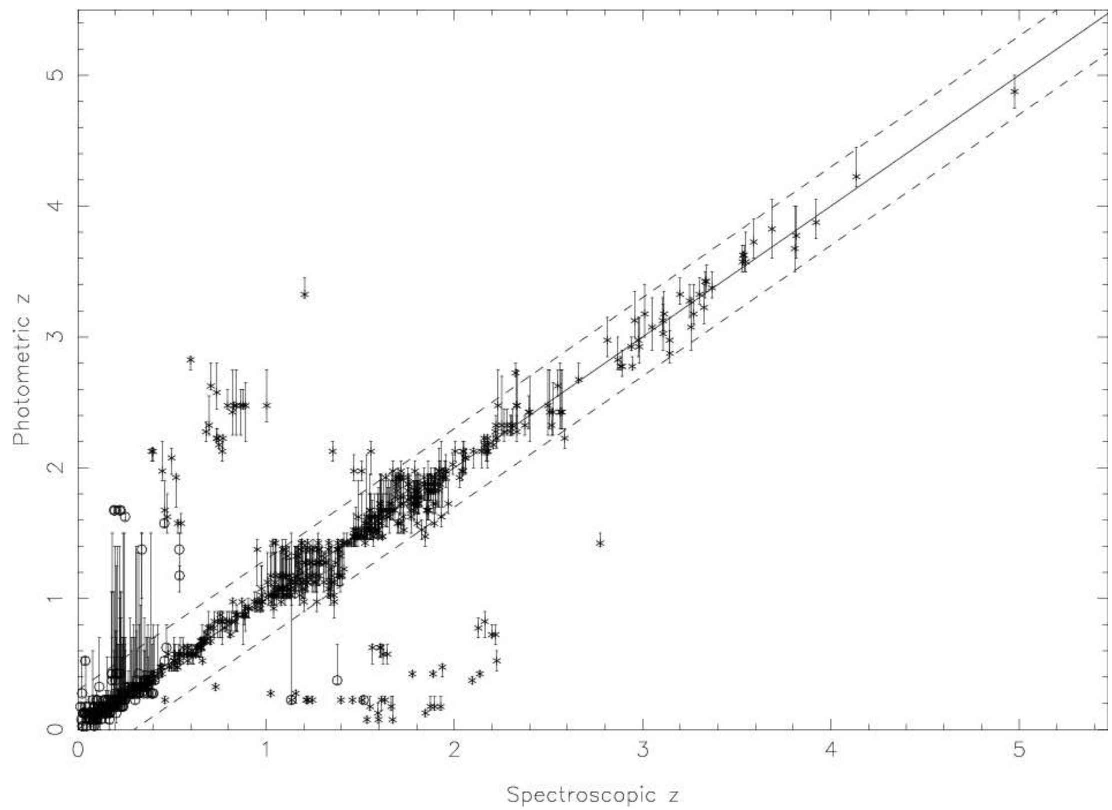


Figure 3.14: The spectroscopic and photometric redshift of each counterpart in the 2XMM sample with $P_{id} \geq 0.9$. The photometric redshift is as assigned by the modified photometric redshift code using the QSO catalogue for point counterparts (asterisks) and galaxy catalogue for extended counterparts (open circles). The solid line shows the path along which the points lie if the assigned photometric redshift matches the spectroscopic redshift. Dashed lines show $|\Delta z| = 0.3$.

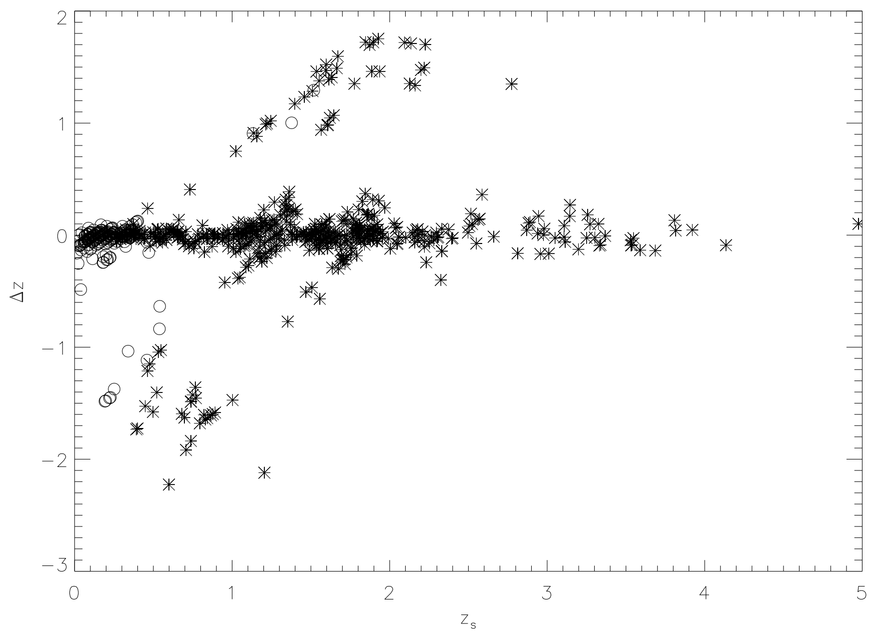


Figure 3.15: The difference between spectroscopic and photometric redshift (Δz) of each counterpart in the 2XMM sample with $P_{id} \geq 0.9$. The photometric redshift is as assigned by the modified photometric redshift code using the QSO catalogue for point counterparts (asterisks) and galaxy catalogue for extended counterparts (open circles).

wide area survey and thus contains mainly bright sources for which the accuracy of the code is better.

3.8 Results

Table 3.4 shows results of the two methods for the Lockman Hole and 2XMM samples, and also those of Kitsionas *et al.* (2005) as a comparison (as Kitsionas *et al.* 2005 used a similar method to the one employed here using X-ray selected objects from the

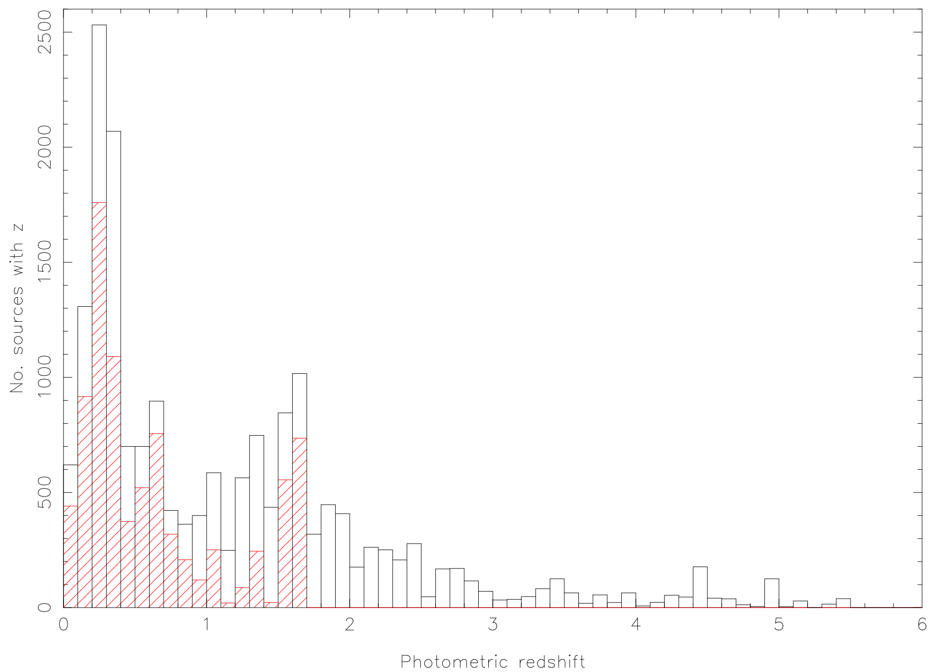


Figure 3.16: The photometric redshift distribution of counterparts with $P_{id} \geq 0.9$. The photometric redshift is as assigned by the modified photometric redshift code using the QSO catalogue for point counterparts and galaxy catalogue for extended counterparts. Red hatched regions show the extended counterpart distribution and unfilled regions the point counterpart distribution.

XMM/2dF survey). It is easily seen that both versions of the code are an improvement on the method used by Kitsionas *et al.* (2005), despite the fact that the Lockman Hole sample used here probes fainter objects, with a larger percentage of objects with $|\Delta z| \leq 0.3$ and also a smaller average difference between spectroscopic and photometric redshifts. The large improvement in average redshift range (the difference between the upper and lower limits) by using the modified version of the code instead of the original version is also obvious (see figure 3.12). Despite the drop in the number of sources classed as 'correct' when using the modified code, smaller error bounds are important if useful science is to be done with the resulting photometric redshifts.

Table 3.4: Photometric redshift statistics

	LH Orig $P_{id} \geq 0.9$ $P_{phot} \geq 0.8$	LH χ^2 $P_{id} \geq 0.9$	2XMM Orig $P_{id} \geq 0.9$ $P_{phot} \geq 0.8$	2XMM χ^2 $P_{id} \geq 0.9$	Kitsionas *et al.* (2005)	Gandhi *et al.* (2004)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
% 'correct'	86.49	53.19	94.08	66.80	-	-
% with $\lvert\Delta z\rvert \leq 0.1$	48.65	48.94	74.23	70.34	30.68	54.55
% with $\lvert\Delta z\rvert \leq 0.2$	59.46	59.57	87.84	82.68	53.41	63.64
% with $\lvert\Delta z\rvert \leq 0.3$	64.86	63.83	93.12	87.40	65.91	72.73
$\overline{\lvert\Delta z\rvert}$	0.34	0.34	0.14	0.21	0.46	0.26
$\overline{z_{range}}$	0.88	0.29	0.54	0.21	-	-
% of outliers	10.81	10.64	4.32	8.66	18.18	9.09
σ_{rms}	0.57	0.56	0.34	0.47	0.77	0.46

NOTES:

$$\sigma_{rms} = \sqrt{\frac{\sum(z_s - z_p)^2}{N}} \qquad (3.14)$$

where z_s is the spectroscopic redshift and z_p is the photometric redshift.

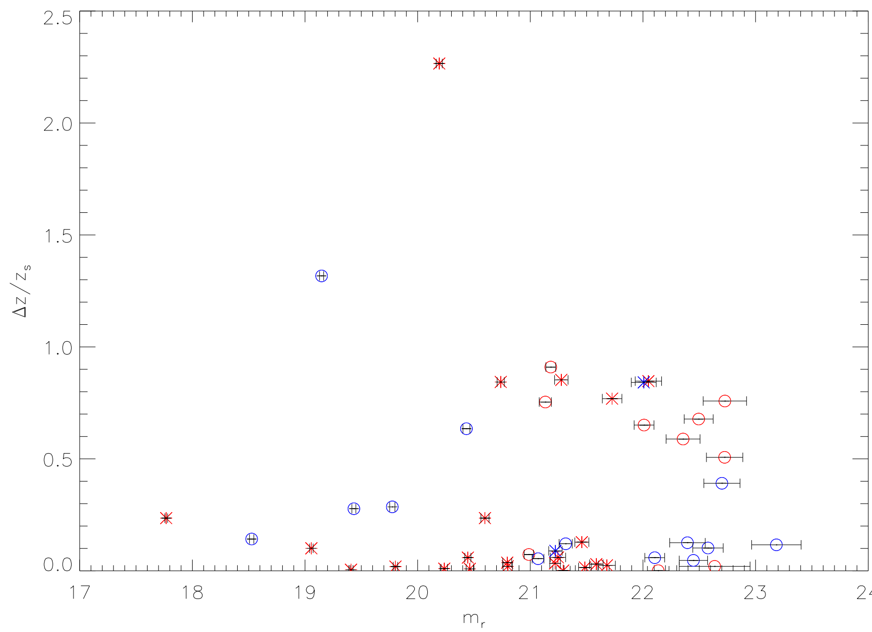


Figure 3.17: The fractional difference between spectroscopic and photometric redshift plotted against the SDSS r band magnitude of the Lockman Hole sample sources.

As of DR5, SDSS provides photometric redshifts for sources. Two different methods are used, both intended for use with galaxies: a template fitting method (photoz1, Csabai *et al.* 2003) and one which uses a neural network with a training set (photoz2, Oyaizu *et al.* 2008). Results found here for extended counterparts and type 2 AGN in the Lockman Hole sample (using the galaxy catalogue as training set) are comparable to the estimates available from SDSS with the obvious advantage that estimates for point counterparts with the QSO catalogue can also be calculated here.

Weinstein *et al.* (2004), the developers of the original code used here, found 83% of sources in their sample had photometric redshift estimates with $|\Delta z| \leq 0.3$ when they applied their code to the SDSS EDR QSO catalogue (using the catalogue as both training

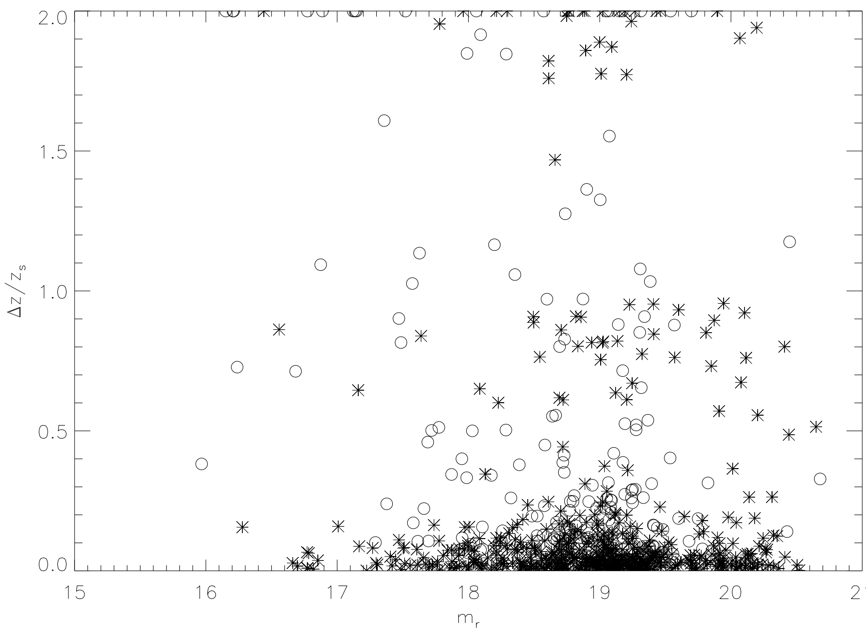


Figure 3.18: The fractional difference between spectroscopic and photometric redshift plotted against the SDSS r band magnitude of the 2XMM sample sources. The small percentage of very large outliers ($|\Delta z|/z_s \geq 2.0$) are all plotted at $|\Delta z|/z_s = 2.0$ so the distribution of the majority of sources can be seen clearly.

set and test set). The accuracy found with the original code used here with a combination of galaxy and QSO training catalogues with X-ray selected sources ($P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$) is lower than this (65% with $|\Delta z| \leq 0.3$) but the overall percentage of 'correct' sources (i.e. sources with spectroscopic redshift within the upper and lower ranges of photometric redshift) is comparable - 86% found here and 83% for Weinstein *et al.* (2004). The results of the 2XMM sample are a great improvement on the Weinstein *et al.* (2004) result - 93% with $|\Delta z| \leq 0.3$ and 94% 'correct' (using the original method and sources with $P_{id} \geq 0.9$ and $P_{phot} \geq 0.8$).

A fairer comparison to make is between the results obtained here and those of other photometric redshift studies which used X-ray selected sources. The method used here is superior to studies by Gandhi *et al.* (2004) (see table 3.4) and Gonzalez and Maccarone (2002) who both found difficulties in accurately estimating photometric redshifts for X-ray selected sources dominated by light from the central AGN. Kitsionas *et al.* (2005) found 68% of sources fit with QSO templates in their sample have $|\Delta z| \leq 0.3$ comparable to 73% found here using the original method and probability cuts and 68% with the χ^2 method. For extended sources they find that for blue objects ($g - r < 0.5$), the templates do not provide accurate photometric redshifts with only 27% of sources with $\Delta z \leq 0.3$. In comparison 33% of blue 'galaxy' sources here satisfy this criteria with the χ^2 method. Table 3.4 shows that, over the full sample, the results found here are superior to the results found by Kitsionas *et al.* (2005) in terms of the percentage of sources with $|\Delta z| \leq 0.2$ and with lower $\overline{\Delta z}$ using our code.

Many papers use the statistic σ_{rms} as a measure of the accuracy of the method. Kitsionas *et al.* (2005) found $\sigma_{rms} = 0.77$ and Gandhi *et al.* (2004) found $\sigma_{rms} = 0.46$ with their methods. The method used here compares favourably with these results with $\sigma_{rms} = 0.47$ for the modified version of the code with the 2XMM sample. The photometric redshift method developed here by modifying the code of Weinstein *et al.* (2004) performs well in comparison to other studies of X-ray selected sources and with considerably smaller error ranges than the original code. Even with the faint sample of Lockman Hole sources the results found are an improvement on the most similar study of X-ray selected AGN (Kitsionas *et al.* 2005). The brighter sample taken from 2XMM demonstrates the high level of accuracy possible with the code.

3.9 Conclusions

A photometric redshift code was used to determine redshift estimates of X-ray selected sources from *XMM-Newton* with optical counterparts taken from SDSS. Quasar and galaxy catalogues were created from SDSS sources and used to determine the colour redshift relation for point and extended counterparts respectively. Tests of this method on sources in the deep Lockman Hole field yielded accuracy of 65% of sources with $|\Delta z| \leq 0.3$ with appropriate cuts on P_{id} and P_{phot}.

Rather than using the arbitrary threshold in probability to determine the upper and lower limits of the photometric redshift values, the code was modified to use the χ^2 values to allocate standard 1σ error bars to each estimate. This altered the average range of the photometric redshift $(\overline{z_{high} - z_{low}})$ from 0.88 to 0.29 while leaving the percentage of sources with $|\Delta z| \leq 0.3$ etc. approximately the same. The use of 1σ errors is useful when using the photometric redshifts to calculate other quantities such as luminosity (as will be done in Chapter 4).

From table 3.4 it is obvious that the methods used here for photometric redshift are equal to and in many ways better than other photometric redshift methods that have previously been used for samples of X-ray selected AGN. The σ_{rms} value for the modified method presented here is 0.47 which is a strong level of accuracy for a photometric redshift method used with X-ray selected sources.

Use of the code with the 2XMM catalogue showed that even greater accuracy is possible when used with large bright samples. 93% and 87% of sources were found to have $|\Delta z| \leq 0.3$ with the original and modified methods respectively for a sample of around 700 sources with spectroscopic redshifts. A sample of around 15000 sources can currently be obtained with 2XMM and SDSS coverage providing a useful resource for survey science.

As the UKIDSS infra-red database increases in size, IR data could be incorporated into this method to further improve results either by using $z - J$ colour as a discriminator and using a 'red' catalogue as proposed in section 3.6 or by incorporating J band magnitudes into the code itself.

Chapter 4

Variability of serendipitous source light curves in the 3C 273 field

4.1 Introduction

Variability across all bands of the electromagnetic spectrum is a fundamental characteristic of AGN. The most rapid variability is observed in the X-ray band. Rapid variability implies that the source of X-ray emission is in the innermost region of the central engine. Investigating X-ray variability can therefore reveal information about the properties of the central engine.

X-ray observations of AGN in the 1980s established that variability in the X-ray band is a common feature of AGN (e.g. Marshall *et al.* 1981). Studies using simple tools such as the flux doubling timescale (the time for the emitted flux to change by a factor of two), e.g. Barr and Mushotzky (1986), found correlations between the variability timescale and X-ray luminosity.

Later studies of X-ray variability in AGN used the power spectral density function (power spectrum). The power spectrum is a method of identifying characteristic timescales but, until very recently, no legitimate periodicities had been discovered in AGN (a significant Quasi Periodic Oscillation (QPO) has recently been detected in RE J1034+396 by Gierliński *et al.* 2008). However, an inverse correlation between normalised variability amplitude (the square root of the power at a specific frequency, normalised to the mean count rate of the light curve) and X-ray luminosity was detected (Green *et al.* 1993) from the power spectra analysis. Lawrence and Papadakis (1993) also found an anti-correlation between power spectrum amplitude and X-ray luminosity but with large scatter around the fit. Green *et al.* (1993) and Lawrence and Papadakis (1993) used *EXOSAT* observations for their studies of AGN light curves. *EXOSAT* was a European X-ray satellite which flew from May 1983 to April 1986. It had a highly eccentric orbit which allowed for up to 76 hours of uninterrupted observation time with high signal to noise. This characteristic made its observations ideally suited to use for power spectrum analysis of AGN. For good quality power spectra, high quality data with near even sampling is required. This is very costly to achieve and is near impossible over long periods with many of the currently available telescopes because of short orbits and unavoidable gaps e.g. the South Atlantic Anomaly. Since the demise of *EXOSAT*, X-ray light curves of AGN have generally been unevenly sampled (although in recent years, observations from *RXTE* and *XMM-Newton* have been used for power spectral analysis of some sources, e.g. Vaughan *et al.* 2003b) and so a different method of quantifying variability is therefore required.

Nandra *et al.* (1997) used normalised excess variance of X-ray light curves as a measure of variability. They compiled a sample of 18 local ($z < 0.05$) Seyfert 1 galaxies observed by *ASCA* with variability on timescales of minutes to hours. The normalised excess variance was calculated for each $128\,\mathrm{s}$ bin light curve:

$$\sigma^2_{rms} = \frac{1}{N\mu^2}\sum_{i=1}^{N}[(R_i - \mu)^2 - \sigma_i^2] \tag{4.1}$$

where N is the number of bins making up the light curve, R_i is the count rate of the i^{th} bin, μ is the unweighted mean count rate and σ_i is the error in the count rate of the i^{th} bin. Nandra *et al.* (1997) used this method to confirm the anti-correlation between variability and luminosity previously detected with power spectrum analysis. With larger bins ($5760\,\mathrm{s}$) an even stronger correlation was detected. Nandra *et al.* (1997) characterised the anti-correlation by the equation:

$$\sigma^2_{rms} \propto L_x^{-0.71\pm0.03} \tag{4.2}$$

Turner *et al.* (1999) expanded on the work of Nandra *et al.* (1997) by increasing the sample to include narrow line Seyfert galaxies (NLS1s, which they define as having FWHM $H\beta < 2000\,\mathrm{km\,s^{-1}}$). They found that the anti-correlation between variability and X-ray luminosity remained but the scatter around the fit increased due to the different types of galaxy in the sample.

In 2000, Almaini *et al.* (2000) used a different method of variability analysis, using a maximum likelihood technique, with a higher redshift sample ($0.1 < z < 3.2$) of 86 QSOs from a deep *ROSAT* survey. These sources typically had very low signal to noise ratios so Almaini *et al.* (2000) used ensemble light curves. By treating each light curve as a snapshot of the QSO population and assuming that all light curves have the same underlying distribution, light curves from multiple sources were combined (after dividing each light curve by its mean flux) and the variance in the resulting distribution was calculated. Taken individually, there appeared to be no correlation between luminosity and variance but when the sources were split into six luminosity bins, the ensemble values of variance showed the anti-correlation with luminosity detected in previous studies. A possible upturn to a positive correlation at high luminosity ($\approx 5 \times 10^{44}\,\mathrm{erg\,s^{-1}}$) was also detected. Splitting the sample into low ($z < 0.5$) and high ($z > 0.5$) redshift sources showed the standard anti-correlation between luminosity and variance for low redshift ensemble sources but the high redshift objects showed a potential positive correlation.

Lu and Yu (2001) used a sample comprising 22 Seyfert 1 galaxies and QSOs and found an anti-correlation between normalised excess variance and central black hole mass. They suggested that the trend between luminosity and variability previously detected may actually be driven by the black hole mass. This trend was confirmed in 2003 by Bian and Zhao (2003) using a sample of 41 broad and narrow line AGN. As was found by Turner *et al.* (1999), the inclusion of NLS1s in the sample caused a larger scatter in the correlation compared to a broad line only sample.

The anti-correlation between normalised excess variance and X-ray luminosity and in turn between variance and black hole mass is therefore well established in clearly defined samples. Here, a serendipitous sample of AGN detected by *XMM-Newton* is used to investigate whether the trend is still detected in lower quality data over a large redshift range.

4.2 Method

4.2.1 Data

3C 273 is a calibration target for *XMM-Newton* and has therefore been observed over twenty times since the launch of the satellite. The wide field of *XMM-Newton* means that many faint objects in the field of view around 3C 273 have therefore also been observed. By combining the multiple observations, an effective exposure time of around $300\,\mathrm{ks}$ can be built up giving a deep look at the field. The observations used for this study are given in table 4.1

SAS version 6.5.0 was used with automated scripts written by Mr Mark Simpson to reduce the data from the ODF files for each of the MOS cameras in the following way. $12-15\,\mathrm{keV}$ light curves were generated to identify and filter out flaring events (periods when the count rate exceeded around 0.3 counts per second). The good time intervals (GTIs) determined from the flaring analysis were used to filter the event lists. Events from the central CCD (i.e. 3C 273) were also filtered from the event lists as they dom-

Table 4.1: Observations

Observation ID	Start date	Duration (s)	Revolution
(1)	(2)	(3)	(4)
0126700201	13/06/00	26395	0094
0126700301	13/06/00	73901	0094
0126700601	15/06/00	31471	0095
0126700701	15/06/00	36728	0095
0126700501	16/06/00	32471	0095
0126700801	17/06/00	73901	0096
0136550101	13/06/01	90958	0277
0112770101	16/12/01	6808	0370
0112770201	22/12/01	6771	0373
0112770601	07/07/02	6694	0472
0112770801	17/12/02	6679	0554
0112770701	05/01/03	6679	0563
0136550501	05/01/03	10000	0563
0112771001	18/06/03	7000	0645
0112770501	08/07/03	9600	0655
0112771101	14/12/03	10000	0735
0136550801	30/06/04	64000	0835

inated the field due to the source's brightness. The resulting event lists were used to create images and exposure maps in various bands for each observation. All images and exposure maps in each band were then stacked using `emosaic`.

Source detection was performed using the images and exposure maps with a detection mask created from the broad band ($0.2-12$ keV) exposure map. `Eboxdetect` created an initial source list which was used by `esplinemap` to create background maps. The background maps were then used with a second run of `eboxdetect` to get a refined source list. Finally, `emldetect` was used to narrow the source list down to just those sources with minimum likelihood of detection (det_ml) greater than twelve. The 2XMM catalogue uses det_ml greater than six for its source detection so the large value used here ensures very few spurious sources are included. Visual inspection of these sources was performed to ensure a high level of quality, leaving 116 sources in the final list. Figure 4.1 shows the stacked full band ($0.2 - 12$ keV) image of the field. Sources with optical counterparts (see section 4.2.2) are circled and labelled.

Although exposures were taken with the pn camera for each of the observations listed in table 4.1, they were not used here because, in the majority of cases, they were taken in small window mode and so did not cover enough of the field of view. In the case of the MOS cameras, the outer ring of CCDs is always in standard imaging mode and it is only the central CCD which operates in different modes. As the central CCD was not included in the analysis performed here, all observations using MOS cameras could be used.

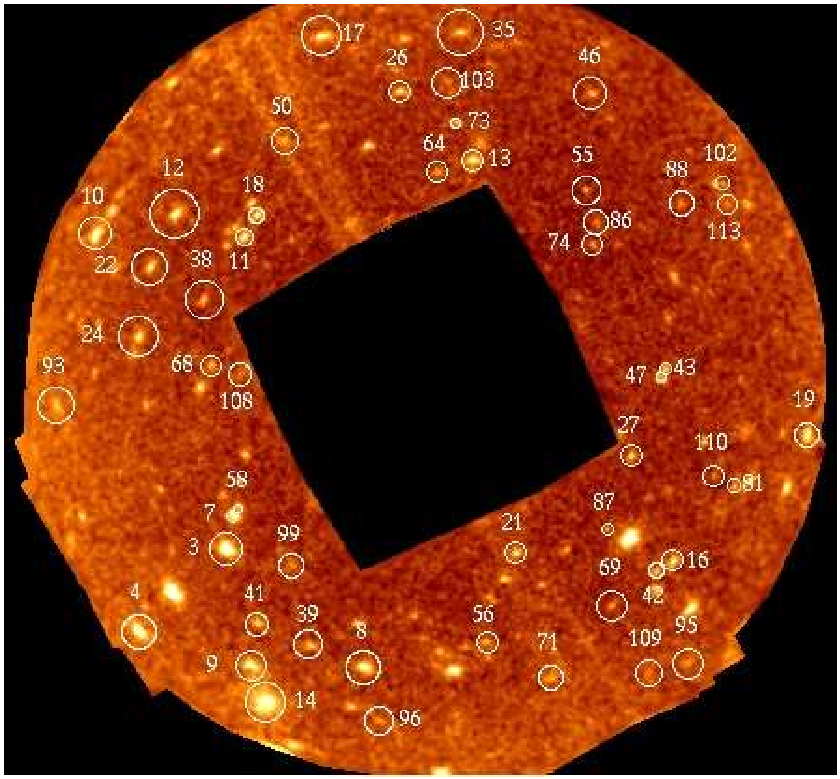


Figure 4.1: Stacked full band image of the 3C 273 field. Only sources with optical counterparts with $P_{id} \geq 0.9$ are identified.

4.2.2 Cross correlation

Previous studies (e.g. Almaini *et al.* 2000, Nandra *et al.* 1997) have detected an anti-correlation between excess variance and X-ray luminosity in AGN. To test whether the sources in the 3C 273 field follow that pattern, redshift is needed in order to convert flux to luminosity. Spectroscopic redshifts were unavailable for the majority of the detected sources so photometric redshifts were determined using the method developed in Chapter 3. To do this, optical data for each source was required and so cross correlation with SDSS DR6 was performed (as outlined in Chapter 3, section 3.5). 90 of the 116

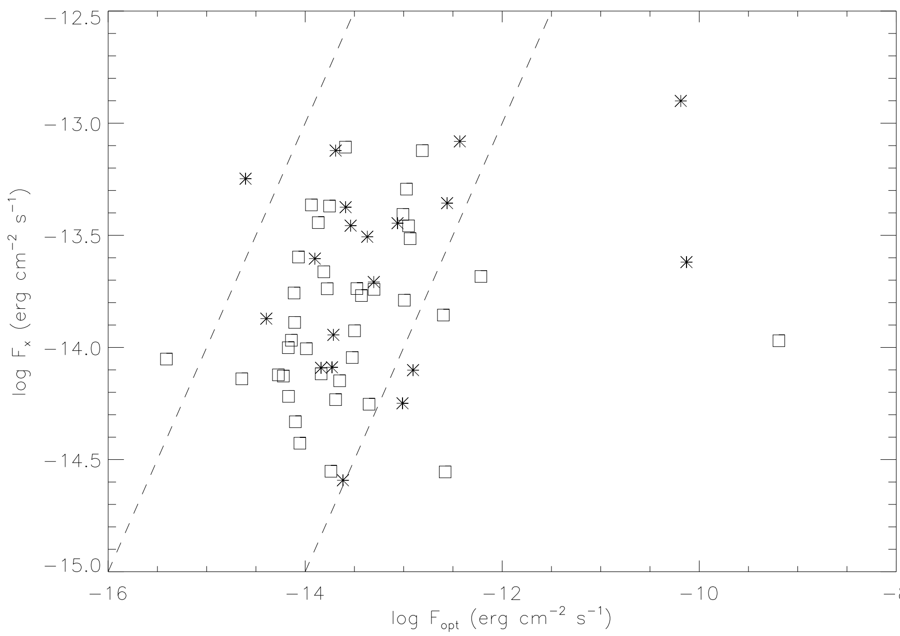


Figure 4.2: Optical and X-ray flux of the 3C 273 sources with optical counterparts ($P_{id} \geq 0.9$). Squares represent extended optical counterparts (classified as galaxies by SDSS) and asterisks represent point optical counterparts. The dashed line on the left is at $\log(F_x/F_{opt}) = 1$ and the line on the right is at $\log(F_x/F_{opt}) = -1$.

detected sources were determined to have optical counterpart candidates of which 58 were assigned optical counterparts with $P_{id} \geq 0.9$ (see section 3.5 for the definition of P_{id}).

An F_x/F_{opt} plot was made to determine possible classifications of the 58 sources with $P_{id} \geq 0.9$ optical counterparts (see figure 4.2). Eight sources with $\log(F_x/F_{opt}) < -1$ were removed from the sample as they are potential stars. All other sources were assumed to be AGN as previous studies have shown the majority of X-ray sources at high Galactic latitudes ($|b| \geq$ 20deg) are active galaxies (e.g. Barcons *et al.* 2007).

4.2.3 Photometric redshifts and luminosity

The SDSS 'ugriz' magnitudes of the optical counterparts for X-ray sources with $P_{id} \geq 0.9$ were used to calculate photometric redshifts using the method described in Chapter 3. The modified version of the photometric redshift code was used because of its smaller error range compared to the original version (Weinstein *et al.* 2004). 'Extended' counterparts (those classed as galaxies by SDSS) were fitted with the galaxy catalogue as a training set and point counterparts were fitted with the quasar catalogue as a training set.

The full band ($0.2 - 12\,\mathrm{keV}$) X-ray luminosity was calculated for each source using the X-ray flux (determined by SAS source detection tasks) and the assigned photometric redshift. The angular diameter distance, D_A, was calculated:

$$a(z) = \frac{1}{1+z} \tag{4.3}$$

$$Z = \int_{a(z)}^{1} \frac{da}{a\sqrt{X}} \tag{4.4}$$

$$\sqrt{X} = \sqrt{\Omega_k + \Omega_\Lambda a^2 + \frac{\Omega_M}{a} + \frac{\Omega_r}{a^2}} \tag{4.5}$$

$$D_A = \frac{cZa(z)}{H_0} \tag{4.6}$$

where z is the photometric redshift, H_0 is the Hubble constant ($70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) and c is the speed of light ($3 \times 10^8\,\mathrm{m\,s^{-1}}$). The Ω values are different components of the density parameter: Ω_k is the curvature density (0.0), Ω_Λ is the vacuum density (0.7), Ω_M is the matter density (0.3) and Ω_r is the radiation density (0.0) (values in line with the *WMAP* concordance cosmology, Spergel *et al.* 2003).

This was converted to luminosity distance, D_L:

$$D_L = (1+z)^2 D_A \tag{4.7}$$

The X-ray luminosity, L_x, is then:

$$L_x = 4\pi F_x D_L^2 \tag{4.8}$$

where F_x is the $0.2 - 12\,\mathrm{keV}$ X-ray flux.

4.2.4 Light curve extraction

The tool `evselect` from SAS version 6.5.0 was used to extract $0.2 - 12\,\mathrm{keV}$ light curves with $100\,\mathrm{s}$ bins. The extraction region for each source was a circle with radius initially set to enclose 90% of the source energy. The radii of overlapping source regions were then shrunk until they no longer overlapped. The background region for each source was an annulus with inner radius of $1'$ and outer radius of $3'$. Any area within

$1'$ of a neighbouring source that overlapped the background annulus was omitted from the background region. Source and background light curves were forced to have the same start and stop times for ease of background subtraction. Light curves were kindly supplied by Mr Mark Simpson.

The 100 s light curves were then rebinned to 1000 s bins by the following method. Each 100 s bin was flagged as 'good' if 80% of it is within a GTI. This step was necessary as `evselect` does not carry out exposure correction of time bins so some bins may only partially overlap with the GTI. Sets of ten consecutive 100 s bins were taken for each 1000 s bin. If six or more of the bins were flagged as 'good' then the 1000 s bin was generated by summing the counts in the 'good' 100 s bins to get the total count (C_T). If less than six bins were 'good' the set was discarded and the next ten bins were taken starting from the bin after the first 'bad' bin in the discarded set. The time assigned to each 1000 s bin is the midpoint between the first and last component 100 s bins. This binning was also performed for background light curves. The number of 100 s bins used to make up each 1000 s bin was noted.

At this point source 4 and source 95 were removed from the sample due to lack of useable light curve data leaving 48 sources in the sample.

For each observation of each source an average background was calculated:

$$\overline{C_{B100}} = \frac{\sum \frac{C_{B1000}}{N_{100}}}{N_{1000}} \tag{4.9}$$

where $\overline{C_{B100}}$ is the average background count in a $100\,\mathrm{s}$ bin, C_{B1000} is the background count in a $1000\,\mathrm{s}$ bin, N_{100} is the number of $100\,\mathrm{s}$ bins making up the $1000\,\mathrm{s}$ bin and N_{1000} is the total number of $1000\,\mathrm{s}$ bins. The average background in a $100\,\mathrm{s}$ bin is required because not all $1000\,\mathrm{s}$ bins are made up of ten $100\,\mathrm{s}$ bins. Background subtracted $1000\,\mathrm{s}$ bin source counts were then calculated:

$$C_S = C_T - \overline{C_{B100}} N_{100} \frac{A_S}{A_B} \tag{4.10}$$

where C_S is the background subtracted source count, C_T is the total count, $\overline{C_{B100}} N_{100}$ is the average background count corrected for the true length of the $1000\,\mathrm{s}$ bin, A_S is the source area and A_B is the background area. Error in the source count was also calculated:

$$\Delta C_S = \sqrt{C_T + \overline{C_{B100}} N_{100} \left(\frac{A_S}{A_B}\right)^2} \tag{4.11}$$

except in cases where $C_T = 0$ in which case:

$$\Delta C_S = \sqrt{1 + \overline{C_{B100}} N_{100} \left(\frac{A_S}{A_B}\right)^2} \tag{4.12}$$

Count rate was then calculated for each $1000\,\mathrm{s}$ bin:

$$R = \frac{C_S}{100 N_{100}} \tag{4.13}$$

The error in the rate is:

$$\sigma = \frac{\Delta C_S}{100 N_{100}} \tag{4.14}$$

All rates and times for each source from all observations were combined into one list so parameters for each source could be calculated taking all observations as one light curve. Light curves for MOS1 and MOS2 cameras were analysed separately.

Due to the low number of counts for many sources, the number of counts in each $1000\,\mathrm{s}$ bin was not large enough for methods using Gaussian statistics to be appropriate (i.e. fewer than around 20 counts). The $1000\,\mathrm{s}$ light curves were therefore rebinned to $15000\,\mathrm{s}$ bins which, for most sources, provided enough bins and counts per bin (before background subtraction) for Gaussian statistics to be appropriate. This was performed in a similar way to the $1000\,\mathrm{s}$ binning. Starting from the first $1000\,\mathrm{s}$ bin the start and stop time for the larger bin was set:

$$start = time(1000\,\mathrm{s}\; bin) - 500\,\mathrm{s} \tag{4.15}$$

$$stop = start + 15000\,\mathrm{s} \tag{4.16}$$

Any $1000\,\mathrm{s}$ bins within the start and stop times were used to create the larger time bin. The number of $100\,\mathrm{s}$ bins making up each larger bin was noted using the information from the $1000\,\mathrm{s}$ binning. More than ninety $100\,\mathrm{s}$ bins were required to make up each $15000\,\mathrm{s}$ bin. If this threshold was not met, the bin was discarded. The time associated with the larger bins is the midpoint between the first and last 1000 s bin included.

Parameters of sources for which 15000 s light curves were created are given in table 4.2.

Having extracted M1 and M2 light curves separately, it was decided that the light curves from the two cameras should be combined to create a single light curve for each source to improve the signal to noise (S/N) ratio. For the 1000 s light curves this was done by taking the 1000 s M1 and M2 light curves and adding the rates for bins with the same time (to within 350 s either way). To account for potential differences between cameras in the number of 100 s bins making up each 1000 s bin, all 1000 s bin counts were interpolated up to ten bins before addition of rates.

The combined camera (M1+M2) 1000 s light curve was then binned up to 15000 s bins in the same way as that of the separate cameras. Combined light curves were only created for those sources where Gaussian statistics were appropriate for both cameras in the separate camera case.

The 1000 s and 15000 s light curves (for M1 and M2 cameras) of source 3 and source 24 are shown in figure 4.3 and figure 4.5 respectively as an example of variable (source 3) and non-variable (source 24) light curves. The combined M1+M2 light curves (1000 s and 15000 s) of source 3 and source 24 are shown in figure 4.4 and figure 4.6 respectively. Source 3 is statistically variable in all cases and source 24 is statistically non-variable in all cases. The x axis in each case is artificial in order to show the light curves of all observations on one plot. The dotted line in each plot shows the mean and vertical dashed lines separate the observations. Errors are not shown in the 1000 s bin plots for

Table 4.2: Source parameters

Source	L_x	z_{phot}	Morph.	Mean rate	Total counts	Total counts
	$(\mathrm{erg\,cm^{-2}\,s^{-1}})$			(10^{-3} counts s^{-1})	M1	M2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
3	$2.3\pm0.7\times10^{45}$	$1.975\pm^{0.075}_{0.125}$	0	11.4	1687	1627
8	$4.1\pm^{4.1}_{2.9}\times10^{44}$	$1.375\pm^{0.075}_{0.275}$	0	3.8	614	565
9	$1.7\pm0.7\times10^{45}$	$2.225\pm^{0.025}_{0.075}$	0	2.8	568	309
10	$7.6\pm^{4.9}_{3.3}\times10^{42}$	0.225 ± 0.025	1	2.9	604	359
11	$5.9\pm^{17}_{4.2}\times10^{42}$	$0.275\pm^{0.125}_{0.075}$	0	2.2	351	318
12	$7.2\pm^{4.7}_{3.4}\times10^{44}$	1.825 ± 0.075	0	3.5	560	434
13	$1.2\pm^{2.2}_{0.9}\times10^{43}$	0.325 ± 0.075	0	2.7	364	475
14	$1.8\pm0.3\times10^{46}$	5.375 ± 0.025	0	7.9	1456	1024
16	$5.9\pm^{205}_{4.2}\times10^{42}$	$0.225\pm^{0.475}_{0.025}$	1	1.9	259	270
17	$6.3\pm^{26}_{4.5}\times10^{44}$	$1.525\pm^{0.575}_{0.125}$	0	1.9	293	239
19	$1.5\pm0.5\times10^{45}$	1.675 ± 0.025	1	1.1	321	-

Table 4.2: continued

Source	L_x	z_{phot}	Morph.	Mean rate	Total counts	Total counts
	($\mathrm{erg\,cm^{-2}\,s^{-1}}$)			(10^{-3} counts $\mathrm{s^{-1}}$)	M1	M2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
21	$9.5\pm^{10}_{6.0}\times10^{44}$	$2.475\pm^{0.175}_{0.275}$	0	2.3	345	304
22	$2.4\pm^{9.3}_{2.0}\times10^{44}$	$1.025\pm^{0.325}_{0.275}$	1	1.2	224	147
24	$3.3\pm^{2.8}_{2.4}\times10^{44}$	1.675 ± 0.025	1	1.8	325	173
35	$2.0\pm^{3.5}_{1.9}\times10^{44}$	$1.975\pm^{0.075}_{0.325}$	0	0.6	71	162
38	$3.4\pm^{11}_{3.2}\times10^{44}$	$2.325\pm^{0.325}_{0.425}$	0	0.6	177	-
39	$2.6\pm^{8.2}_{2.3}\times10^{42}$	0.175 ± 0.075	1	0.3	104	-
46	$1.8\pm^{315}_{1.7}\times10^{42}$	$0.225\pm^{0.775}_{0.025}$	1	0.4	112	30
50	$4.8\pm^{600}_{4.6}\times10^{42}$	$0.375\pm^{0.975}_{0.025}$	1	1.4	165	224
55	$5.0\pm^{4000}_{5.0}\times10^{41}$	$0.175\pm^{0.875}_{0.175}$	1	0.5	81	67
56	$2.7\pm^{5.1}_{2.5}\times10^{44}$	$1.525\pm^{0.025}_{0.525}$	1	0.9	-	143
64	$1.0\pm^{6.9}_{1.2}\times10^{43}$	$0.625\pm^{0.275}_{0.025}$	1	0.9	148	127

Table 4.2: continued

Source	L_x	z_{phot}	Morph.	Mean rate	Total counts	Total counts
	$(\mathrm{erg\,cm^{-2}\,s^{-1}})$			(10^{-3} counts s^{-1})	M1	M2
(1)	(2)	(3)	(4)	(5)	(6)	(7)
68	$1.2\pm^{5.6}_{1.2}\times10^{43}$	$0.625\pm^{0.225}_{0.025}$	1	0.9	163	119
69	$4.5\pm^{53}_{4.4}\times10^{43}$	$1.575\pm^{0.125}_{1.075}$	1	0.3	74	50
71	$1.3\pm^{2.5}_{0.8}\times10^{44}$	$1.675\pm^{0.025}_{0.125}$	1	1.1	232	135
86	$6.7\pm^{33}_{6.4}\times10^{41}$	0.125 ± 0.075	1	0.4	82	106
88	$6.2\pm^{102}_{12}\times10^{42}$	$0.625\pm^{0.325}_{0.075}$	1	0.3	102	-
93	$3.5\pm^{12}_{2.8}\times10^{45}$	$2.425\pm^{0.625}_{0.475}$	0	0.1	33	-
96	$1.8\pm^{62}_{1.8}\times10^{43}$	$0.675\pm^{0.525}_{0.575}$	1	0.3	8	76
99	$4.0\pm^{171}_{3.5}\times10^{42}$	$0.275\pm^{0.525}_{0.025}$	1	1.0	188	146
103	$8.4\pm^{57}_{12}\times10^{41}$	0.225 ± 0.075	1	0.4	94	108
109	$1.8\pm^{478}_{1.8}\times10^{42}$	$0.275\pm^{0.925}_{0.275}$	1	0.1	73	14

NOTES: Morph. is morphology as classified by SDSS: 0=point, 1=extended. Mean rate is from the 1000s M1+M2 light curve unless only one camera is used in which case it is taken from the single camera light curve. Total counts are after background subtraction, rounded to the nearest whole count.

ease of viewing.

4.2.5 Variability calculations

For each 15000 s bin light curve of each source, two parameters were calculated - a χ^2 estimation of whether the light curve is variable and the excess variance. The χ^2 value tested against the hypothesis that the light curve is constant is:

$$\chi^2 = \sum_{i=1}^{N} \frac{(R_i - \mu)^2}{\sigma_i^2} \tag{4.17}$$

where R_i is the i^{th} count rate in the light curve, σ_i is the error on the rate of the i^{th} bin and μ is the unweighted mean:

$$\mu = \frac{\sum_{i=1}^{N} R_i}{N} \tag{4.18}$$

The p value associated with each χ^2 value was then calculated. For sources with $p \leq 0.1$, the null hypothesis of a constant light curve is rejected at the 90% confidence level and the light curve is described as variable.

For variable light curves (i.e. those with $p \leq 0.1$) the normalised excess variance was calculated:

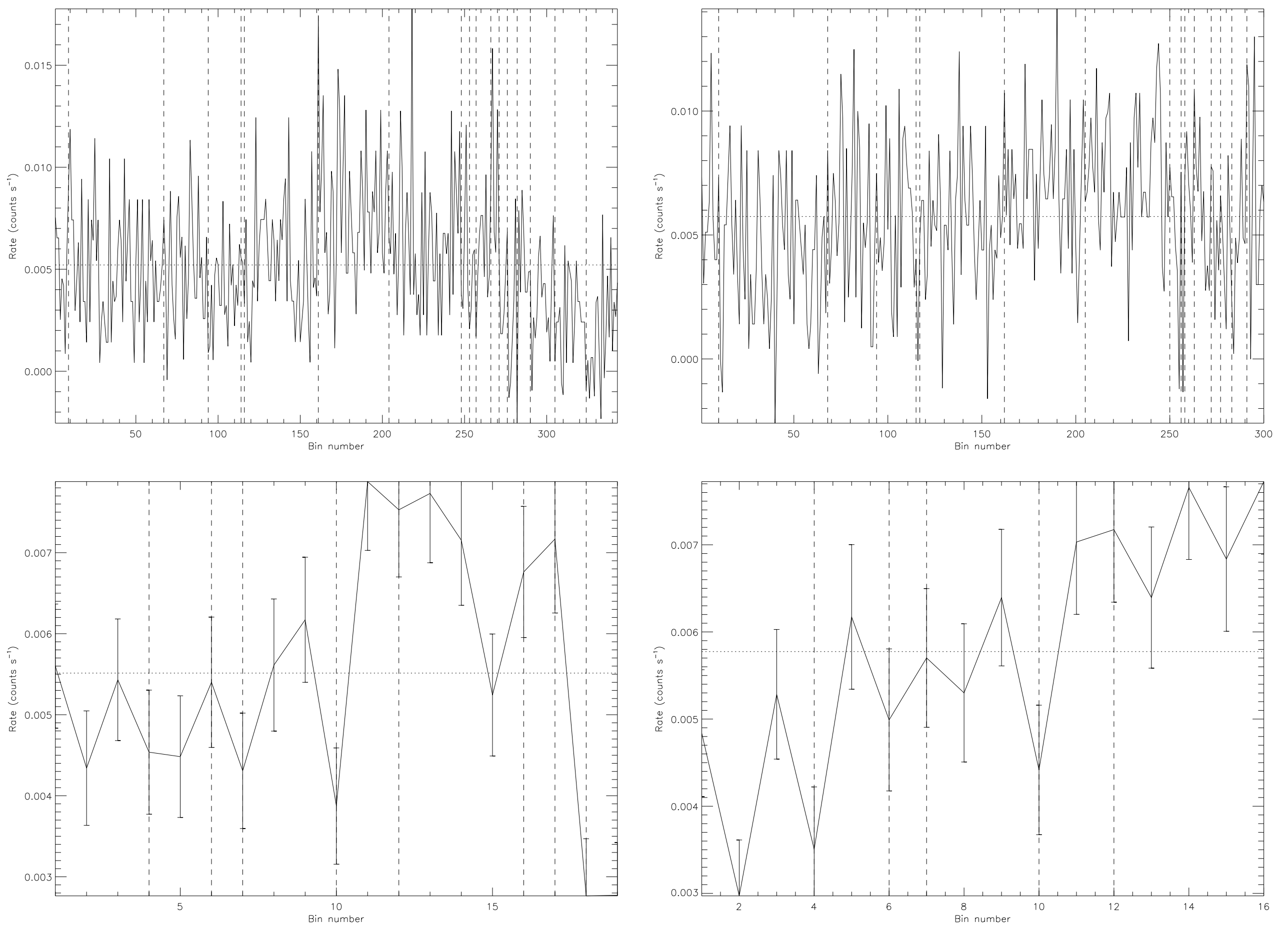


Figure 4.3: Source 3 light curves. M1 light curves are on the left hand side and M2 on the right for 1000 s (top) and 15000 s (bottom) bins.

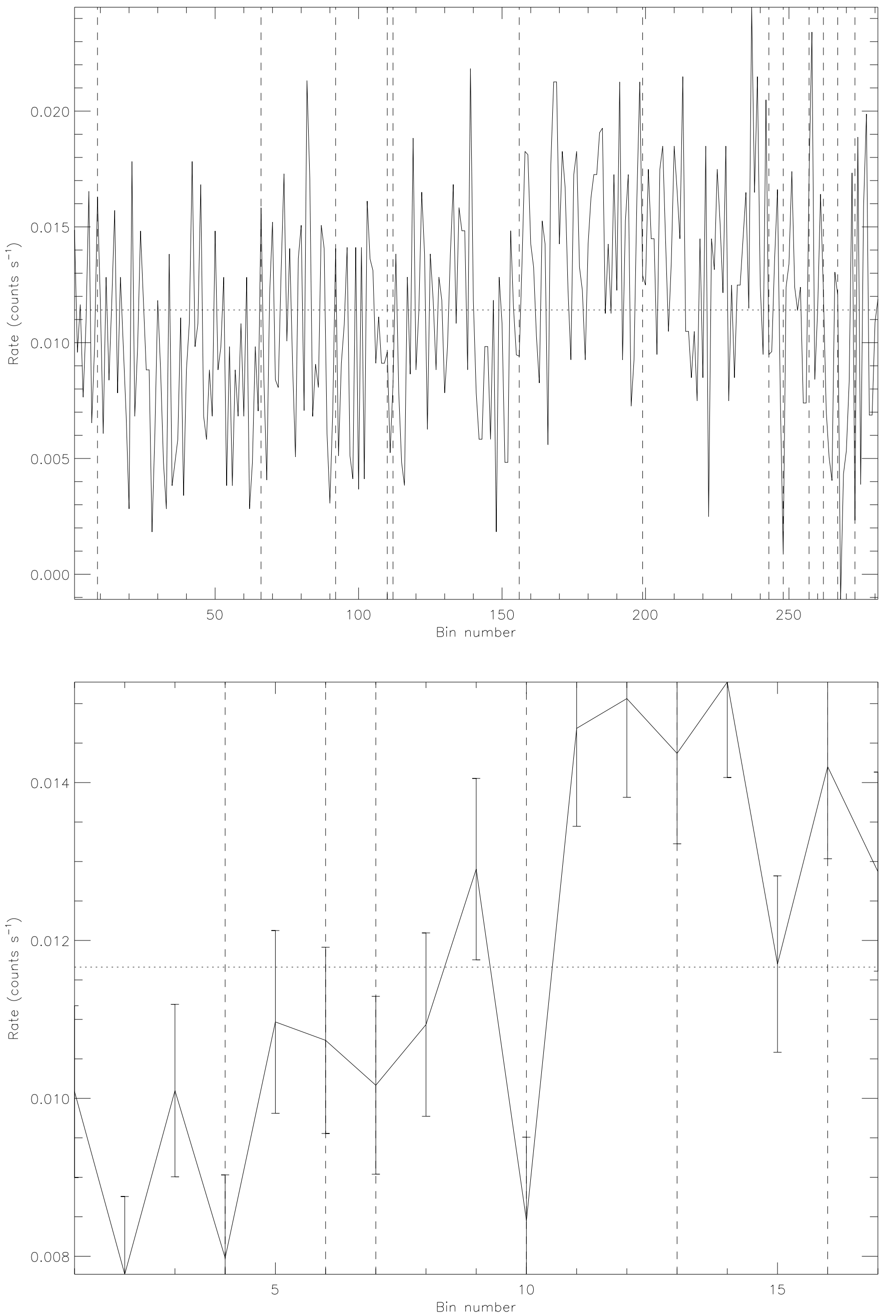


Figure 4.4: Combined M1+M2 light curves for source 3. Top: 1000 s bins. Bottom: 15000 s bins.

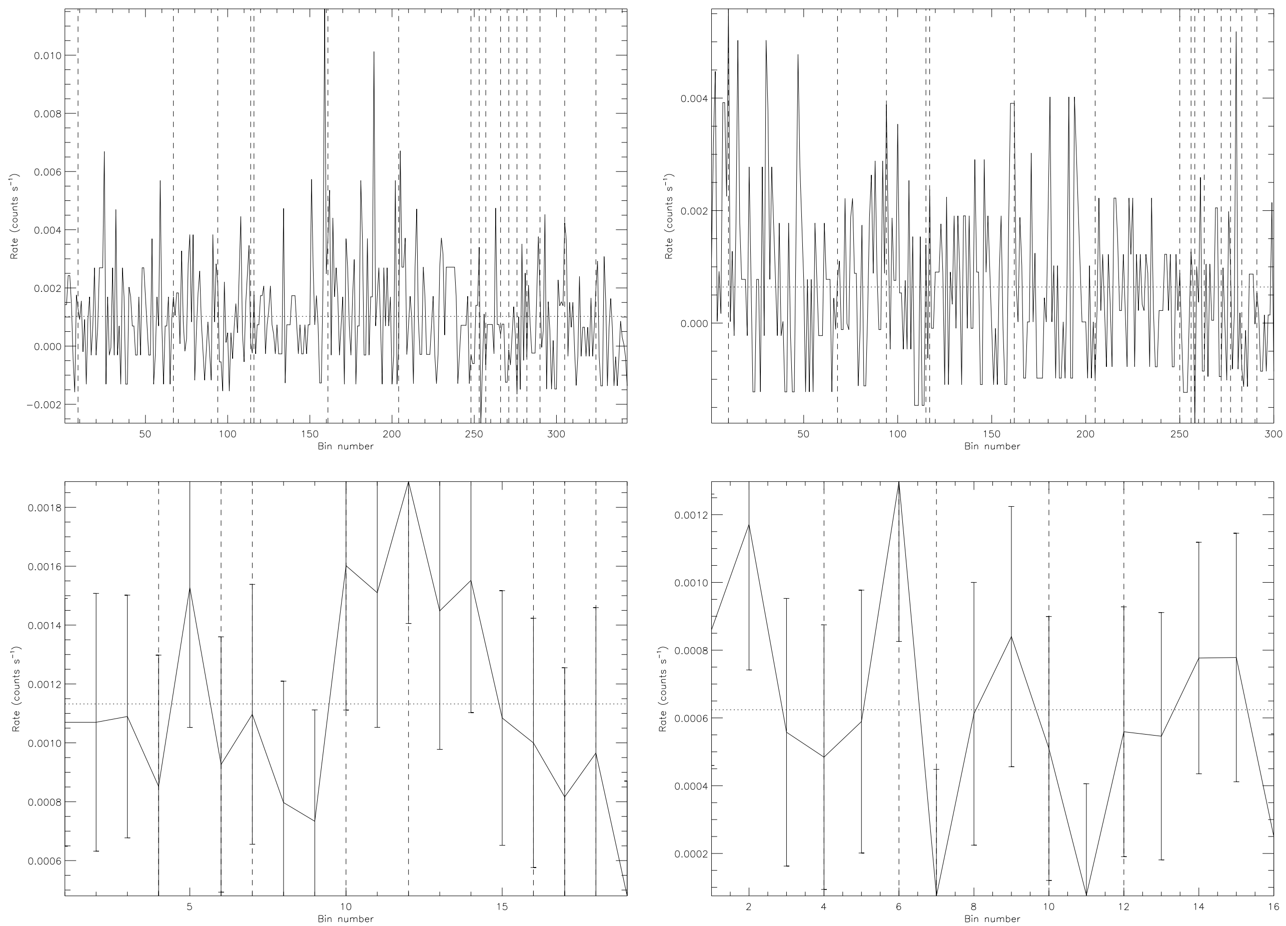


Figure 4.5: Source 24 light curves. M1 light curves are on the left hand side and M2 on the right for 1000 s (top) and 15000 s (bottom) bins.

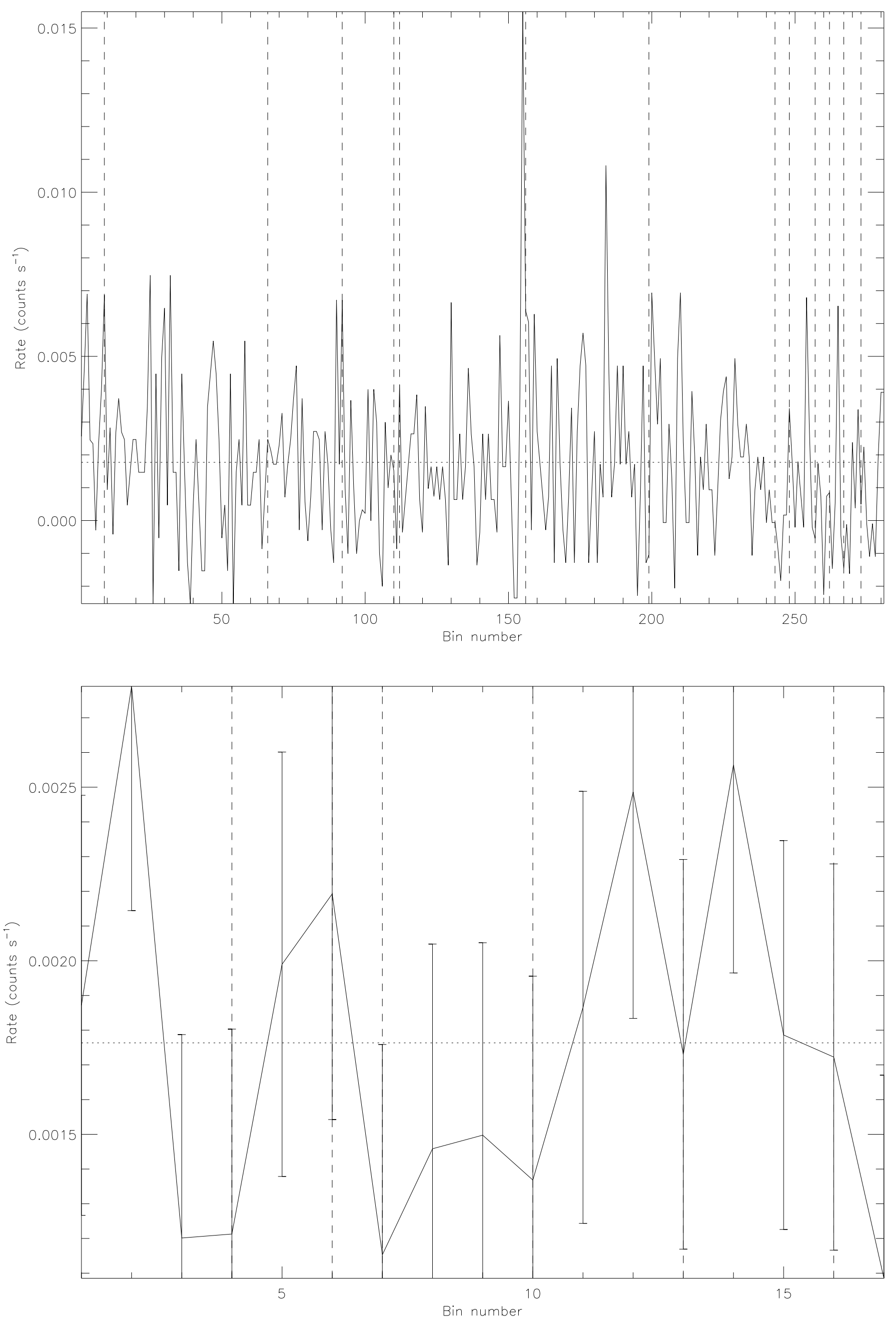


Figure 4.6: Combined M1+M2 light curves for source 24. Top: 1000 s bins. Bottom: 15000 s bins.

$$\sigma_{nxs}^2 = \frac{S^2 - \overline{\sigma_{err}^2}}{\overline{x}^2} \tag{4.19}$$

where

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (R_i - \mu)^2 \tag{4.20}$$

and

$$\overline{\sigma_{err}^2} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \tag{4.21}$$

The error on the excess variance, as defined in Vaughan *et al.* (2003a), is:

$$\Delta\sigma_{nxs}^2 = \sqrt{\left(\sqrt{\frac{2}{N}} \frac{\overline{\sigma_{err}^2}}{\mu^2}\right)^2 + \left(\sqrt{\frac{\overline{\sigma_{err}^2}}{N}} \frac{2F_{var}}{\mu}\right)^2} \tag{4.22}$$

where F_{var} is the fractional rms variability amplitude:

$$F_{var} = \sqrt{\sigma_{nxs}^2} \tag{4.23}$$

Note that the error on σ_{nxs}^2 defined above accounts only for the measurement error in the light curve and not for the random scatter intrinsic to the light curve.

These values were calculated for each source and camera (M1, M2 and combined

M1+M2) for which a $15000\,\mathrm{s}$ bin light curve was extracted (i.e. those for which Gaussian statistics were appropriate). The results of the $15000\,\mathrm{s}$ light curve analysis are shown in table 4.3 for the variable sources ($p \leq 0.1$).

Some previous studies, e.g. Nandra *et al.* (1997) and Turner *et al.* (1999), have used a slightly different equation for the estimation of excess variance:

$$\sigma^2_{rms} = \frac{1}{N\mu^2}\sum_{i=1}^{N}[(R_i - \mu)^2 - \sigma_i^2] \tag{4.24}$$

Using this equation instead of the one presented above made little difference to the results of the sample (typically around -0.03 to σ^2_{nxs}).

4.3 Results

Of the 50 sources with optical counterparts ($P_{id} \geq 0.9$), 32 (29, 28) have enough counts for Gaussian statistics to be appropriate for $15000\,\mathrm{s}$ bin light curves in the MOS1 (MOS2, both) camera and are therefore eligible for the light curve analysis. The percentage of sources that are deemed to be variable at the 90%, 95% and 99% confidence level (i.e. p values of $\leq 0.1, 0.05$ and 0.01 respectively) for each camera is shown in table 4.4.

The normalised excess variance of the $15000\,\mathrm{s}$ light curves is plotted against the full band ($0.2 - 12\,\mathrm{keV}$) X-ray luminosity for the variable sources ($p \leq 0.1$) in figure 4.7.

Table 4.3: Variability results of $10000\,$s bin light curves

	M1		M2		M1+M2	
Source	p	σ^2_{nxs}	p	σ^2_{nxs}	p	σ^2_{nxs}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
3	1.6×10^{-9}	0.062±0.016	2.0×10^{-6}	0.041±0.014	2.4×10^{-10}	0.036±0.009
8	0.0009	0.09±0.04	0.0737	0.07±0.04	1.5×10^{-5}	0.11±0.03
9	0.0001	0.10±0.04	0.6355	-	0.0067	0.07±0.03
10	0.0002	0.23±0.07	0.7412	-	0.0439	0.07±0.04
11	0.4575	-	0.1831	-	0.1635	-
12	0.0111	0.15±0.08	0.4629	-	0.0019	0.13±0.06
13	0.1499	-	5.6×10^{-9}	0.20±0.06	1.6×10^{-5}	0.10±0.04
14	0.0712	0.020±0.013	2.0×10^{-12}	0.21±0.05	1.6×10^{-8}	0.072±0.019
16	0.3573	-	0.2168	-	0.1176	-
17	0.8540	-	0.1235	-	0.2151	-

Table 4.3: continued

	M1		M2		M1+M2	
Source	p	σ^2_{nxs}	p	σ^2_{nxs}	p	σ^2_{nxs}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
19	0.3659	-	N/A	N/A	N/A	N/A
21	0.0011	0.25±0.09	0.1337	-	0.0054	0.08±0.04
22	0.0078	0.8±0.4	0.3073	-	0.0703	0.31±0.18
24	0.8443	-	0.7175	-	0.7917	-
26	0.1293	-	0.8981	-	0.6508	-
35	0.9167	-	0.3400	-	7554	-
38	0.6995	-	N/A	N/A	N/A	N/A
39	0.0217	1.3±0.6	0.4970	-	0.1577	-
46	0.0564	1.1±0.6	0.3926	-	0.0647	1.2±0.7
50	0.0282	0.44±0.25	0.0126	0.20±0.11	0.0060	0.20±0.10
55	0.0291	1.5±1.0	0.5905	-	0.0593	1.9±1.0
56	N/A	N/A	0.1194	-	N/A	N/A
64	0.9973	-	0.3693	-	0.8094	-
68	0.3783	-	0.5510	-	0.3866	-
69	0.2251	-	0.8509	-	0.5638	-
71	0.0001	0.79±0.24	0.9901	-	00097	0.50±0.18
86	0.2432	-	0.5463	-	8974	-
88	0.8955	-	N/A	N/A	N/A	N/A
93	0.2212	-	N/A	N/A	N/A	N/A

Table 4.3: continued

	M1		M2		M1+M2	
Source	p	σ^2_{nxs}	p	σ^2_{nxs}	p	σ^2_{nxs}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
96	0.7841	-	0.9408	-	0.8520	-
99	0.3641	-	0.1908	-	0.4221	-
103	0.6360	-	0.2548	-	0.4803	-
109	0.4890	-	0.6900	-	0.5486	-

NOTES: p is an estimate of the likelihood of the resulting χ^2 for a constant light curve. If $p \leq 0.1$, the source is deemed to be variable and normalised excess variance, σ^2_{nxs} was calculated. N/A indicates that there were not enough counts to perform the analysis.

Table 4.4: Variable sources (10000s light curves)

	Percentage of sources detected to be variable		
Camera	$> 90\%$ conf.	$> 95\%$ conf.	$> 99\%$ conf.
(1)	(2)	(3)	(4)
MOS1	41	34	22
MOS2	17	14	10
M1+M2	46	36	32

The top plot shows the results for single camera light curves and the bottom plot is the combined camera result.

Despite the large scatter in the points, the previously observed anti-correlation between normalised excess variance and X-ray luminosity is clearly visible. A Spearman's rank test of the results gives correlation coefficients of -0.759 and -0.720 for the separate camera and combined camera light curves respectively. The p values for the correlations are 2.6×10^{-4} and 5.5×10^{-3} for separate cameras and combined cameras respectively suggesting that the correlation between normalised excess variance and X-ray luminosity is significant. Following previous studies, a fit of the form:

$$\sigma_{rms}^2 \propto L_x^{-\alpha} \tag{4.25}$$

is made to the data. For the separate camera case, $\alpha = 0.17 \pm 0.04$ and for combined camera, $\alpha = 0.22 \pm 0.05$, although the χ^2 goodness of fit tests show that these are formally very poor fits (due to the large errors on the points and significant scatter) - $\chi^2 \approx 58$ (29) for the separate (combined) camera plots. The fits are shown as a solid line in figure 4.7.

Explanations for the anti-correlation between luminosity and variability are based on the assumption that more luminous sources have larger emitting regions. Variability takes longer to propagate through a larger region and the variability is therefore slower. The emitter may be either a single coherent emitter or, as proposed by Abramowicz *et al.* (1991), made up of independent hotspots which cause variations as they rotate through

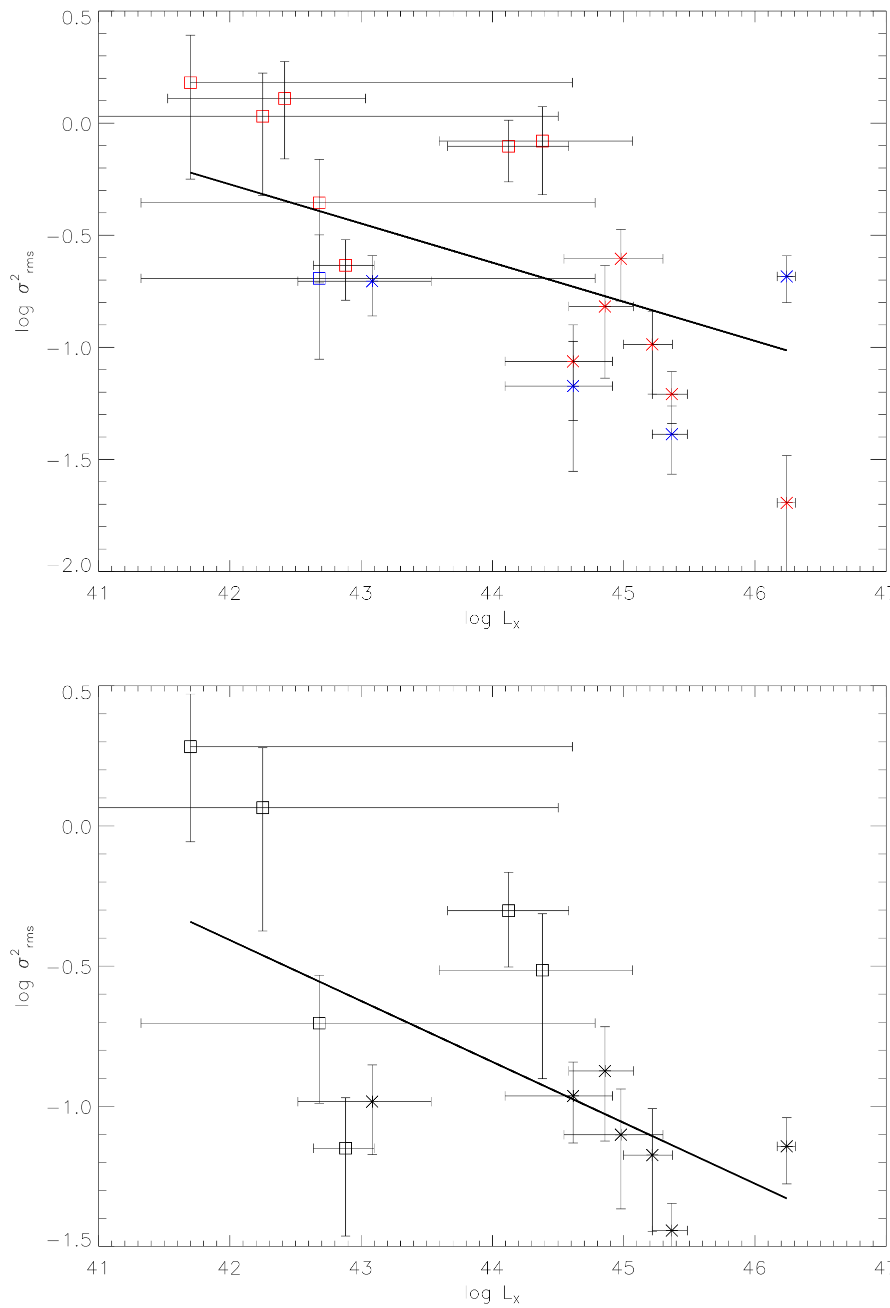


Figure 4.7: X-ray luminosity and normalised excess variance of 15000s bin light curves for variable ($p \leq 0.1$) sources. Top: Results from separate cameras. Red symbols are from M1 light curves and blue symbols from M2 light curves. Bottom: Combined camera results. Asterisks represent point counterparts and squares represent 'extended' counterparts. The line on each plot is the linear fit to the points.

the observer's line of sight. More luminous objects would have more hotspots leading to smaller rms fluctuations. Almaini *et al.* (2000) investigated these models and found that, depending on the parameters and geometry chosen, either one could explain the observed anti-correlation. An alternative model was proposed by Abrassart and Czerny (2000) where the variability is caused by the X-ray emission being partly obscured at times by optically thick clouds.

The power law indices found for the correlations are flatter than that determined by Nandra *et al.* (1997) who found $\alpha = 0.71 \pm 0.03$ with ≈ 2 minute bins. Markowitz and Edelson (2004) however found that the slope of the power law flattened as the timescale sampled increased up to around the 200 day timescale after which the slope was constant. They found $\alpha = 0.316 \pm 0.024$ for one day timescale decreasing to $\alpha = 0.223 \pm 0.012$ for a six day timescale. The sampling used here fits in with this finding, being flatter than the much shorter timescale result of Nandra *et al.* (1997) but steeper than the much longer timescale (six day) result of Markowitz and Edelson (2004). The possible explanation for the flattening of the power law towards longer timescales comes from power spectrum analysis. Studies of Seyfert 1 power spectra (e.g. Markowitz *et al.* 2003) have shown that the power spectrum can be described by a broken power law, flatter below the break and steeper above it. The break timescale scales linearly with the mass of the central black hole, with more massive black holes having lower frequency breaks. Markowitz and Edelson (2004) suggest that for their sample, with its large range of luminosities and black hole masses, at the longest timescales probed they are exploring variability below the power law break in most sources and so the variability amplitudes of most sources are at their flattest and the slope of the luminosity-variability slope is therefore also flatter. This effect could be the explanation for the flatter slope found here compared to the slope of Nandra *et al.* (1997).

In comparison to previous studies (e.g. Nandra *et al.* 1997, Turner *et al.* 1999) it is clear that the scatter in the detected anti-correlation is large. The large scatter in luminosity is explained to some extent because it is determined using photometric redshifts and not spectroscopically. Turner *et al.* (1999) explain the greater scatter in the correlation found using their sample compared to that of Nandra *et al.* (1997) as being due to the inclusion of narrow line Seyfert 1 galaxies (NLS1s) instead of just sampling broad line Seyferts. Here, there is no determination of the type of AGN included in the sample and this may also contribute to the scatter.

Almaini *et al.* (2000) found a possible change in the correlation with increasing redshift. The samples of Nandra *et al.* (1997) and Turner *et al.* (1999) are both on low redshift sources ($z < 0.1$). The use of sources here with redshifts out to around 2.5 may also impact on the correlation. To investigate this, the sample was split into low ($z \leq 0.5$) and high ($z > 0.5$) redshift sources. Figure 4.8 shows the plot of luminosity against normalised excess variance. These plots show tentative evidence that the correlation is flatter for high redshift sources than for low redshift sources. The values of α for the separate (combined) camera cases are 0.84 ± 0.26 (1.00 ± 0.27) for the low redshift sources and 0.26 ± 0.08 (0.32 ± 0.10) for high redshift sources. As with the earlier fits, χ^2 goodness of fit tests show that the fits are formally poor.

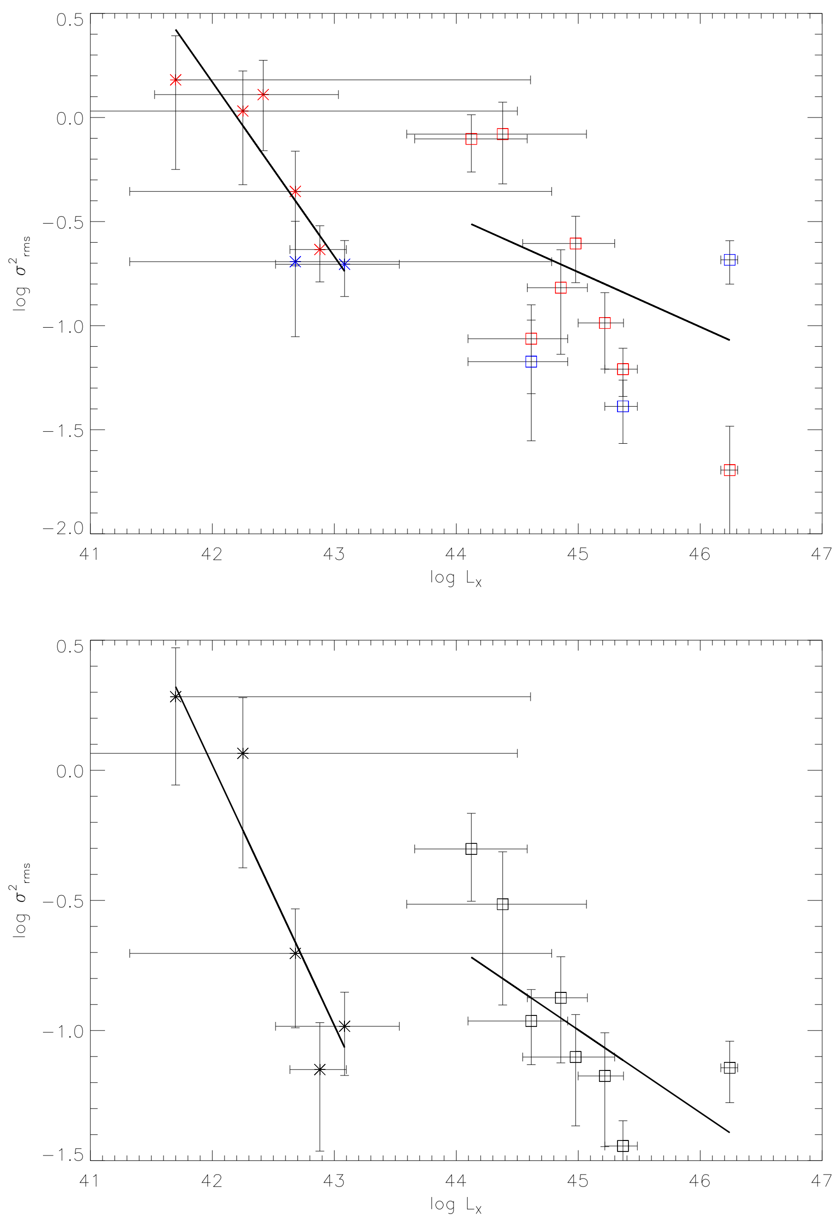


Figure 4.8: X-ray luminosity and normalised excess variance of 15000s bin light curves for variable ($p \leq 0.1$) sources. Top: Results from separate cameras. Red symbols are from M1 light curves and blue symbols from M2 light curves. Bottom: Combined camera results. Asterisks represent low redshift ($z \leq 0.5$) sources and squares represent high redshift ($z > 0.5$) sources. The lines on the plot are the fits to the low redshift (left) and high redshift (right) sources respectively.

To deal with the low count rates encountered with the sources in the 3C 273 field, the light curves were 'binned up' until enough counts were in each bin to allow Gaussian statistics and therefore the normalised excess variance and χ^2 variability tests to be used. Other approaches to deal with low count statistics are also possible. Rather than binning the data until Gaussian statistics are appropriate, the C-statistic (Cash 1979) can be used with any number of counts as it is based on Poisson statistics, appropriate for low count rates. Almaini *et al.* (2000) used ensemble light curves because of the low signal to noise ratio in many of the individual light curves in their sample. They found that sources with greater flux were more likely to be variable (a result that is repeated with the sources used here) and noted that those that were found to be 'non-variable' probably just had too low a signal to noise ratio for the variability to be detectable. The number of useable light curves sampled in this study limits the usefulness of this approach here but use of ensemble light curves with bigger samples may provide further insight into the relation between X-ray luminosity and excess variability.

As was found with the Lockman Hole sample in Chapter 3, use of deep observations such as the one used here means that very faint sources are detected (down to flux of $2.6 \times 10^{-15}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in this case). At these fluxes, the accuracy of the photometric redshifts is reduced, leading to large errors on X-ray luminosity, and the S/N of the light curves is low. Use of the technique used here with brighter samples may lead to more statistically significant results.

4.4 Conclusion

The work in this chapter has shown that various techniques developed in this thesis (e.g. photometric redshifts, cross correlation with SDSS) can be used with archive data of serendipitous sources to find significant results. The light curves of the serendipitous sources in the 3C 273 field were used here, despite having low count rates, to confirm the anti-correlation between excess variance and X-ray luminosity previously detected in samples of low redshift AGN.

Flatter slopes of the correlation than observed in previous works (e.g. Nandra *et al.* 1997) were found. This may be due to the sample covering a large range of luminosity and therefore black hole mass with light curves extending to longer timescales than before. Some tentative evidence was found, as in Almaini *et al.* (2000), for a change in the correlation between low and high redshift samples. This finding would benefit from further study with larger samples with more accurate luminosity and redshift determinations.

The work presented in this chapter can be seen as a pilot for bigger projects of this kind, for example studies of other fields in the *XMM-Newton* Science Archive. Similar analysis could be performed for bright objects in 2XMM with automatically extracted light curves.

Chapter 5

X-ray and optical variability of Seyfert 1 galaxies as observed with *XMM-Newton*

5.1 Introduction

Active galactic nuclei (AGN) emit radiation over a broad range of wavelengths from radio through to X-rays and are variable over this entire spectral range. It is thought that the variability in different bands may be connected or driven by a single mechanism. This chapter examines the potential connection between optical and X-ray variability in Seyfert 1 galaxies.

Optical and ultraviolet (UV) continuum emission[1] in AGN is believed to be thermal

[1]Throughout this chapter, the term 'optical' is used to refer to the optical-UV bandpass of the Optical Monitor on *XMM-Newton*.

emission originating in the accretion disk surrounding the central supermassive black hole (SMBH). The X-ray emission, which shows the most rapid variability, is thought to come from the hot 'corona' close to the SMBH. There are two main theories regarding a connection between these two bands: reprocessing of X-rays into thermal optical emission or Compton upscattering of optical photons to X-ray energies. In the X-ray reprocessing scenario, X-rays from the corona are absorbed in the disk, heating it, which then re-radiates this energy as optical photons (e.g. Guilbert and Rees 1988). The converse theory is that photons from the disk enter the corona and are Compton upscattered by hot electrons to be emitted as X-ray photons (e.g. Sunyaev and Titarchuk 1980, Haardt and Maraschi 1993). It is currently unclear how important these mechanisms are in forming the continuum emission from AGN.

If either of these processes is important in AGN then their effect should be observable in the variability of the source. If optical photons drive the X-ray photons then an increase in flux in the optical light curve should be repeated in the X-ray band a short while later and vice versa. By cross correlating X-ray and optical light curves to look for such a correlation, possible connections between the different emission regions can be inferred. The traditional way to achieve this is to schedule simultaneous observations of a source on multiple telescopes. *XMM-Newton* provides a simpler way of performing these observations as it carries the Optical Monitor as well as three X-ray telescopes and can therefore perform X-ray and optical observations simultaneously. This capability has been exploited here to examine a small sample of objects and study the relationship between X-ray and optical variability.

Previous studies of this subject have provided a confusing array of results. In some cases, no correlation is detected on short timescales but light curves are correlated when averaged over a longer period (e.g. NGC 4051; Peterson *et al.* 2000). NGC 4051 was also studied by Shemmer *et al.* (2003) who found evidence that the X-rays were led by the optical emission by ≈ 2 days. For Ark 564, on the other hand, Shemmer *et al.* (2001) found that the UV lags the X-ray emission by 0.4 days and optical lags UV by ≈ 2 days. There are also cases where the X-ray and optical light curves are correlated with no lag. In a six year monitoring programme of NGC 5548 by Uttley *et al.* (2003), a strong correlation between X-ray and optical light curves was detected at lag 0 ± 15 days from data averaged at 30 day intervals. For further examples see Uttley (2005). It is obvious that further work needs to be done to resolve the confusion surrounding reprocessing.

Only two previous studies have used the simultaneous observing capabilities of *XMM-Newton* to examine variability correlations in AGN. Mason *et al.* (2002) used a 1.5 day long *XMM-Newton* observation of NGC 4051 supplemented with data from the *Rossi X-ray Timing Explorer* (*RXTE*) and found three prominent features in which the X-ray leads the optical emission. Using Monte Carlo simulations they derived a confidence of 85% in these features. Arévalo *et al.* (2005) analysed the long ($300\,$ks) observation of MCG$-$6-30-15 and reported a significant correlation with the UV leading the X-rays by 1.8 days. These two objects were re-examined here along with data from six other Seyfert 1 galaxies.

5.2 Data reduction

5.2.1 Sample definition and observations

The satellite *XMM-Newton* carries three X-ray telescopes and a co-aligned Optical Monitor (OM; Mason *et al.* 2001), comprising a $30\,\mathrm{cm}$ diameter Ritchey-Chrétien telescope with a micro-channel plate intensified CCD. This gives a field of view of $17\mathrm{arcmin}^2$ and spatial resolution of $\approx 1''$, with $0.48''$ pixels. The OM can take images using any of six broad band filters sensitive over the range $1800 - 6000$ Å, thereby providing optical imaging and timing information simultaneous with the X-ray observations. The combination of long X-ray exposures with simultaneous optical data makes *XMM-Newton* ideal for probing the relation between X-ray and optical variability on short timescales. For more details on the OM see section 2.2.2.

The *XMM-Newton* Science Archive (XSA) contains all available data products from the satellite (see section 2.3.1). The XSA was used to find long observations ($> 30\,\mathrm{ks}$) of nearby ($z < 0.1$) Seyfert 1 galaxies. Objects were selected if there were fifteen or more exposures taken with a single filter of the OM. Eight objects were found to satisfy these criteria: 1H 0707-495, Ark 120, Ark 564, MCG−6-30-15, Mrk 766, NGC 3783, NGC 4051 and NGC 7469. Four of these sources had observations in two or more XMM revolutions in the same filter. A summary of the observations of these sources is given in table 5.1.

Table 5.1: Observation summary

Object	Redshift	On time	Revolution	OM filter	No. OM exposures	Bin size	p	OM rms	X-ray rms
		(s)				(s)		%	%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1H 0707-495	0.0411	46018	0159	UVW2	15	1296	0.067	2.0	40.1
Ark 120	0.0322	112130	0679	UVW2	80	1296	0.051	0.7	2.0
Ark 564	0.0246	101774	0930	UVW1	49	2160	0.088	0.3	22.8
MCG−6-30-15	0.0077	349286	0301–0303	U	275	1123	<0.01	1.9	28.2
Mrk 766	0.0129	129906	0265	UVW1	75	1296	0.95	0.0	25.4
NGC 3783 (1)	0.0097	40412	0193	UVW2	30	1296	0.36	0.4	5.8

Table 5.1: Observation summary

Object	Redshift	On time	Revolution	OM filter	No. OM exposures	Bin size	p	OM rms	X-ray rms
		(s)				(s)		%	%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
NGC 3783 (2)		275633	0371–0372	UVW2	53	4752	<0.01	2.9	20.7
NGC 4051 (1)	0.0023	121958	0263	UVW1	43	2160	<0.01	1.5	47.6
NGC 4051 (2)		51866	0541	UVW1	15	2246	<0.01	0.8	43.8
NGC 7469	0.0163	164128	0912–0913	UVW2	111	1296	<0.01	1.9	7.7

NOTES: Column (6) is the bin size of the X-ray and optical light curves (where the X-ray light curves have been binned to match the optical sampling). Column (7) is the χ^2 p value of the light curve. A value of < 0.01 indicates variability at 99% confidence. Columns (8) and (9) contain the root mean square percentage variability amplitude of the optical and X-ray light curves respectively. Where multiple, non-consecutive observations of an object were made they are labelled with separate numbers to avoid confusion e.g. NGC 3783 (1)

5.2.2 Predicted time delays for the sample

The eight targets selected all have published black hole mass estimates, allowing simple predictions to be made for the inter-band time delays that might be expected from reprocessing models. First, all heating in the disk was assumed to be generated by the viscosity of the disk itself. Different radii of the disk are at different temperatures and can be modelled as blackbodies with a particular peak wavelength. Using the effective wavelength (λ_{peak}) of the appropriate OM filter, the temperature (T) of a blackbody spectrum that peaks in the observed band was calculated using Wien's Law:

$$\lambda_{peak} = \frac{2.898 \times 10^{-3} mK}{T} \tag{5.1}$$

This temperature was then used with the standard formulae for a geometrically thin, optically thick accretion disk to determine the approximate radius from which the emission originates:

$$T(r) \approx 6.3 \times 10^5 \left(\frac{\dot{M}}{\dot{M}_E}\right)^{1/4} M_8^{-1/4} \left(\frac{r}{R_S}\right)^{-3/4} \tag{5.2}$$

where $\dot{M}/\dot{M}_E$ is the accretion rate as a fraction of the Eddington accretion rate, M_8 is the mass of the black hole in units of $10^8\,\mathrm{M}_\odot$ and r/R_S is the radius in terms of the Schwarzchild radius (Peterson 1997). Using values of 0.01 and 0.1 for $\dot{M}/\dot{M}_E$, and published black hole masses (see table 5.2), a range for the radius of the optical region being observed was derived. Assuming the X-ray emission comes from close to the central black hole, the light travel time to the radius of optical emission region gives a rough estimate for the time delay between these two regions.

As another check on the delay expected, the assumption was made that the disk is heated entirely externally by X-rays. In this case, the energy of the blackbody peaking in the OM filter is comparable to the total X-ray luminosity. The energy received by the disk from the X-ray source is:

$$E_{in} = F_x A = \frac{L_x}{4\pi r^2} A \tag{5.3}$$

where F_x is X-ray flux, A is the area of the disk and L_x is the X-ray luminosity. From Stefan's Law, the energy flowing out of the disk is:

$$E_{out} = A\sigma T^4 \tag{5.4}$$

where σ is Stefan's constant. Assuming that all the X-ray energy is reprocessed:

$$E_{in} = E_{out} \tag{5.5}$$

$$\frac{L_x}{4\pi r^2} = \sigma T^4 \tag{5.6}$$

Using the temperature as calculated above and the X-ray luminosity obtained from the

$0.1-10$ keV X-ray spectrum, the radius of the optical emission region can be calculated. Again, the radius was converted to an approximate time delay in light days. The results of these calculations are shown in table 5.2. The majority of sources are expected to show time delays of less than a day which should theoretically be observable with *XMM-Newton*.

5.2.3 Optical light curves

The Observation Data Files (ODFs) for each source were extracted from the XSA and processed using the task `omichain` of the *XMM-Newton* Science Analysis System (SAS v6.5.0). `Omichain` takes the ODFs and performs all standard OM processing steps on them to produce the final products - images and source lists. Once the images have been created, `omichain` performs source detection and photometry routines. Finally, the source positions are converted from detector to sky coordinates and a combined source list from all exposures is created. An example of a processed image from the OM is shown in figure 5.1.

To create an optical light curve, photometry must be performed on each exposure of each object. Although `omichain` automatically performs photometry on each image, the presence of host galaxy emission in some images causes the routine to flip between 'extended' and 'point source' modes of photometry, and so does not perform a uniform extraction on all the images. As the aim is to obtain accurate photometry of the nucleus and not include the host emission this is obviously undesirable. In order to ensure

Table 5.2: Estimated time delays

Object	M_{bh}	Int. delay	L_x	Ext. delay
	$(10^6\,\mathrm{M}_\odot)$	(days)	$(10^{44}\,\mathrm{erg\,s^{-1}})$	(days)
(1)	(2)	(3)	(4)	(5)
1H 0707-495	2.3[1]	0.02–0.03	0.42	0.16
Ark 120	150 ± 19[2]	0.25–0.54	20.1	1.10
Ark 564	1.09[1]	0.01–0.03	4.5	0.98
MCG−6-30-15	$2.9\pm^{1.8}_{1.6}$[3]	0.03–0.07	1.14	0.69
Mrk 766	0.63[4]	0.01–0.02	1.83	0.62
NGC 3783	29.8 ± 5.4[2]	0.09–0.18	2.55	0.39
NGC 4051	1.91 ± 0.78[2]	0.02–0.04	0.03	0.08
NGC 7469	12.2 ± 1.4[2]	0.05–0.10	3.9	0.48

[1]Zhou and Wang (2005); [2]Peterson *et al.* (2004); [3]McHardy *et al.* (2005); [4]Botte *et al.* (2005)

NOTES: Values in column (3) are based on the assumption of internal disk heating by viscosity. The lower limits use a value of $\dot{M}/\dot{M}_E = 0.01$ and the upper limits use 0.1. X-ray luminosity (column (4)) is for the band $0.1 - 10\,$keV. Values in column (5) are based on the assumption of external disk heating by X-rays.

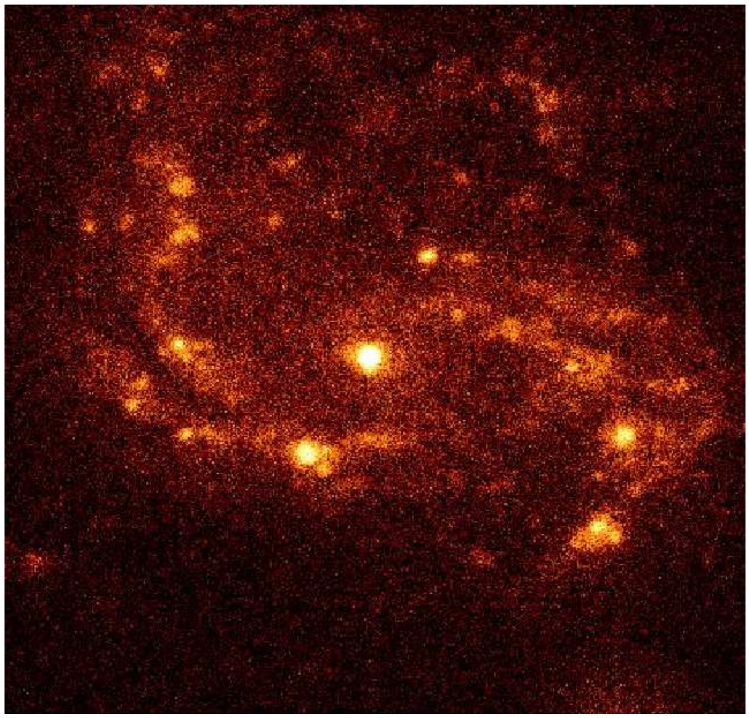


Figure 5.1: 1 ks OM observation of NGC 4051 in the UVW1 filter.

uniform data reduction, custom made scripts were used to perform aperture photometry and apply the appropriate OM corrections to the resulting data, modelled closely on the SAS processing tasks `omdetect` and `ommag`.

Simple aperture photometry was applied to the images produced by `omichain` (after the modulo-8 correction had been applied by the task `ommodmap`). Based on the user defined source position (in detector coordinates), the exact source location was obtained by repeatedly centroiding on the counts within a small box, then the counts within a circular aperture were accumulated. For all sources, except MCG−6-30-15, a circle of radius twelve pixels was used as the source aperture to match the calibration (see below). In the case of MCG−6-30-15, a twelve pixel radius aperture overlapped a nearby star so, to avoid this, an aperture of seven pixels was used for all exposures of MCG−6-30-15. For most sources, an annulus placed around the source aperture was used to extract the background count rate. The background annulus was placed at least 30 pixels out from the source to avoid the extended PSF of the UV filters. In the cases of Ark 564, MCG−6-30-15 and NGC 4051, the host galaxy is visible around the AGN. As the host galaxy should not be varying in flux and it is the flux variability, not absolute flux, that is being studied, no attempt was made to subtract the host galaxy flux from the flux of the central AGN. To get an accurate value for the actual background of the field, a circular background aperture away from the host galaxy was used in these cases.

In order to produce accurate count rates from aperture photometry of OM images there are five corrections that must be applied - point spread function (PSF1), coincidence loss (CL), deadtime (DT), a UV specific PSF correction (PSF2) and time-dependent sensitivity degradation (TDS). Coincidence loss is a loss of flux that occurs when more than one event is collected in the same pixel in a single CCD frame, and the CL correction is a non-linear correction that has been calibrated empirically and found to be self-consistent when applied to an aperture radius of $6''$ (twelve pixels). Before the CL correction can be applied, the PSF1 correction must therefore be implemented to scale the count rate from the aperture used to an effective aperture radius of twelve pixels. When the CCD is read out there is a short amount of time, the frame transfer time, when any incoming photons are not detected. The DT factor corrects for this loss. In the UV filters, the wings of the PSF are more extended than in the optical filters so a further correction is made to count rates from exposures in these filters to account for the loss of count rate from the wings. Finally, the time-dependent degradation factor takes account of the fact that the performance of the OM gradually decreases over time. The uncertainties on each flux point were derived by propagating the Poissonian error on the source and background count rates.

5.2.4 X-ray light curves

X-ray light curves were produced using data from the EPIC pn camera onboard *XMM-Newton*. Event lists for each observation were extracted using the standard processing procedure `epchain`. In order to check for background flaring, $10-15$ keV light curves were extracted from off-source regions and examined by eye for periods of high, variable background. Where such intervals occurred, they were removed by filtering out all time intervals during which the count rate in the high-energy light curve exceeded a critical value. The critical value was chosen to be slightly above the quiescent level, but was slightly different for each observation. After filtering out background flares, source light curves were extracted from a $40''$ circular region using only single or double-pixel events in the range $0.2-10$ keV, and background subtracted using a nearby source-free region. The light curves were binned such that the X-ray sampling is comparable to the sampling of the OM light curves.

Figure 5.2 shows the optical/UV and X-ray light curves of the eight sources. MCG$-$6-30-15, NGC 3783 and NGC 7469 all had consecutive observations which have been combined.

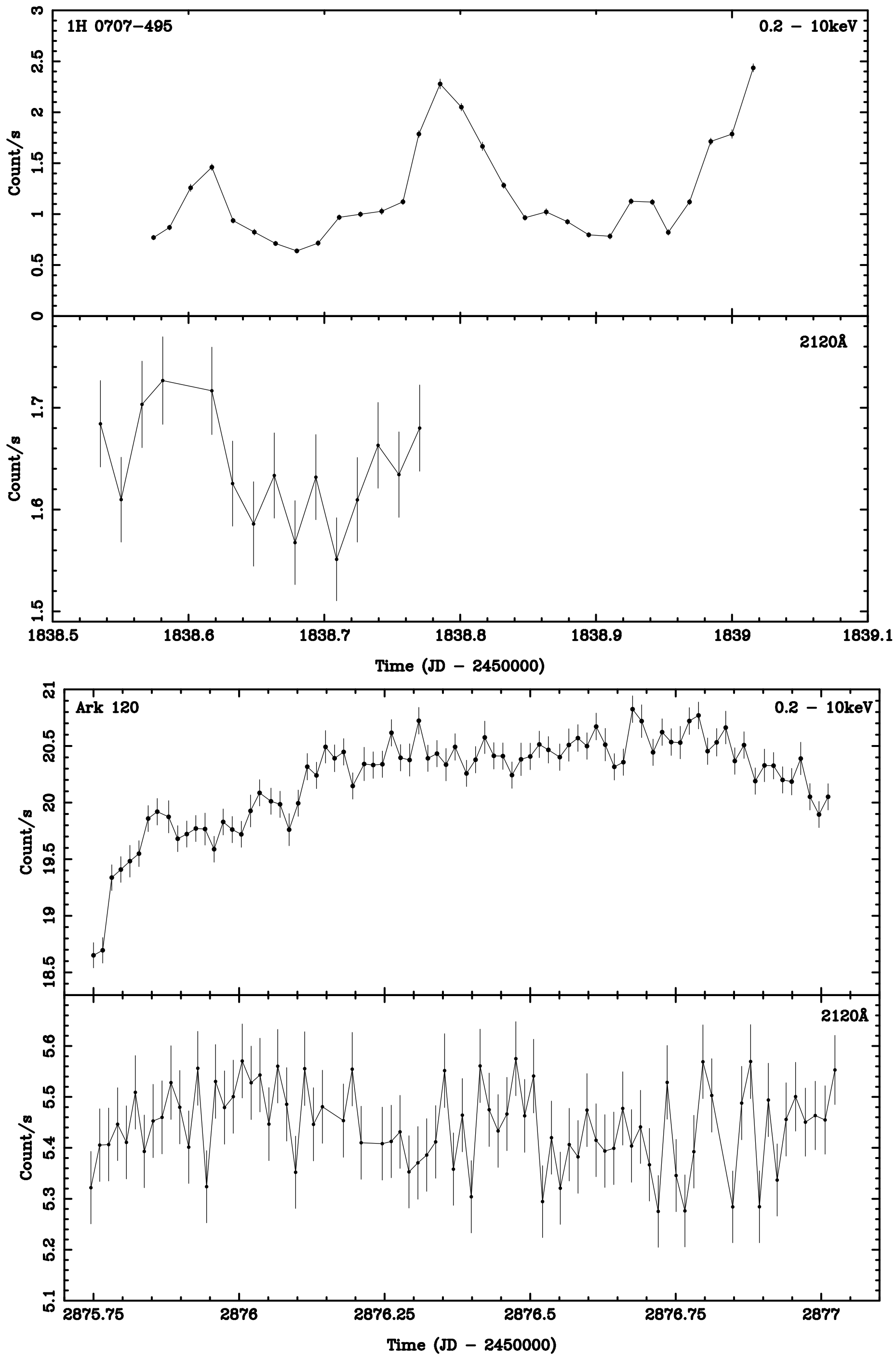


Figure 5.2: X-ray ($0.2 - 10\,\mathrm{keV}$) and optical light curves of the selected sources. X-ray light curves are binned to match the optical sampling. Error bars are included for all count rate measurements but, for several of the X-ray light curves, the error bars are smaller than the data point symbols.

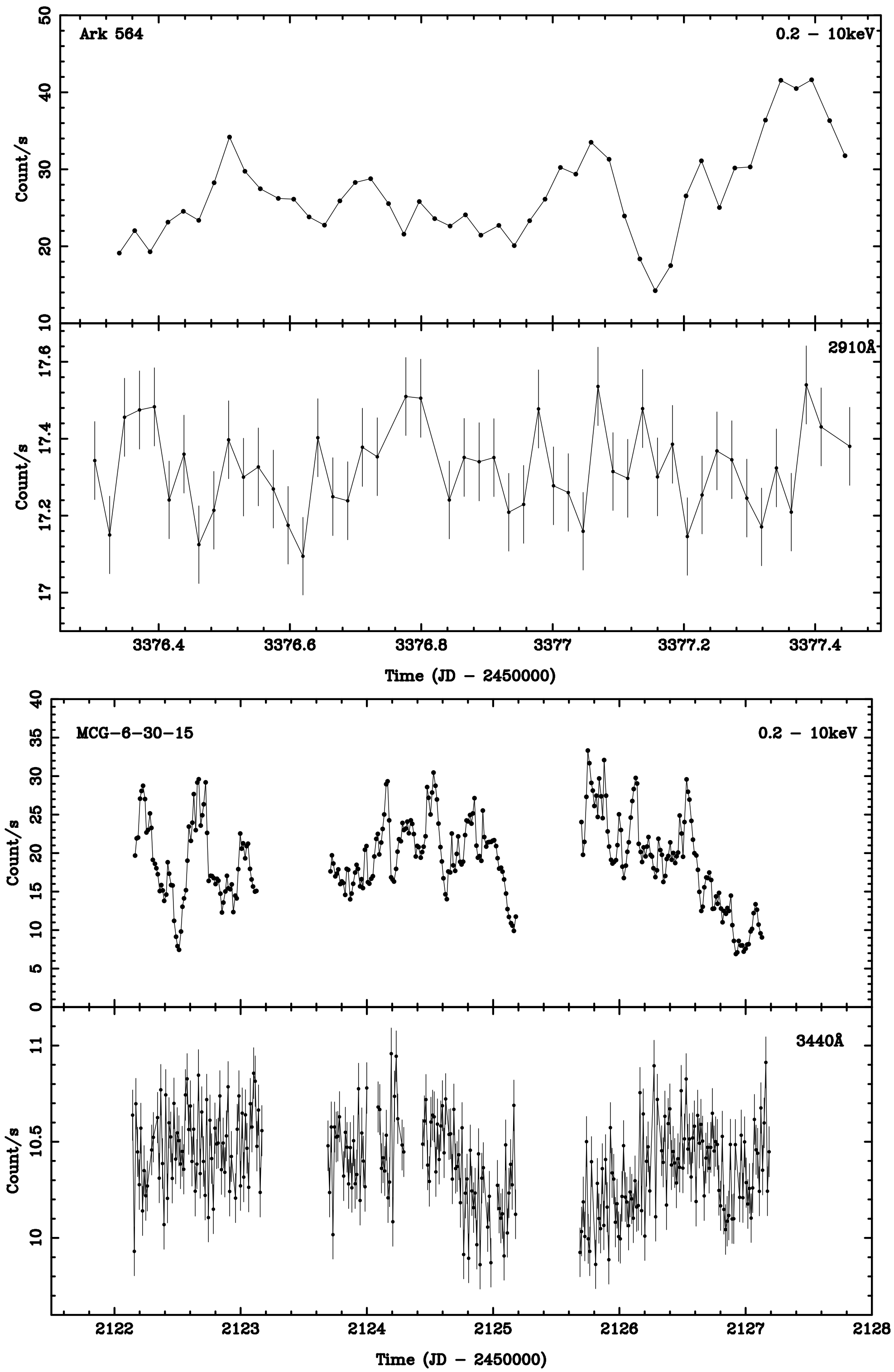


Figure 5.2: continued

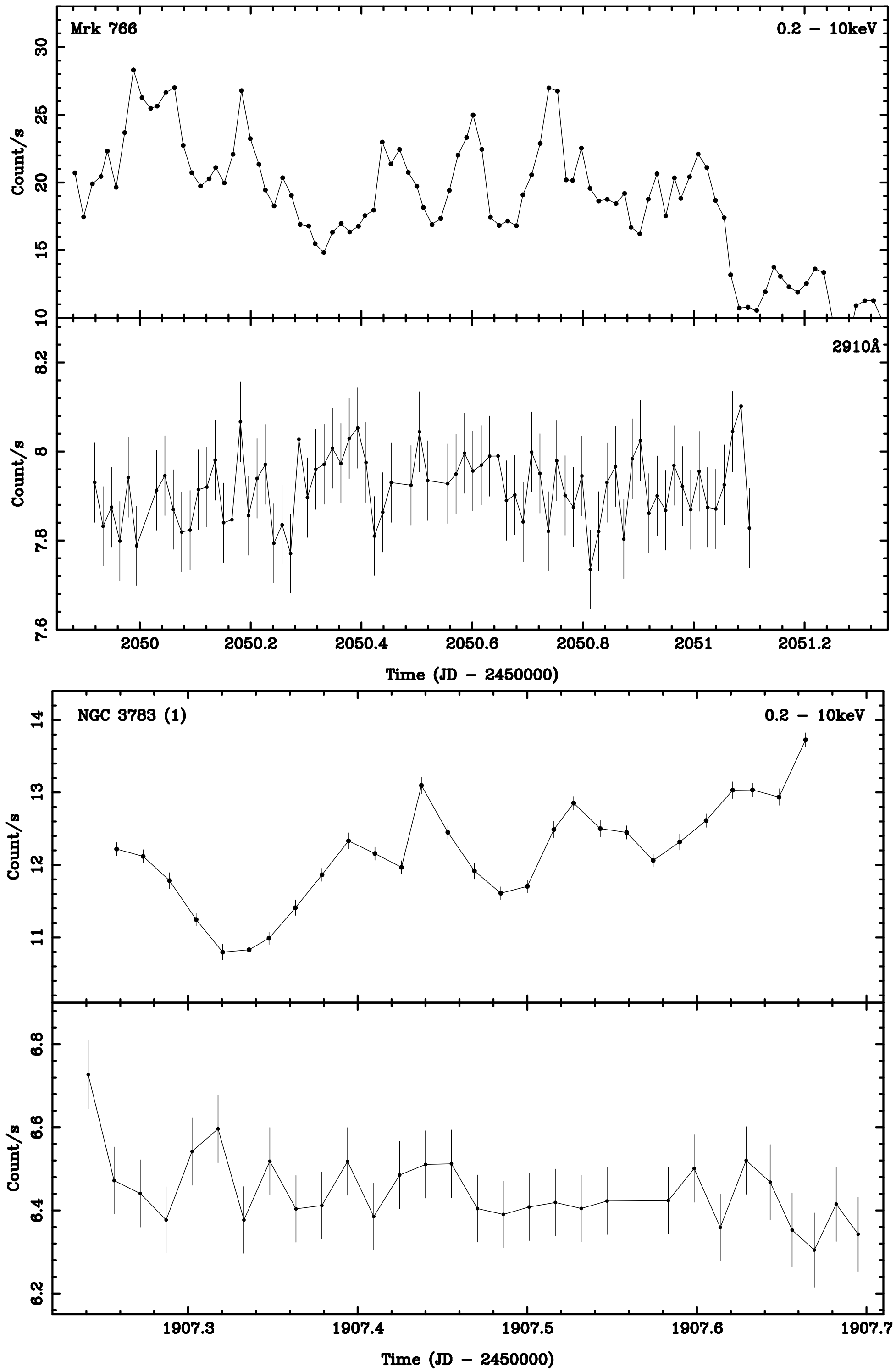


Figure 5.2: continued

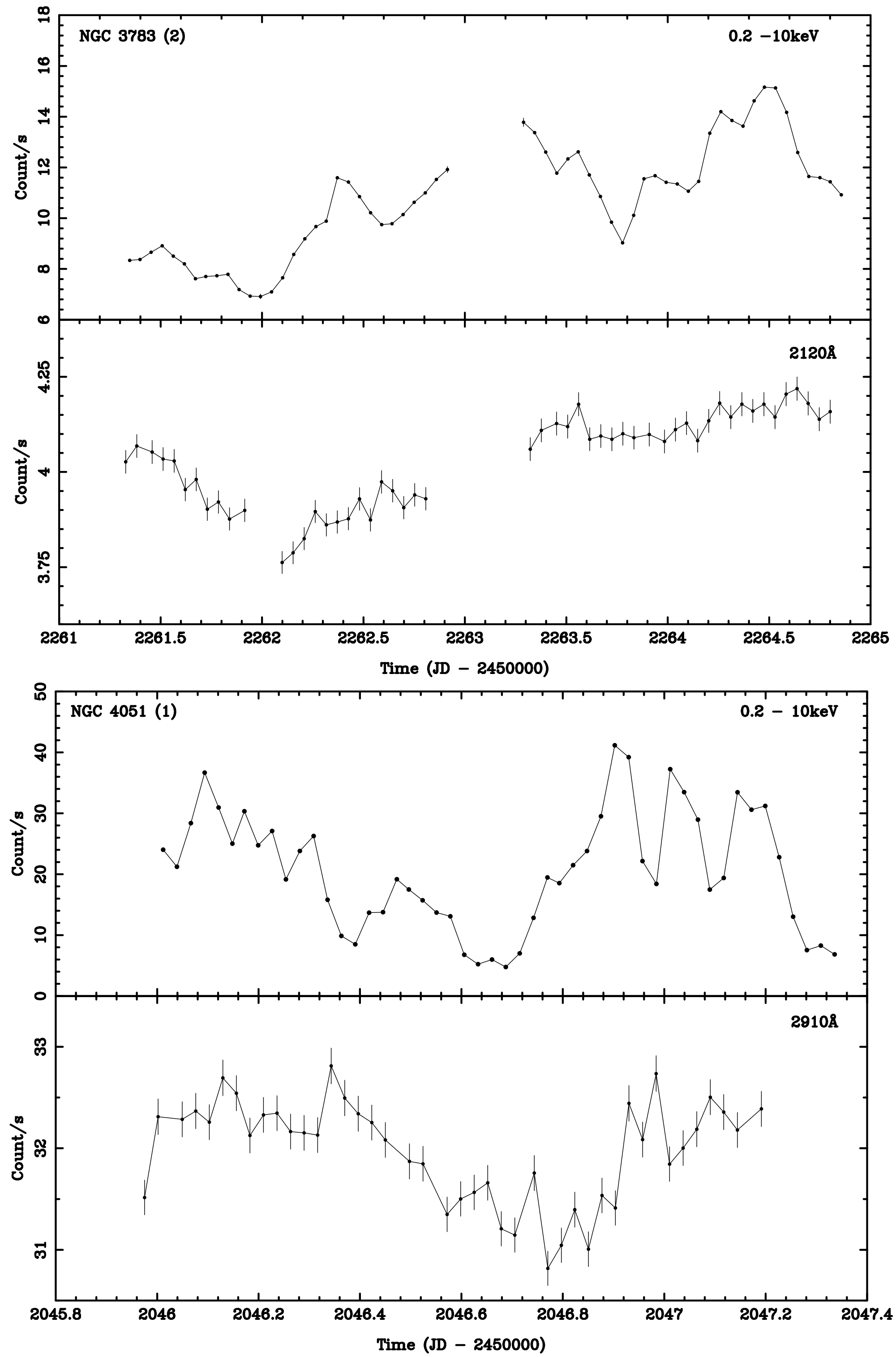


Figure 5.2: continued

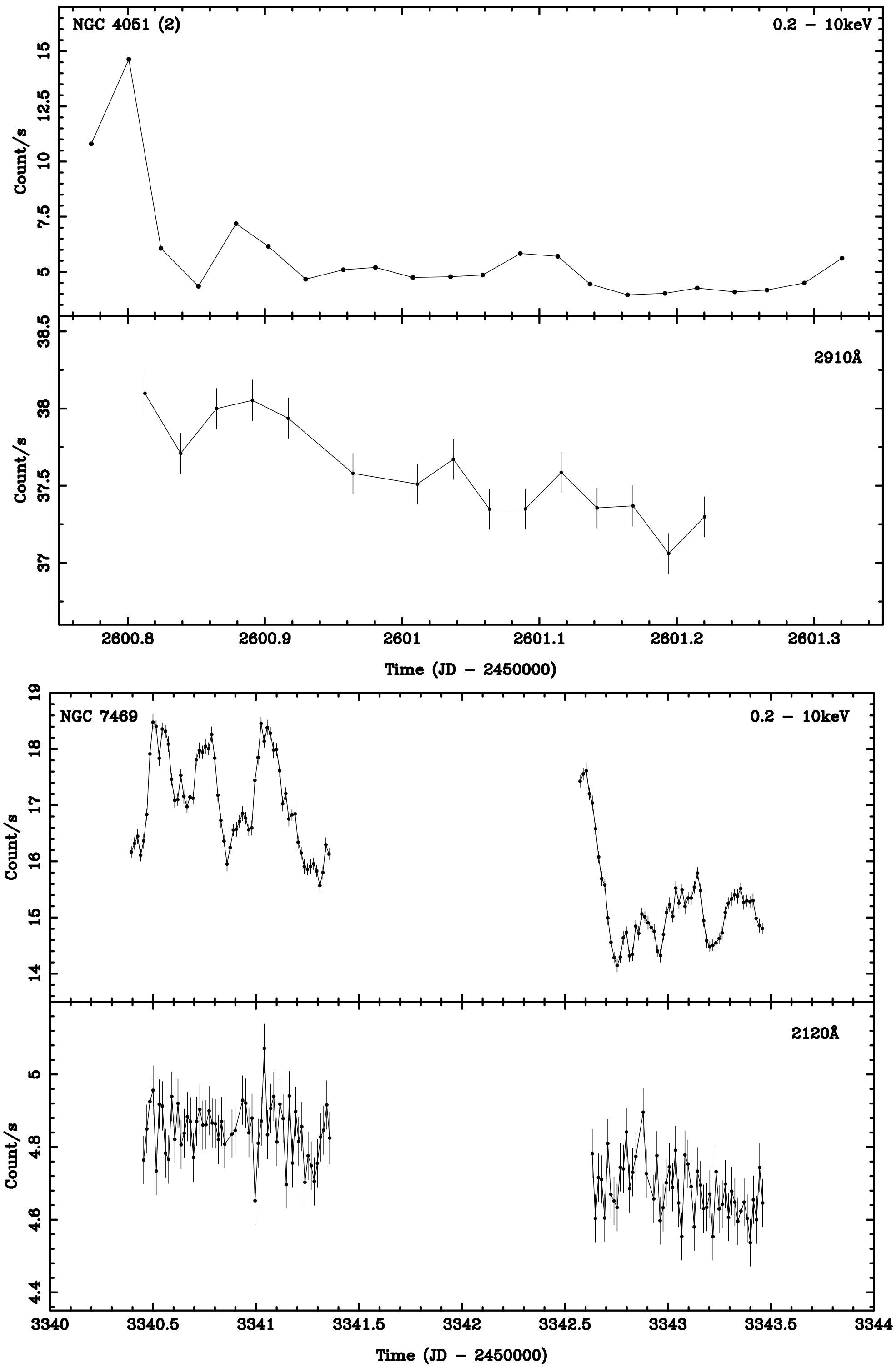


Figure 5.2: continued

5.3 Testing the method

5.3.1 Monte Carlo tests

In order to test the robustness of the light curve extraction method (outlined above) to PSF losses and host galaxy contamination, random data were simulated and subjected to the same aperture photometry procedure. As the spacecraft pointing can change during an observation by as much as a few arcseconds, causing the target source to drift several pixels across the image, this effect was also included in the simulations. The simulations were designed to match the MCG$-$6-30-15 data, chosen because it is the longest data set and shows the most contamination from its host galaxy. The simulation procedure, kindly performed by Dr Simon Vaughan, was as follows.

The 'true' source position on a 224×224 pixel detector grid was defined as (x_0, y_0) based on the average detector position of MCG$-$6-30-15 in the real OM images. The effect of random pointing changes was simulated by giving a random offset to the precise position of the point source in each image. A point source was then added to the image at the new position $(x_0 + \Delta x, y_0 + \Delta y)$ using the PSF model obtained from the latest calibration files. In practice, a differential PSF model was calculated from the cumulative PSF model listed in the calibration files, and linearly interpolated to produce a function sampled with 0.1 pixel resolution out to twelve pixels. This was used to define the image of the point source from a smooth PSF projected onto a relatively coarse pixel grid. The host galaxy emission was modelled as an ellipsoid with an exponential surface density profile, with spatial parameters and normalisation (relative to the point source) taken from the spatial analysis of Arévalo *et al.* (2005), kindly supplied by Dr Arévalo. The source image was then normalised to match the typical counts per image from the observations, and a typical (flat) background level (≈ 3.3 counts per pixel) was added. The counts per pixel were then randomly drawn from a Poisson distribution to simulate random counting statistics. The simulated images were then subjected to aperture photometry. Specifically, repeated centroiding was used to find the exact source position in each image, then counts were accumulated from a circular region and the PSF1 correction was applied exactly as for the real data. The effects of the other corrections, namely CL, TDS and DT, were not included. The latter two corrections are simply constant factors determined for each image, and the CL is well calibrated for aperture photometry (assuming the PSF1 correlation is correct).

The simulations were designed to reproduce the effects of aperture photometry on a constant point source, the position of which is slightly different for each image, superposed on a realistic host galaxy image and flat constant background level, all subjected to Poisson variations. The end result of 200 simulations was a time series that could be fitted with a constant to test for spurious variability. Using random position offsets $(\Delta x, \Delta y)$ drawn from a Gaussian of mean zero and standard deviation 0.2 pixels, the result was a fit statistic of $\chi^2 = 211.29$ with 199 degrees of freedom (dof), a perfectly acceptable fit ($p = 0.262$), indicating no statistically significant variability was present given 200 images. A longer simulation using 5000 images did show significant variability ($\chi^2 = 5432.56$ for 4999 dof, $p = 1.2 \times 10^{-5}$) but only at the 0.4% level, which is too small to be detected in less than 200 images. This excess variability is also far smaller than the measured amplitudes in any of the observations (see below).

Allowing for a 10 pixel drift in both x and y directions over the course of 200 images, comparable to the true drift during the MCG$-$6-30-15 observation, plus a random offset for each image, gave similar results: $\chi^2 = 205.19$ for 199 dof ($p = 0.39$). There was no statistically significant correlation ($r = 0.044,\ p = 0.53$) between the measured flux and the offset between the centroid and the stationary 'first guess' position (x_0, y_0) given to the photometry algorithm, indicating that the photometry method is not biased by even fairly large changes in spacecraft pointing. The same results were obtained when the host galaxy was not included in the simulations, proving that contamination by host galaxy light within the source aperture does not affect the variability. Changing the random offsets $(\Delta x, \Delta y)$ to have a standard deviation of 0.5 pixels in each direction also gave entirely consistent results.

5.3.2 Empirical tests using field stars

Satisfied that the method performed well using simulated data, it was then tested on actual data from the OM. For several of the sources in the sample (Ark 564, Mrk 766, MCG$-$6-30-15), stars were visible in the field of view. To prove that the method does not introduce spurious variability, the process was tested on a selection of these field stars. The field around Ark 564 contains many objects. For the field star tests, two objects of similar brightness to Ark 564 were chosen which were far enough removed

Table 5.3: Field sources used for testing

Field target	RA (degrees)	Dec (degrees)	p
(1)	(2)	(3)	(4)
Ark 564	340.593	29.725	0.02
	340.701	29.756	0.70
MCG−6-30-15	203.966	-34.137	0.21
Mrk 766	184.611	29.813	0.80
NGC 4051	180.786	44.539	0.21
	180.793	44.541	0.56
	180.763	44.541	0.77

from other objects in the field to allow accurate light curves to be extracted. The other fields examined here contained few objects other than the target, so all field stars (unless contaminated by being too close to the target) were tested.

Table 5.3 summarises the results of these tests. A p value of $\gg 0.01$ indicates that a light curve is non-variable. Three of the four stars tested proved to be non-variable proving that the method used here does not introduce spurious variability. One of the stars tested in the Ark 564 field has a p value of 0.02 implying possible variability in the source. Cross correlation analysis (CCF, see later) between this object and the Ark 564 light curve was performed to ensure the detected variability was not a systematic problem also affecting Ark 564. No correlations were found between the two objects. It is suggested that the star is intrinsically variable and does not cause a problem with the method.

As another test of whether or not the method is robust, light curves for extended sources in the NGC 4051 host galaxy were extracted. These will not vary on the timescales probed here so a non-varying light curve should definitely be produced. The light curves extracted did indeed prove to be non-variable confirming that the method does not introduce spurious variability.

5.4 Results

To determine whether or not the optical light curve is variable, a χ^2 test was performed on each light curve. The resulting p values are given in table 5.1. Values of $p < 0.01$ indicate variability at the 99% confidence level or greater and values between 0.01 and 0.05 are possible detections of variability ($95-99\%$ confidence). Sources with $p > 0.05$ are deemed to be non-variable. Variability was detected in the OM light curves of four of the eight objects.

Also shown in table 5.1 is the optical and X-ray rms variability amplitude for each source. The typical error in these values is approximately 1% so, depending on the brightness of the source, an object with rms amplitude $> 1\%$ should be detectable as varying. The optical rms values shown here are probably underestimated as they include the total flux of the source with no host galaxy subtraction. For the four sources

Table 5.4: Root mean square variability

Object	OM rms	X-ray rms
	$(10^{42}\,\mathrm{erg\,s^{-1}})$	$(10^{42}\,\mathrm{erg\,s^{-1}})$
(1)	(2)	(3)
MCG−6-30-15	0.2	35
NGC 3783 (2)	3.3	51
NGC 4051 (1)	0.08	2.9
NGC 4051 (2)	0.5	0.7
NGC 7469	7.3	30

that show significant variability in both optical and X-ray bands, the rms variability amplitude was calculated in luminosity units, tabulated in table 5.4.

5.4.1 Cross correlation analysis

A standard tool for testing and measuring correlations between two time series is the cross correlation function (CCF), which gives the degree of correlation between the two series as one series is shifted backwards and forwards in time relative to the other. Unfortunately, when the data are not exactly evenly sampled one cannot apply the standard, 'textbook' formulae for estimating the CCF (Box and Jenkins 1976; Priestley 1981; Brockwell and Davis 1996). There are two reasons why the present data are not evenly sampled: the OM images are rather evenly spaced in time but there are a few gaps, and observations spanning more than one *XMM-Newton* revolution will be split by the orbital gap. Suggested solutions to the problem of uneven sampling include interpolating the data onto an even grid, or calculating the correlation using pairs of data points that have a lag falling within a finite range (i.e. $\tau_{ij} = t_x(i) - t_y(j)$ falls in the interval $[\tau, \tau + \Delta\tau]$). Specific algorithms that have been used in AGN monitoring include the interpolated cross correlation function (ICCF; Gaskell and Peterson 1987), the discrete correlation function (DCF; Edelson and Krolik 1988) and the z-transformed discrete correlation function (ZDCF; Alexander 1997). All three of these were used in the present analysis and the results were found to be consistent. For each potential lag, τ, the basic correlation function is:

$$CCF(\tau) = \frac{\mathbf{x.y}}{\sqrt{(\mathbf{x.x} - \mathbf{\Delta x.\Delta x})(\mathbf{y.y} - \mathbf{\Delta y.\Delta y})}} \tag{5.7}$$

where $\mathbf{x} = (\mathbf{x_i} - \mathbf{\overline{x}})$ for the i points in one light curve and likewise for $\mathbf{y}$ for the other light curve. $\mathbf{\Delta x}$ and $\mathbf{\Delta y}$ are the measurement errors on the two light curves. The three different algorithms use different methods to determine the points that contribute to each lag.

A second problem that plagues CCF estimation from AGN data is that the data are often non-stationary, meaning that the mean and/or variance of a short light curve will change over time, due to variations on timescales longer than the length of the observation. Following the advice of Welsh (1999), two precautions were used to mitigate the side-effects of non-stationarity. First, a linear function was fitted to the data (using ordinary least-squares) and removed to suppress any long timescale variation, and secondly, the mean and variance were calculated 'locally', using only the data contributing to a particular lag. Following another suggestion in Welsh (1999), the correlation coefficient was computed using both the raw data and the ranked data (each data point being correlated was replaced with its rank). In principle using the ranked data should improve sensitivity to non-linear correlations and make the statistic more robust to outliers, compared to using the raw data. In practice the two were found to be in close agreement.

Confidence limits on the CCF were computed using Bartlett's formula (Bartlett, 1955) as an approximate but simple test of the significance of any correlations. In particular, the standard deviation of the distribution of CCF points around an expected value of zero was computed using equation 11.1.9 of Box and Jenkins (1976). This assumes that, in reality, the two series are not correlated, and can be used to test against this hypothesis. Under the assumption of no cross-correlation (i.e. the null hypothesis that the expectation value of the CCF is zero), Bartlett's formula may be written as:

$$\sigma^2\{\mathrm{CCF}(\tau_j)\} = \frac{1}{N_{\mathrm{pair}}} \sum_i \mathrm{ACF}_1(\tau_i)\mathrm{ACF}_2(\tau_i) \tag{5.8}$$

where the two ACF functions are the empirical auto-correlation functions and the summation is carried out over all measured lags. Here, N_{pair} is the number of pairs of data points contributing at the lag τ_j.

The 95% confidence region was calculated as $\pm 1.96\sigma$. The confidence region for the correlation coefficient at a given lag scales with $N_{\mathrm{pair}}^{-1/2}$. The bounds are smallest at small time delays, where the two time series overlap the most, and increase towards the edges of the CCF. The CCFs were examined over a range of lags spanning, in either direction, half the observation duration. At longer lags, the number of pairs that contribute to the CCF calculation rapidly decreases leading to highly uncertain CCF estimates. It should be noted that Bartlett's formula is strictly only valid when the number of data points, N, is large and the data are stationary. As the X-ray/optical data do not satisfy either condition well, the confidence limits should be considered as rough estimates of the size of the true confidence region. Bartlett's formula is discussed in most standard references for time series analysis including: Jenkins and Watts 1968 sect 8.2.1; Box and Jenkins 1976, sect 11.1.3; Priestley 1981, sect 9.5; Bendat and Piersol 1986, sect 8.4; Kendall and Ord 1955, sect 11.3 or Brockwell and Davis 1996, sect 7.3.4.

Previous papers discussing inter-band correlations in AGN variability have made use of Monte Carlo simulations to test the significance of any possible correlations. In principle the Monte Carlo approach should be robust but in practice it does rely on the power spectra of the two light curves being well understood in order to produce realistic simulations. Technically the CCF is not a 'pivotal' statistic, meaning it is sensitive to 'uninteresting' parameters such as the power spectral shape, and so simulations must be made as carefully as possible. For the present analysis it was not possible to obtain reliable power spectra from the limited OM data, and so Bartlett's formula (which uses the empirical auto-correlation function) was used as a guide to significance. It is also worth noting that the results of Monte Carlo tests are sensitive to which of the pair of light curves are randomised – whether the X-ray, or optical, or both datasets are randomised can substantially affect the apparent significance, and as yet there is no accepted preference. This choice is not required for Bartlett's formula to be used.

Figure 5.3 shows the CCF plots for the objects studied which displayed X-ray and optical variability. A positive lag in the plots implies the X-ray variations lead those in the optical and vice versa for a negative lag. The 95% and 99% confidence limits are indicated by dotted and dot-dashed lines respectively. In the case of NGC 4051 (1, Rev. 263) and NGC 3783 the 95% confidence region is sufficiently broad that these data do not impose useful constraints on the possibility of a correlation. The error bounds, as computed using the Bartlett formula, were large in these cases because the empirical ACFs for both optical and X-ray time series were very broad. Particularly in the case of NGC 3783, the X-ray power spectrum over the timescales studied here is rather steep, leading to smooth light curves and a broad ACF and hence a large error in the Bartlett formula. This accounts for the fact that realisations of processes with steeper power spectra (or equivalently, broader ACFs) are more 'weakly non-stationary'. In the cases of NGC 4051 (2, Rev. 541), NGC 7469 and MCG$-$6-30-15 the CCF is sufficiently well constrained to exclude a strong correlation, i.e. 95% confidence limit of $|\mathrm{CCF}(\tau)| < 0.5$, over a plausible range of τ.

Some previous studies (e.g. Nandra *et al.* 2000) have detected a correlation between the optical light curve and X-ray photon index but no correlation between optical and X-ray light curves. In light of this the optical light curves of the objects studied here were cross correlated with the $0.7 - 2/2 - 10\,\mathrm{keV}$ hardness ratio (which should track the photon index) but the cross correlation functions were found to be very similar to the ones between X-ray and optical flux.

5.5 Discussion

5.5.1 Summary of results

Using data from the OM and X-ray telescopes aboard *XMM-Newton*, a sample of eight objects have been examined to look for evidence of reprocessing between wavebands. Optical variability on short timescales was detected at 99% confidence in four out of eight objects and rms variability amplitudes in all cases show that the X-ray variability is stronger than the optical variability in both fractional terms and absolute luminosity units. In three out of the four objects showing detectable optical variability, the CCF is sufficiently well constrained to exclude a strong X-ray/optical correlation. While this suggests that reprocessing on timescales of < 1 day is not a dominant mechanism, there may be several other reasons why a correlation is not detected. Light travel time effects can smear out variability on short time scales so the size of the optical region may prevent the correlation being detectable. It is also possible that only some of the optical emission is due to reprocessing of X-rays and so the variability due to the X-ray emission could be masked. Other issues such as the uncertain geometry of the emitting regions can also have an effect on the correlation: is the X-ray emission isotropic?, is all of the accretion disk visible to the X-ray source? etc. As well as these physical explanations it must also be kept in mind the caveats that were discussed above regarding the

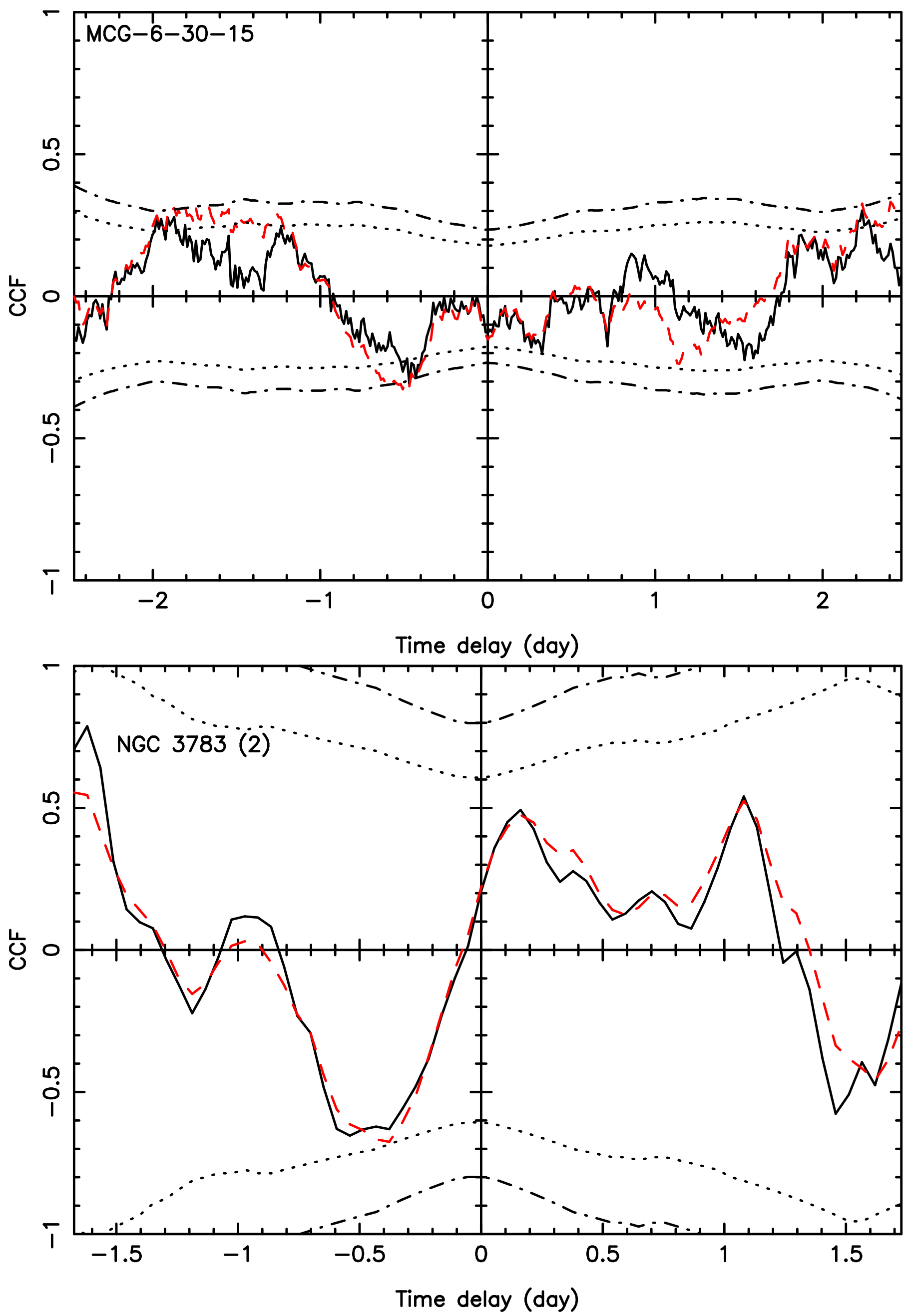


Figure 5.3: CCF plots of sources detected as being optically variable. The DCF is plotted in black and the ICCF is the red dashed line. The dotted and dot-dashed lines represent the 95 and 99% confidence limits respectively.

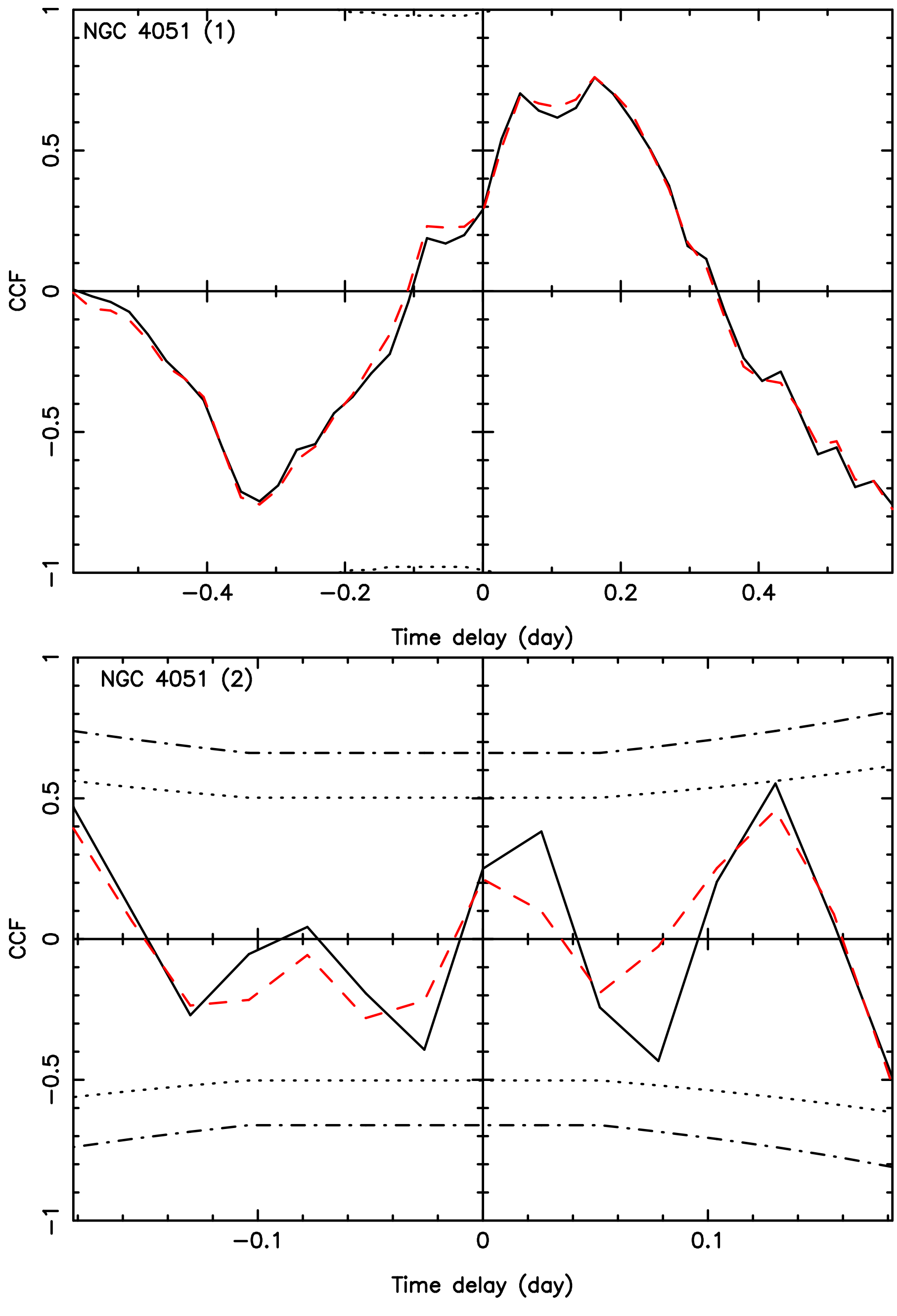


Figure 5.3: continued

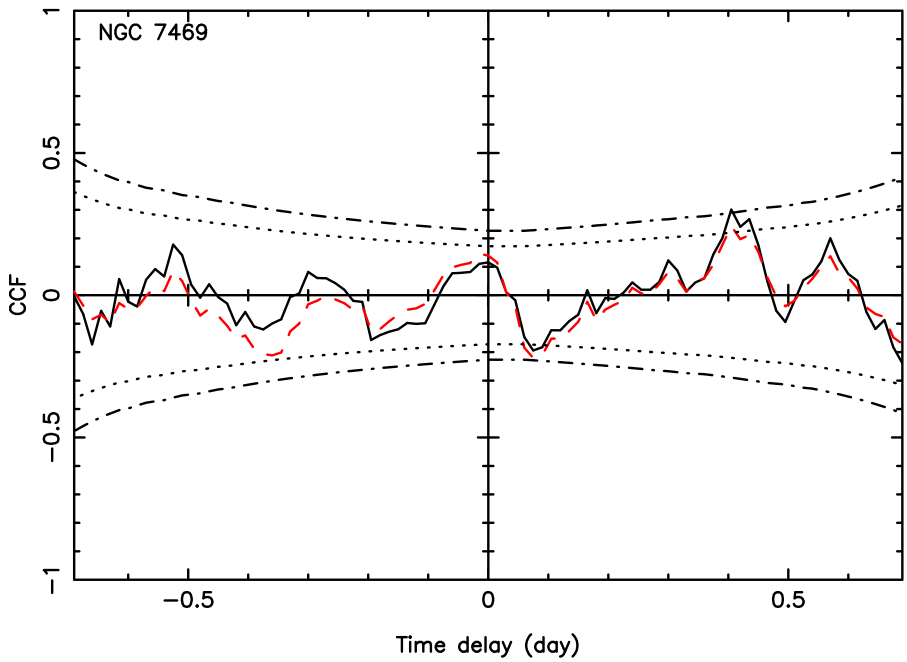


Figure 5.3: continued

sparse sampling of the light curves and the use of Bartlett's formula as a rough estimate of the significance of the CCF.

5.5.2 Comparison with previous studies

Observations using *XMM-Newton* data

The observation of NGC 4051 in Revolution 0263 was also investigated by Mason *et al.* (2002) for inter-band correlations. They supplemented the *XMM-Newton* data with X-ray observations from *RXTE* and their CCF analysis covers positive lags up to two days. Over the region that their CCF plot overlaps with that presented here ($0 - 0.6$ days), the results are consistent. A peak in the CCF around 0.2 days is observed with a value of

0.75, similar to their peak of ≈ 0.6. Using simulations of uncorrelated red-noise light curves they place a confidence of 85% in the correlation. The peak detected here lies between 68.3 and 95% confidence levels, certainly not significant at 99%. However, as mentioned in section 5.4.1, the confidence regions in this case are not well enough constrained to comment on the correlation.

The OM observation of MCG$-$6-30-15 was previously analysed by Arévalo *et al.* (2005), who reported a positive correlation with $DCF_{max} = 0.82$ at $160\,$ks (around 1.8 days) with the UV leading the X-ray data. Using Monte Carlo simulations they estimated a confidence limit of 98.5% (i.e. a p-value of 0.015). By contrast, the analysis presented here found a much lower CCF peak (0.35) which was not considered a significant detection of correlation.

There are several differences between the analysis of Arévalo *et al.* (2005) and that of the present analysis. In terms of the light curve extraction, Arévalo *et al.* (2005) performed photometry on attitude corrected images by fitting an empirical PSF model to the source image, while simultaneously modelling the host galaxy emission. The present analysis used aperture photometry on (detector space) images based on the SAS processing pipeline tasks, including corrections for coincidence loss, deadtime and the counts/frame dependence of the PSF model, none of which were included by Arévalo *et al.* (2005). These differences in extraction technique explain the apparent difference between the two reductions of the same data. Arévalo *et al.* (2005) also used coarser binning of the light curves, but applying the same binning to light curves produced here did not substantially decrease the discrepancy. As shown in section 5.3, the aperture photometry method employed in this analysis is robust to host galaxy presence and to pointing changes but, unlike the analysis of Arévalo *et al.* (2005), the present method includes all the known time-dependent flux corrections that should be applied to OM data. Therefore, in principle at least, the reduction presented here should be the more reliable.

In practice most of the difference between the correlation strengths is due to the more conservative approach employed in the present analysis for calculating the CCF. Arévalo *et al.* (2005) presented the DCF after normalising by the noise-subtracted variances (i.e. the variances used in the denominator of the correlation coefficient formula had the mean square error subtracted). The present analysis used ICCF, DCF and ZDCF methods without the noise-subtraction term, employing linear and rank-order correlation coefficients, the latter of which should be more robust to outliers. Including the noise-subtraction term increased the CCF value by ≈ 0.2 at the claimed lag, but this only affects the absolute value of the CCF and not its significance (since the effect should be calibrated out during the significance test).

A second difference is that, in the present analysis the time series were 'detrended' before calculating the CCF by subtracting off a linear function. As recommended by Welsh (1999) this helps remove any bias caused by very long timescale variations manifested as quasi-linear trends over the data that can lead to spuriously high correlation coefficients. Applying the detrending reduced the CCF by ≈ 0.2 at the claimed lag. If the correlation at ≈ 1.8 days was robust it should not be affected by removal of longer timescale (≈ 5 day) variations.

Observations from other telescopes

Several other sources discussed here have also been studied previously: Ark 564, NGC 4051 and NGC 7469.

Ark 564 was observed by Papadakis *et al.* (2000) over a period of 21 days at the Skinakas Observatory in the optical waveband and simultaneously in X-rays with *ASCA*. With once-nightly optical observations, Papadakis *et al.* (2000) probed slightly longer timescales than examined here but, in line with the trend reported here, they found no significant correlations between the two bands. Shemmer *et al.* (2001) examined Ark 564 over a longer period (50 days) in UV with the *Hubble Space Telescope* (*HST*) and X-ray with *RXTE* and *ASCA*. They found evidence of UV emission following X-ray by ≈ 0.4 days.

Shemmer *et al.* (2003) studied NGC 4051 and found evidence for a correlation between X-ray and optical bands with a lag of less than one day. Their campaign covered a period of three months with approximately daily observations and so probed a longer timescale than examined here. NGC 4051 was also investigated by Peterson *et al.* (2000). On short timescales they found no evidence for correlations. Having smoothed the data in 30 day bins however, a correlation consistent with zero lag is detected.

A 30 day observation of NGC 7469 by Nandra *et al.* (1998) found that at least two variability mechanisms are required to describe their result. Cross correlation of UV and X-ray light curves gave evidence for a lag of around four days. This lag explained the differing times of maxima in the two light curves but failed to describe how the minima in the light curves occurred simultaneously.

As none of the studies described above probe the same short timescales examined here it is difficult to make a comparison between them and the results described in this chapter. An overall picture seems to be emerging of no correlation between X-ray and optical emission over sub-day periods but significant correlations appearing on much longer timescales.

5.5.3 Implications for AGN variability

In line with other work, it is found here that X-ray rms variability exceeds that in the optical in all cases. However, in the objects with well constrained CCF (MCG−6-30-15, NGC 4051 (2, Rev. 0541) and NGC 7469), the absence of a strong correlation implies the majority of the variance in the optical band on short timescales is uncorrelated with the short timescale X-ray variations. Papadakis *et al.* (2000) suggest that this behaviour could be explained if only a small section of the accretion disk is visible to the X-rays or if the X-ray emission is anisotropic. Two of the observations showing optical variability, NGC 3783 and NGC 4051 (1, Rev. 0263) have poorly constrained CCF confidence regions and so a firm conclusion on these objects can not be drawn.

In terms of the overall picture emerging from variability studies, this work solidifies earlier findings (Edelson *et al.* 2000; Uttley *et al.* 2003) that, on short timescales, X-ray and optical variations in Seyfert 1 galaxies are uncorrelated. Many papers (see section

5.5.2) find some evidence for correlations on longer timescales. It seems that rapid X-ray and optical variations in Seyfert 1 galaxies occur through separate mechanisms but long-term variations are somehow connected. As discussed earlier this may be due to short timescale variability being smoothed by light travel time effects. More complex models incorporating combinations of emission mechanisms may be required to properly explain the results of various studies.

Intensive observations with *XMM-Newton* are not suitable for probing longer timescale variations. A current alternative to *XMM-Newton* is the *Swift Gamma Ray Burst Explorer*. *Swift* carries co-aligned X-ray (XRT; Burrows *et al.* 2005) and UV/optical (UVOT; Roming *et al.* 2005) telescopes covering essentially the same bandpass as *XMM-Newton*, but is in a low-Earth orbit making it better suited to regular 'snapshot' monitoring of sources over weeks or months. The combination of *XMM-Newton* and *Swift* is therefore capable of probing X-ray/optical correlations with unprecedented detail on all timescales from hours to weeks.

Another promising possibility is *ASTROSAT*, an Indian Space Research Organisation satellite, scheduled for launch in April 2009. Like *Swift* and *XMM-Newton*, *ASTROSAT* will carry both a UV/optical telescope and several X-ray detectors. One of the main aims of the Large Area Xenon Proportional Counter (LAXPC) instruments onboard *ASTROSAT* is to take very long exposure observations of AGN to look for variability properties. Combined with the optical observations, this has the potential to extend the work presented here and reveal more detail on the connection between optical and X-ray emission in AGN.

Chapter 6

Cosmic variance and bias in X-ray selected quasars from 2XMMp

6.1 Introduction

The cosmological principle states that, on large scales, the Universe is homogeneous and isotropic. On smaller scales however, obvious stucture is observed (e.g. galaxies, clusters, walls and voids). For this structure to exist there must have been some initial deviation from homogeneity. The Universe is therefore believed to have developed a power spectrum of inhomogeneities shortly after the Big Bang. Overdense inhomogeneities grew by attracting neighbouring matter through gravitation. If the size of the inhomogeneities grew to bigger than the Jeans mass, the mass could collapse to form a gravitationally bound object. In this way, structure would form.

The origin of these inhomogeneities is still uncertain but there are two popular theories. The first is that they are the result of a period of inflation in the very early Universe (e.g. Guth and Pi 1982). During this period, the Universe underwent rapid expansion and quantum fluctuations, which would normally go undetected, would be caught in the expansion and stretched. After inflation this would leave a series of irregularities on different scales. A second possibility is that the inhomogeneities are topological defects (e.g. cosmic strings, magnetic monopoles) resulting from phase transitions in the early Universe (e.g. Battye *et al.* 1998). It is currently unclear which, if either, of these mechanisms is correct although inflation has the advantage that it can also solve other problems of cosmology such as the horizon problem.

Once formed, the inhomogeneities grew at a rate dependent on size leading to the structure seen today. The constituents of the Universe (e.g. baryons, dark matter) also have some small scale effects on the growth of structure. The changes in the power spectrum can be tracked forward from the initial spectrum to get the expected power spectrum today. In most models the expected matter distribution is Gaussian. The actual distribution is broadened because observable objects (e.g. galaxies, AGN) do not exactly trace the underlying dark matter. An important product of cosmic variance studies is the bias parameter which measures the ratio of the variance of observable matter to that of the underlying mass. Bias characterises how the observable matter traces the underlying mass e.g.:

$$b_{AGN} = \frac{\sigma_{AGN}}{\sigma_{mass}} \tag{6.1}$$

where b_{AGN} is the AGN bias, σ_{AGN} is the AGN variance and σ_{mass} is the total mass variance. Knowledge of the bias parameter aids in tracing the origin of structure. It can also be used to determine the mass of the dark matter halos in which AGN reside (Sheth *et al.* 2001).

A popular measure of source density is the log N - log S plot: the number of sources in terms of flux density. These plots are consistently fit by a broken power law in different samples but, while the slope of these curves is consistent, the normalisation changes. If objects were uniformly distributed across the sky there would not be a difference in the normalisation. The difference is known as cosmic variance. Cosmic variance is reflected in a field to field variation in the number of sources that is greater than that expected from Poisson noise.

In this chapter, fields from the serendipitous X-ray source catalogue 2XMMp are used to investigate cosmic variance and determine the AGN bias parameter. The cosmology adopted throughout is the *WMAP* (*Wilkinson Microwave Anisotropy Probe*) concordance cosmology: $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Spergel *et al.* 2003).

6.2 Previous studies

The clustering properties of sources have been studied for many years. The most common method of quantifying clustering is the two point correlation function, $\xi(r)$, which measures the excess probability of finding a pair of sources separated by distance r compared to the probability of finding such a pair in a randomly distributed field. This relation can normally be expressed as a power law:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \tag{6.2}$$

where r_0 is the correlation length (the scale at which $\xi(r) = 1$). To calculate $\xi(r)$, the redshifts of the sources are needed. If redshift values are not available, an alternative function can be calculated - the angular correlation function, $w(\theta)$. This can again be expressed as a power law:

$$w(\theta) = \left(\frac{\theta}{\theta_0}\right)^{-\gamma} \tag{6.3}$$

where θ is the angular separation of two sources on the sky. This value can then be converted to an estimate of the two point correlation function using the Limber equation (Limber 1953).

To determine $\xi(r)$ or $w(\theta)$ at various separations, simulations of fields are created that have the correct flux distribution for the observed fields but with randomly distributed sources. An estimator such as the one by Landy and Szalay (1993) is then used to determine $\xi(r)$ or $w(\theta)$ for different separations, e.g.:

$$w(\theta) = \frac{DD - 2DR + RR}{RR} \tag{6.4}$$

where DD, DR and RR are numbers of data-data, data-random and random-random pairs separated by θ respectively. The majority of the following studies have made use of such correlation functions to describe clustering in their samples. The Wiener-Khintchine theorem shows that the correlation function is the Fourier transform of the power spectrum which is used in this chapter to investigate clustering.

Early clustering studies concentrated on galaxies as these were easily identifiable in optical surveys. Tucker *et al.* (1997) found $r_0 = 6.28 \pm 0.27\,\mathrm{h^{-1}\,Mpc}$ and $\gamma = 1.52 \pm 0.03$ for galaxies in the Las Campanas Redshift Survey for separations up to $16.4\,\mathrm{h^{-1}\,Mpc}$. At large separations ($r > 50\,\mathrm{h^{-1}\,Mpc}$), Tucker *et al.* (1997) found that $\xi(r)$ dropped to around zero suggesting a high level of uniformity on large scales. Ratcliffe *et al.* (1998) found similar results for ≈ 2500 galaxies in the Durham/UKST Galaxy Redshift Survey ($r_0 = 6.8 \pm 0.3\,\mathrm{h^{-1}\,Mpc}$ and $\gamma = 1.25 \pm 0.6$).

More recently, since efficient star-quasar separation became possible, large optical surveys of AGN clustering have been conducted. The results of the 2dF Quasar Redshift Survey (2QZ, Croom *et al.* 2005) showed that clustering of AGN was comparable to that of local optically selected galaxies ($r_0 = 5.48 \pm^{0.42}_{0.48}\,\mathrm{h^{-1}\,Mpc}$ and $\gamma = 1.2 \pm 0.1$). Croom *et al.* (2005) calculated a quasar bias of $b_Q = 2.02 \pm 0.07$ at a mean redshift of 1.35. Croom *et al.* (2005) also found evidence that quasar bias is a strong function of redshift with $b_Q = 1.13 \pm 0.18$ at $z = 0.53$ increasing to $b_Q = 4.24 \pm 0.53$ at

$z = 2.48$ but found no evidence for evolution with luminosity. The Asiago-ESO/RASS QSO Survey (AERQS) investigated clustering of local X-ray selected AGN to determine a 'zero point' for future investigations of the evolution of the correlation function (Grazian *et al.* 2004). Grazian *et al.* (2004) found $r_0 = 8.6 \pm 2\,\mathrm{h}^{-1}\,\mathrm{Mpc}$ with $\gamma = 1.56$, comparable to local elliptical galaxies and EROs (Extremely Red Objects). Comparing their local AGN results to those of an earlier analysis of 2QZ (Croom *et al.* 2001), Grazian *et al.* (2004) also found evidence for evolution of AGN clustering with redshift. Wake *et al.* (2004) used SDSS DR1 to compile a sample of 13605 AGN at $z < 0.2$ and found clustering comparable to that of galaxies. Myers *et al.* (2007) used approximately 300000 photometrically selected QSOs from SDSS DR4 to determine a quasar bias of $b_Q = 2.41 \pm 0.08$ at $z = 1.40$ and also found evidence for increasing quasar bias with redshift.

Out to $z \leq 2.2$, optical AGN selection methods generally use the UV excess technique so optical QSO studies sample mainly unobscured, type 1 AGN. Obscured AGN are missed in these selections because the obscuring material absorbs UV radiation and re-emits it at longer wavelengths. X-ray surveys have the advantage that they can probe the obscured, type 2 AGN regime. There have been several recent studies of the clustering of AGN detected in X-rays. Gilli *et al.* (2005) examined clustering of AGN in the Chandra Deep Fields - CDF North ($0.13\mathrm{deg}^2$ area) and CDF South ($0.1\mathrm{deg}^2$ area). A significant difference was found between the clustering amplitudes in the two fields ($r_0 = 10.3 \pm 1.7\,\mathrm{h}^{-1}\,\mathrm{Mpc}$ for CDF-S and $5.5 \pm 0.6\,\mathrm{h}^{-1}\,\mathrm{Mpc}$ for CDF-N) although the slope of the distribution was found to be consistent between the fields (γ=1.33$\pm$0.14 for

CDF-S and 1.50±0.12 for CDF-N) and also consistent with that of optically selected QSOs. The difference in amplitude was attributed to cosmic variance because the two fields, although deep, cover a small area. The CDF-S in particular showed evidence of two structures in the field indicated by two spikes in the redshift distribution. No statistically significant difference in clustering between hard and soft sources or type 1 and type 2 AGN was found. Variance between Chandra fields was also reported by Yang *et al.* (2003) in their study of $0.4\mathrm{deg}^2$ of the Lockman Hole region. In contrast to the CDF study (Gilli *et al.* 2005), Yang *et al.* (2003) detected a difference between hard ($2-8\,\mathrm{keV}$) and soft ($0.4-2\,\mathrm{keV}$) band clustering. With γ fixed at 1.8, amplitudes of $\theta_0 = 40''\pm 11$ and $\theta_0 = 4''\pm 2$ were determined in the hard and soft band respectively.

Larger surveys such as XMM-COSMOS (Hasinger *et al.* 2007), AXIS (Carrera *et al.* 2007), XMM-LSS (Pierre *et al.* 2004) and ChaMP (Kim *et al.* 2004a) have also investigated AGN clustering. The XMM-COSMOS survey covered an area of $2\mathrm{deg}^2$ and found AGN bias of $1.5-4$ (Miyaji *et al.* 2007). With γ fixed at 1.8, amplitudes of $r_0 = 9.4 \pm 0.8\,\mathrm{h}^{-1}\,\mathrm{Mpc}$, $4.6\pm^{1.7}_{2.5}\,\mathrm{h}^{-1}\,\mathrm{Mpc}$ and $8.4\pm^{1.7}_{2.0}\,\mathrm{h}^{-1}\,\mathrm{Mpc}$ were found for soft ($0.5-2\,\mathrm{keV}$), medium ($2-4.5\,\mathrm{keV}$) and ultra hard ($4.5-10\,\mathrm{keV}$) bands respectively. AXIS is another *XMM-Newton* survey, covering $4.8\mathrm{deg}^2$ with 34 high Galactic latitude ($|b| \geq 20°$) fields. Clustering in the AXIS fields was detected in the soft ($0.5-2\,\mathrm{keV}$, $\theta_0 = 19''\pm^{7}_{8}$, $\gamma = 1.2\pm^{0.3}_{0.2}$) and XID ($0.5-4.5\,\mathrm{keV}$, $\theta_0 = 19''\pm^{7}_{8}$, $\gamma = 1.3\pm^{0.4}_{0.3}$) bands but not in the hard ($2-10\,\mathrm{keV}$) and ultra hard ($4.5-7.5\,\mathrm{keV}$) bands (Carrera *et al.* 2007). The $4.2\mathrm{deg}^2$ *XMM-Newton* LSS (Large Scale Structure) Survey also detected clustering in the soft band ($0.5-2\,\mathrm{keV}$, $\theta_0 = 6.3'' \pm 3$, $\gamma = 2.2 \pm 0.2$) but no clustering in the $2-10\,\mathrm{keV}$ hard band (Gandhi *et al.* 2006). Gandhi *et al.* (2006) used the Limber equation to determine a correlation length of $r_0 = 6 \pm 3\,\mathrm{h}^{-1}\,\mathrm{Mpc}$, similar to that found by studies of local optically selected galaxies and optically selected QSOs at $z \approx 1$. In contrast to the above studies, Kim *et al.* (2004b), found no significant evidence for field to field variation in source density in the ChaMP survey covering 14deg^2.

There is therefore a fairly consistent view from both optical and X-ray studies that AGN clustering is observable and similar to that of local galaxies. There are also several studies suggesting that AGN bias evolves with redshift.

Here, the 2XMMp catalogue is used to investigate clustering in X-ray selected AGN. 2XMMp is a serendipitous X-ray source catalogue covering a large portion of the sky (285deg^2) and containing 123170 unique sources (for more details see section 2.3.2). The large number of objects and wide area should provide advantages over previous X-ray studies allowing large samples covering enough area to get a more accurate picture of cosmic variance. As it is a catalogue of X-ray sources, the sample should be able to probe the obscured AGN sources that are missed by optical surveys.

6.3 Method

The method used here is to determine the number of sources above a specified flux in each field for three energy bands, soft ($0.5-2\,\mathrm{keV}$), hard ($4.5-12\,\mathrm{keV}$) and full ($0.2-12\,\mathrm{keV}$), and compare the measured source distributions to statistical distributions to determine the excess variance observed. Redshift distributions from the XMM Wide Angle Serendipitous survey (XWAS, Tedds *et al.* 2006) are then used to estimate the redshift distribution of 2XMMp from which cosmic variance and bias can be calculated.

6.3.1 Field selection

The first step was to select a uniform sample of fields. Only fields obeying the following strict criteria were selected:

- High Galactic latitude ($|b| \geq 20°$) in order to avoid the Galactic plane and minimise contamination by stars
- Observed in all three cameras (MOS1, MOS2 and pn) so EPIC fluxes could be used
- Observed in full frame mode to keep the area of each field that can be used for serendipitous source detection constant
- Observed with any optical blocking filter except thick as this blocks too much of the soft X-ray radiation

758 of the 2400 fields in 2XMMp satisfied these criteria. These fields were then visually inspected and flagged as follows:

- Flag = 1: No central source and no problem areas

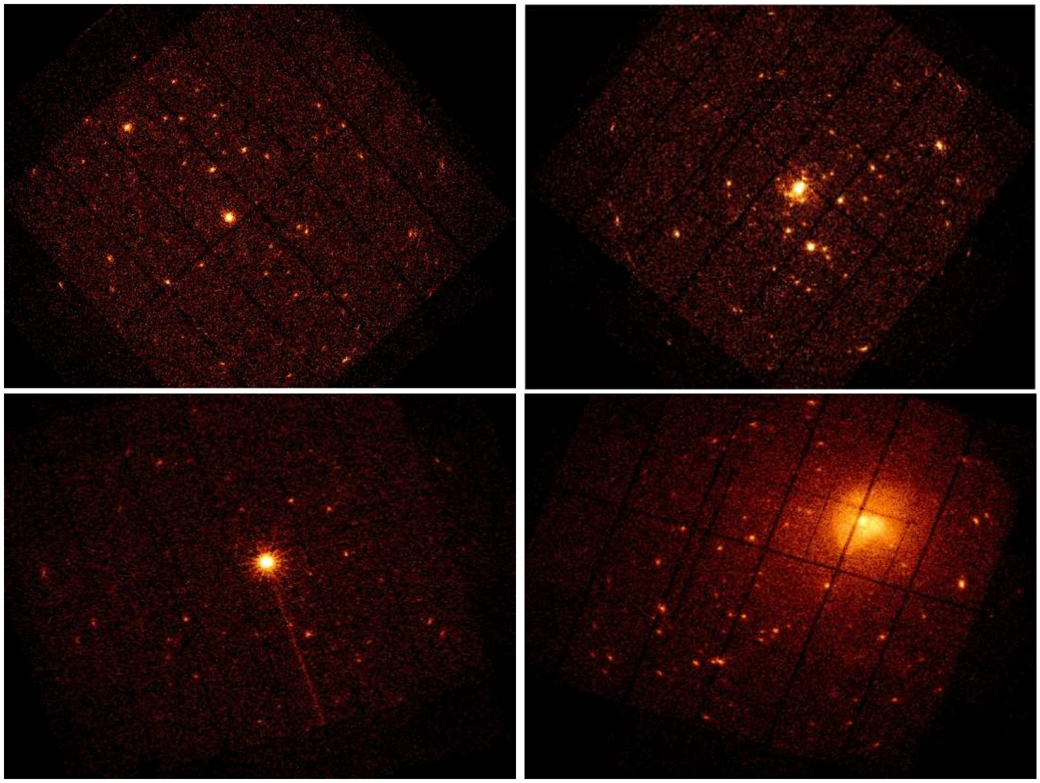


Figure 6.1: Example fields representing the four flags assigned during visual screening. Top left: flag 1, top right: flag 2, bottom left: flag 3, bottom right: flag 4.

- Flag = 2: Small central source
- Flag = 3: Central source, Out Of Time (OOT) stripe (caused by photons hitting the detector during the readout phase)
- Flag = 4: Problems away from centre, regions of extended emission

Figure 6.1 shows an example field for each flag.

All fields in the sample must have the same usable area. If a field with a large region of extended emission was used, the area of the emission would have to be excluded from all fields leading to a loss in potential data. By excluding fields with such areas this problem is avoided and the maximum amount of information from each field is retained. For this reason, only fields with flags 1 and 2 were taken forward for analysis (544 fields). Overlapping fields must be removed to ensure that the same source is not counted in multiple fields. Where fields were found to overlap, the field with the shorter exposure time was removed from the sample. Correction for overlapping left 398 fields in the sample.

6.3.2 Flux and exposure time cuts

For each band, several flux limits were chosen to study. For each flux level, the minimum exposure time required, so that all sources with flux greater than the minimum are detected, must be determined. To do this, sensitivity maps were generated for each field in each band.

Sensitivity maps indicate the minimum count rate that a source must have at a given point in the field in order to be detected at the desired significance level. A minimum likelihood of det_ml $= 6$ was used here to match the value used for source detection in 2XMMp. The sensitivity map is created using background maps, exposure maps and detector masks to take into account quantum efficiency, filter transmission, vignetting etc. Sensitivity maps were created for the energy bands used here giving the point source detection upper limits for each field. The minimum flux observable increases towards the edge of the field (i.e. sensitivity is greatest in the centre of the field) and so the flux level to be used was determined by the flux at the edge. For each field and energy band, the average count rate of the sensitivity map obtained in an annulus of inner radius $10'$ and outer radius $12'$ from the pointing coordinates was measured. This was converted to flux using ECFs (Energy Conversion Factors).

Using the latest calibration matrices, ECFs are calculated by assuming the source spectrum fits an absorbed power law model with absorbing column density $N_h = 3.0 \times 10^{20}\,\mathrm{cm}^{-2}$ and continuum spectral slope of $\Gamma = 1.7$. ECFs are provided on the XMM website for individual cameras, energy bands and filters[1]. ECFs were available for the hard band but had to be calculated for the soft and full bands by combining data from other bands. This was done using the count rates and fluxes expected for the above spectral model for the $0.2-0.5$, $0.5-1$, $1-2$ and $0.5-12\,\mathrm{keV}$ bands (kindly provided by Dr Silvia Mateos). The sum of rates (R) and sum of fluxes (F) for the appropriate bands were calculated and the ECF for the combined band (for separate cameras and filters) was then determined:

$$ECF = \frac{\sum R}{\sum F} \tag{6.5}$$

EPIC ECFs were calculated by summation of ECFs for the appropriate combination of cameras and filters for each field and energy band. With the EPIC ECFs, the average EPIC count rate between $10'$ and $12'$ for each field could be converted to flux:

[1] http://xmmssc-www.star.le.ac.uk/dev/Catalogue/2XMMp/UserGuide_2xmmp.html#TabECFs

$$F_{EPIC} = \frac{R_{EPIC}}{ECF_{EPIC}} \tag{6.6}$$

A plot of exposure time against flux was then made to determine suitable cuts for each energy band (see figure 6.2). Only fields satisfying the exposure time cuts for each band were used.

For each field satisfying the required initial criteria, with flag = 1 or 2 and with exposure time satisfying the cut, parameters for sources in the field were obtained from the 2XMMp source catalogue. For each source, the flux in each energy band and the position of the source in the field was recorded. Sources in each field above the flux threshold and within $12'$ of the field centre were counted and used to create source distribution plots. $12'$ was used as the maximum off axis angle to reduce problems due to PSF distortion which can be significant at larger angles.

6.3.3 Source distributions

398 non-overlapping fields were used for the source count. The overall area covered by the sample (having accounted for chip gaps etc.) is 41.0deg^2. The analysis was performed separately for soft, hard and full band sources and for different fluxes in each energy band. Table 6.1 shows the EPIC flux and exposure time limits and the source count results. Figure 6.3 shows the distribution of sources per field for each band and flux level. The values determined here compare favourably with the more rigorous

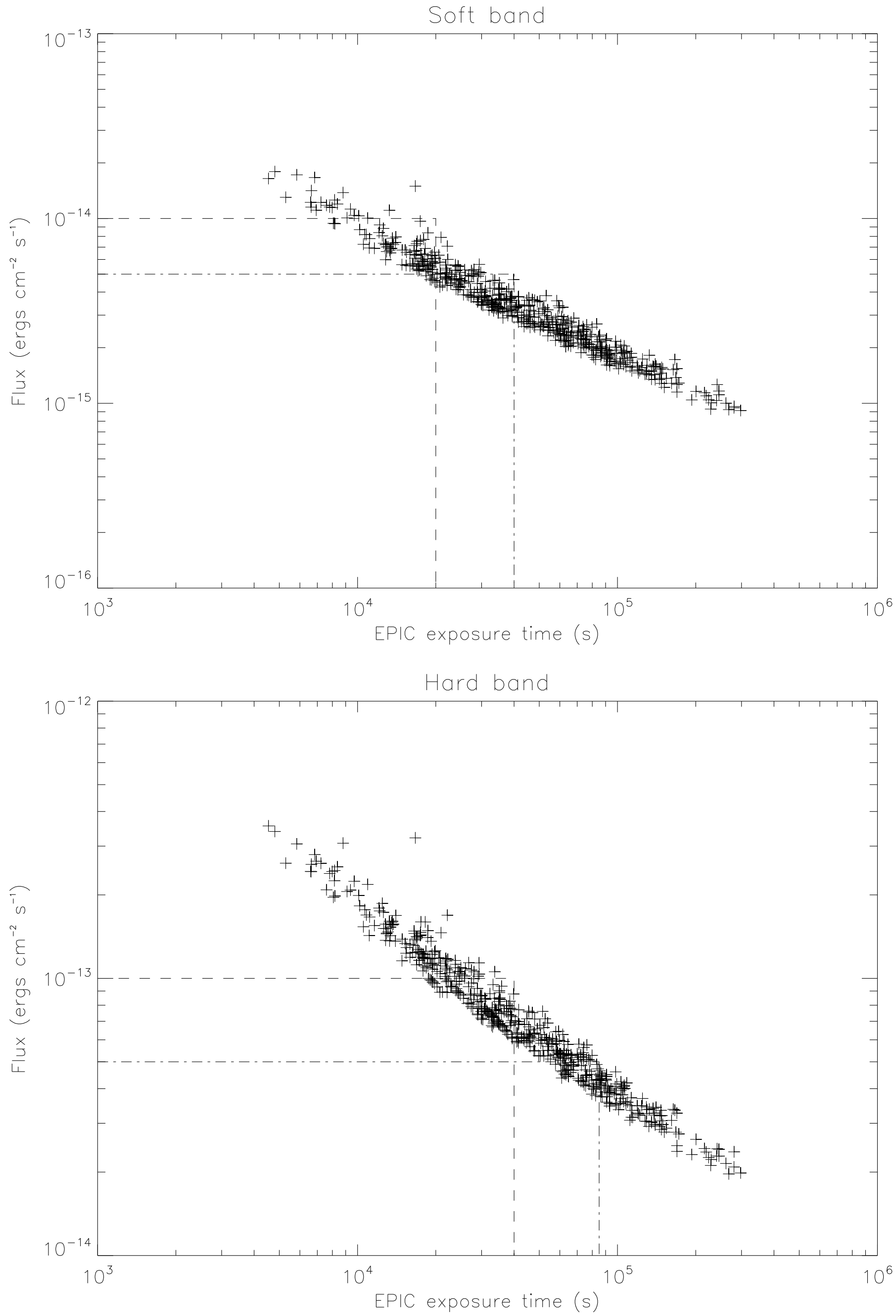


Figure 6.2: Exposure time and flux between $10'$ and $12'$ for each field for the three energy bands (soft, hard and full as labelled). The different line styles represent different cuts (see table 6.1).

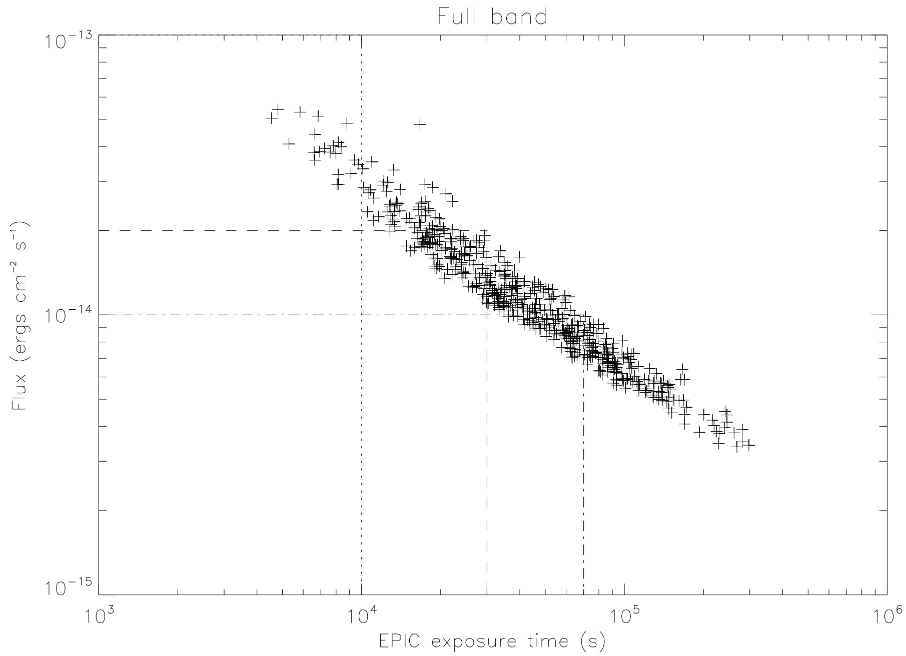


Figure 6.2: continued.

determination of a similar sample of objects made by Mateos *et al.* (2008).

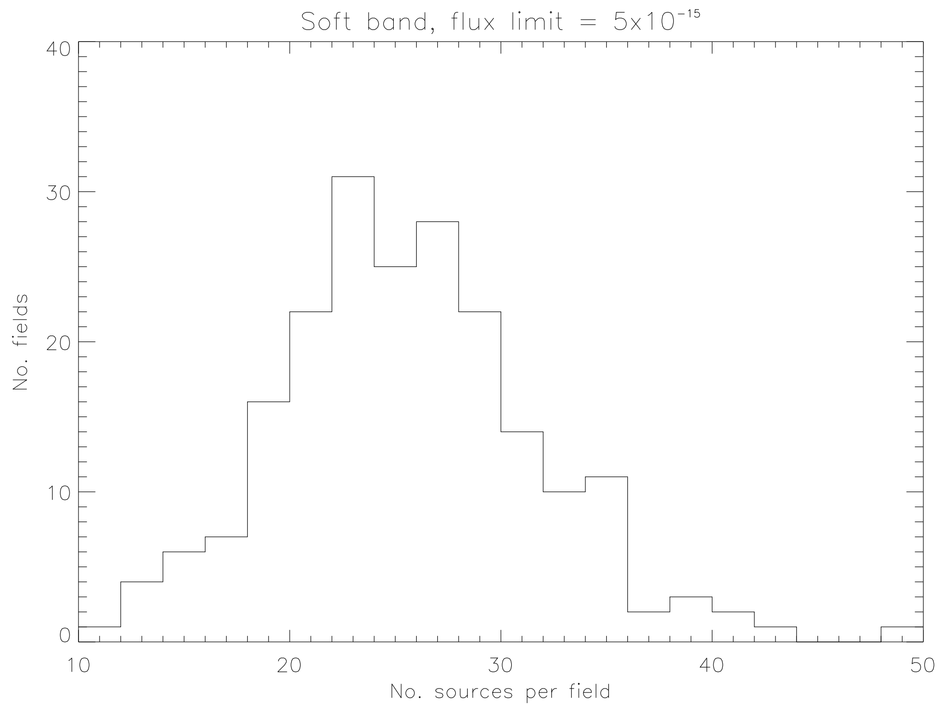
6.3.4 Determining excess variance

If the positions of astronomical objects are random, then plotting the number of fields against the number of sources in a field (with flux greater than S) should statistically give a Poisson distribution at low source counts while with large numbers of sources this will tend towards a Gaussian distribution (i.e. variance equal to the mean, $\sigma^2 = \mu$). If there is cosmic variance this distribution will be broadened. The difference between the measured distribution and the statistical variance expected is the excess variance, σ^2_{ex}. These distributions can therefore be used to determine the excess variance observed due

Table 6.1: Source count results

Band	Minimum flux	Minimum exposure time	No. fields	No. sources	Mean sources
	($\mathrm{erg\,cm^{-2}\,s^{-1}}$)	(s)			per field
(1)	(2)	(3)	(4)	(5)	(6)
Full	1×10^{-14}	70000	114	6788	59.54
	2×10^{-14}	30000	261	7752	29.70
	1×10^{-13}	10000	348	1416	4.07
Soft	5×10^{-15}	40000	206	5389	26.16
	1×10^{-14}	20000	306	3874	12.66
Hard	5×10^{-14}	85000	85	315	3.71
	1×10^{-13}	40000	206	269	1.31

NOTES: Mean sources per field = no. sources / no. fields.



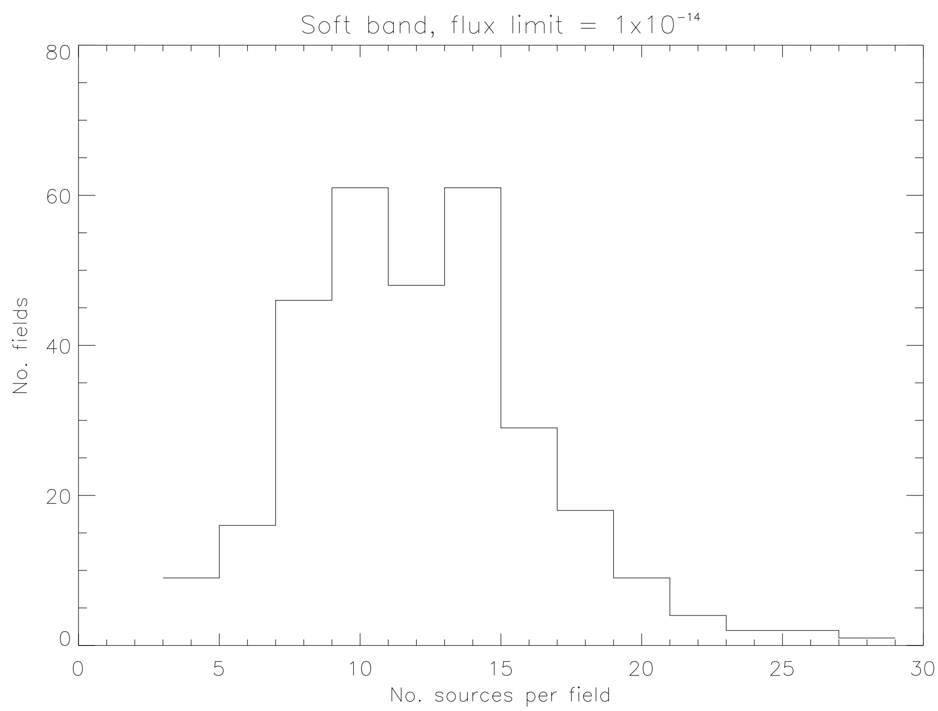
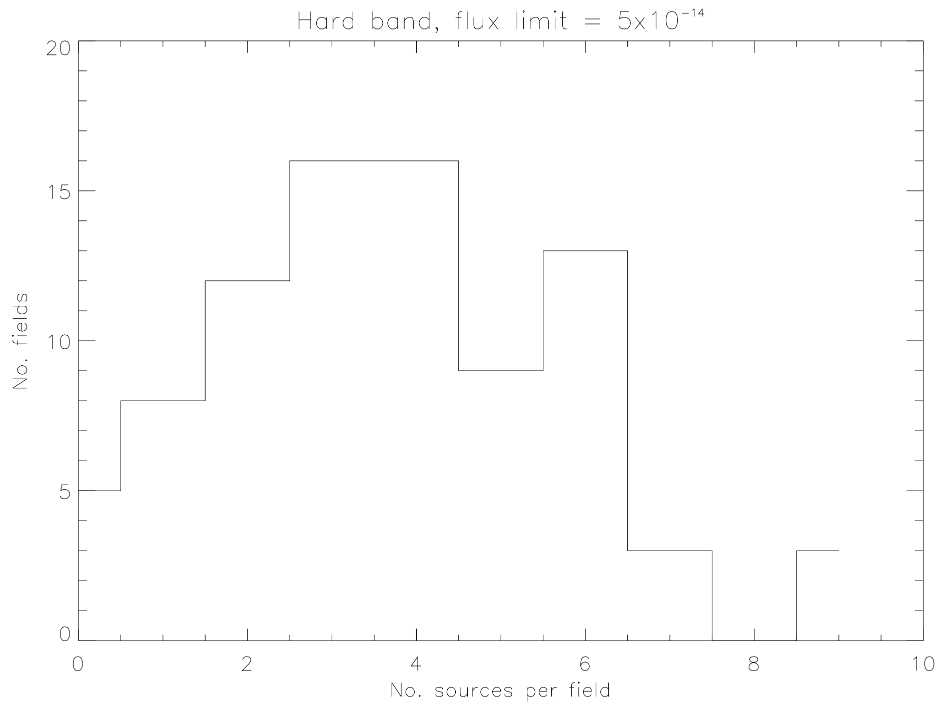


Figure 6.3: Distributions of the number of sources per field for the different energy bands and flux levels (as labelled).



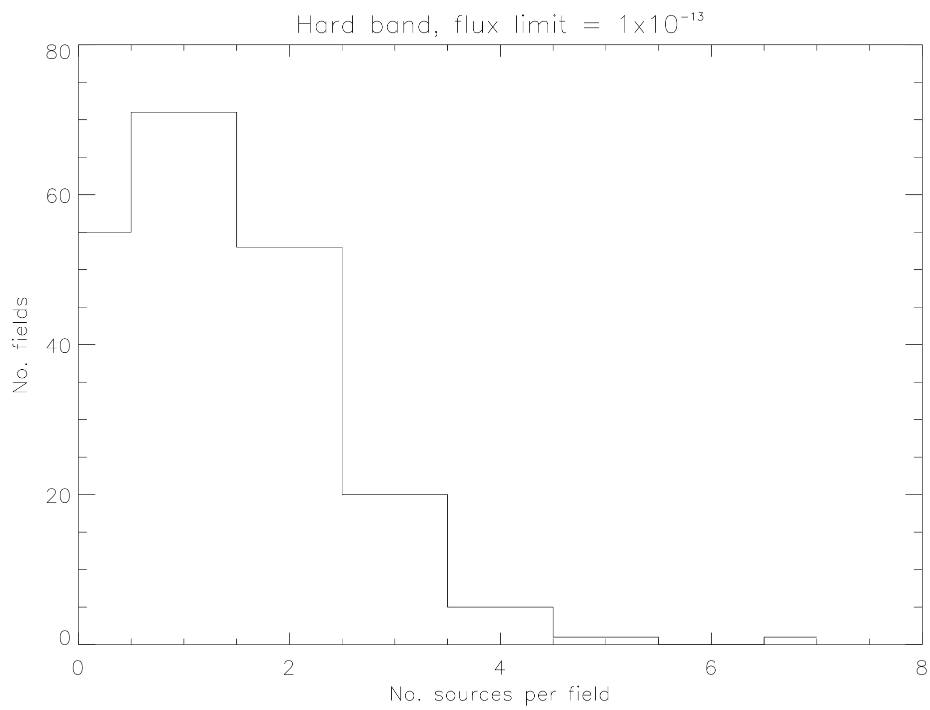


Figure 6.3: continued

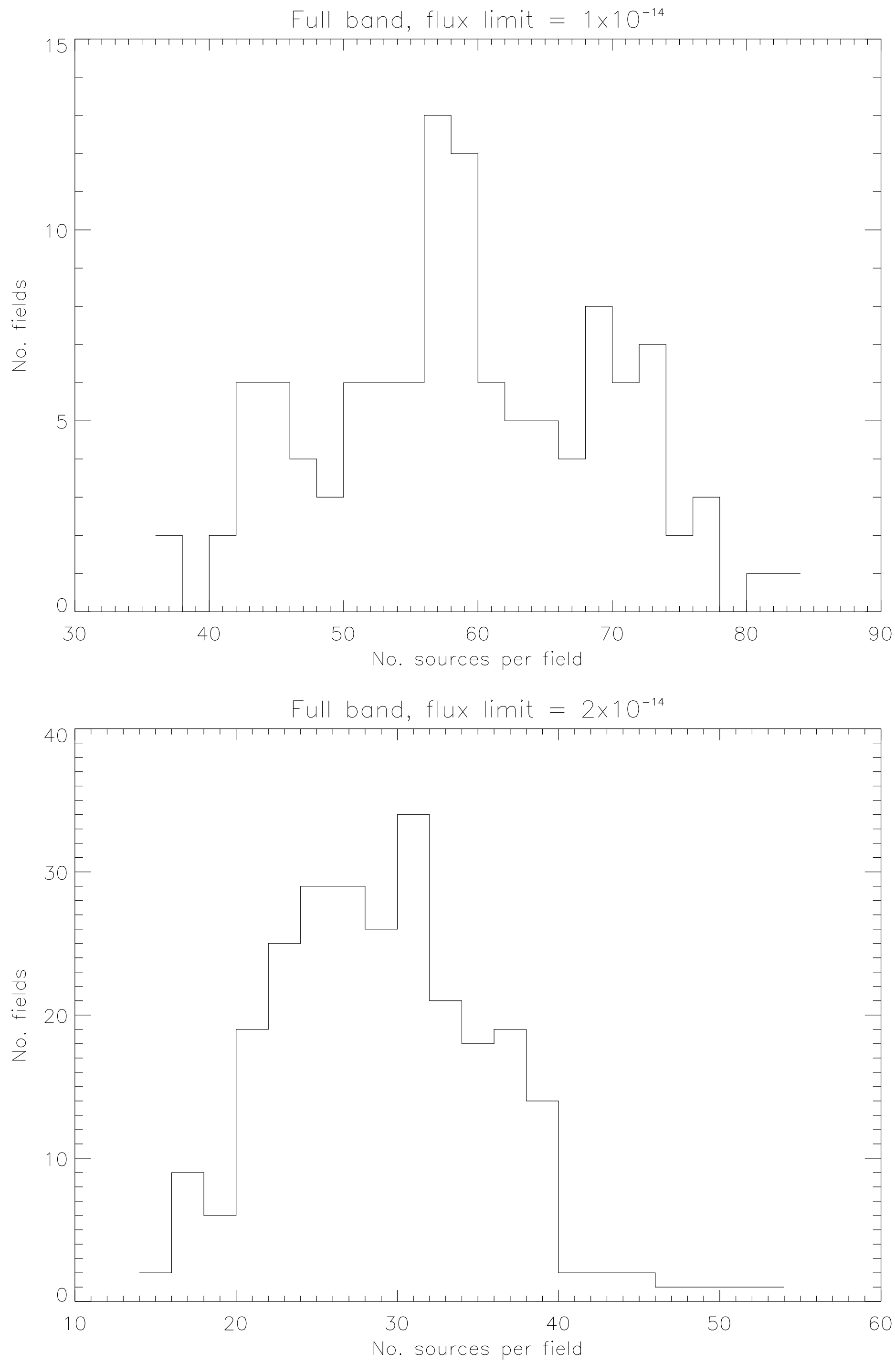


Figure 6.3: continued

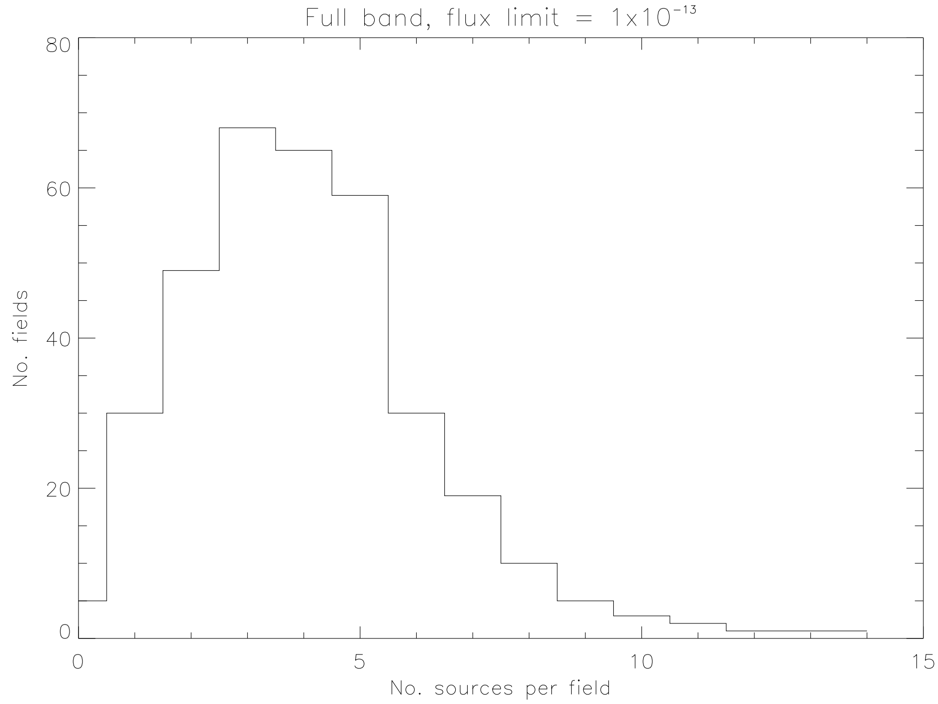


Figure 6.3: continued

to cosmic variance.

For flux limits with large source numbers, fitting a Gaussian distribution to the resulting source distribution provides the excess variance.

$$\sigma^2 = \mu + \sigma_{ex}^2 \tag{6.7}$$

The variance, σ^2, and mean, μ, of the distribution can be determined from the fit and the excess variance can therefore be calculated with the above equation. Figure 6.4 shows the results of this process for the soft band distributions and for the full band distributions with flux limits 1×10^{-14} and $2 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$. The dashed curve is a standard Gaussian distribution using the mean of the underlying distribution to calculate the standard deviation ($\sigma = \sqrt{\mu}$). The solid line is a Gaussian fit to the actual distribution (the histogram) and clearly deviates from the expected Gaussian for that mean number of sources, demonstrating the presence of excess variance. The fitted σ and μ values and the resulting σ_{ex} values are given in table 6.2.

For flux limits with low mean number of sources per field (i.e. both the hard band results and the full band with flux limit $1 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$), if there is no cosmic variance then the source distribution is statistically Poissonian but, because of the excess variance, it is expected to be more accurately described by a convolution of Poisson and Gaussian distributions. Here, the convolution is tested with a range of possible σ_{ex} values. Figure 6.5 shows the χ^2 distribution of the resulting curves. The minimum value is σ_{ex}. Errors on σ_{ex} are the 1σ confidence limits (where $\chi^2 = \chi^2_{min} + 1$).

The excess variance for each band and flux combination is given in table 6.2 with an indication of whether the Poisson (P) or Gaussian (G) method was used.

6.3.5 Calculation of the AGN bias parameter

Recent results (e.g. from WMAP) have helped to constrain theories on the growth of structure. It is now possible to calculate the expected excess variance using these results. The observed cone of each field was divided into redshift slices with width of $\Delta z = 0.05$. The angular diameter distance, D_A, of each slice was calculated:

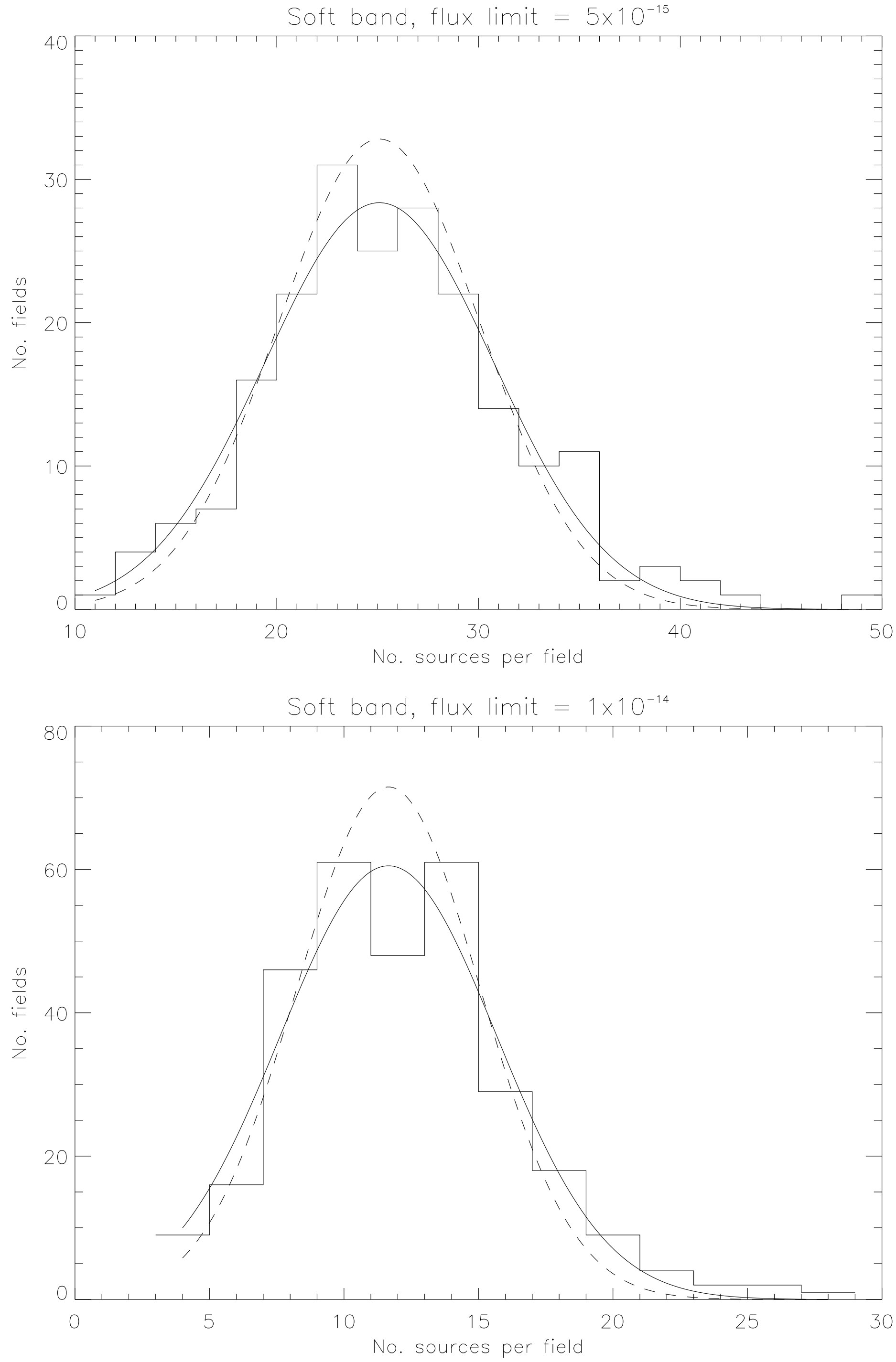


Figure 6.4: Gaussian fitting results for the soft and full bands (as labelled). The histogram is the observed distribution from 2XMMp. The dashed curve shows a Gaussian with $\sigma = \sqrt{\mu}$ and the solid line is the Gaussian fit to the underlying distribution. The histograms are plotted with bin width of two.

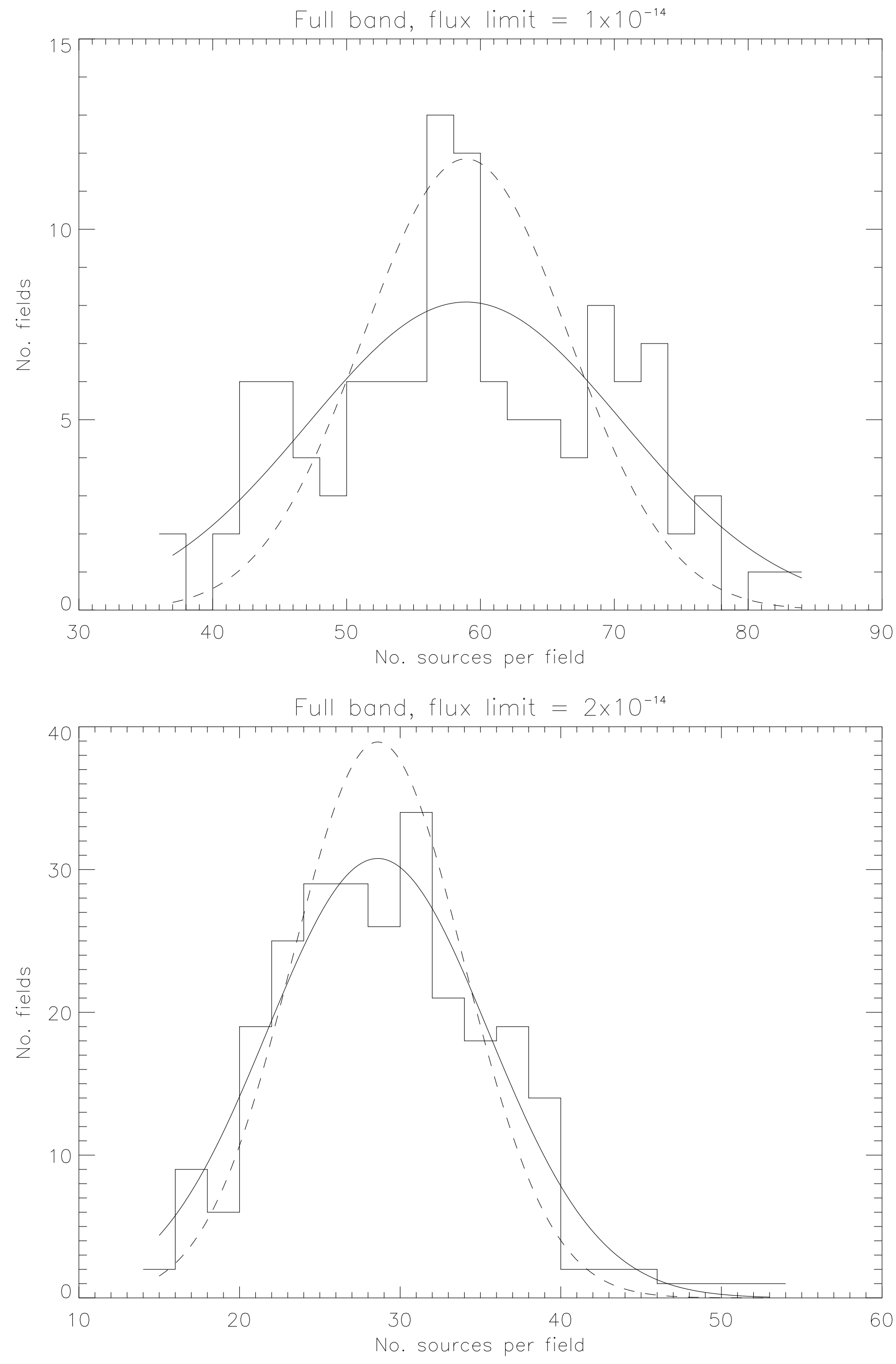


Figure 6.4: continued.

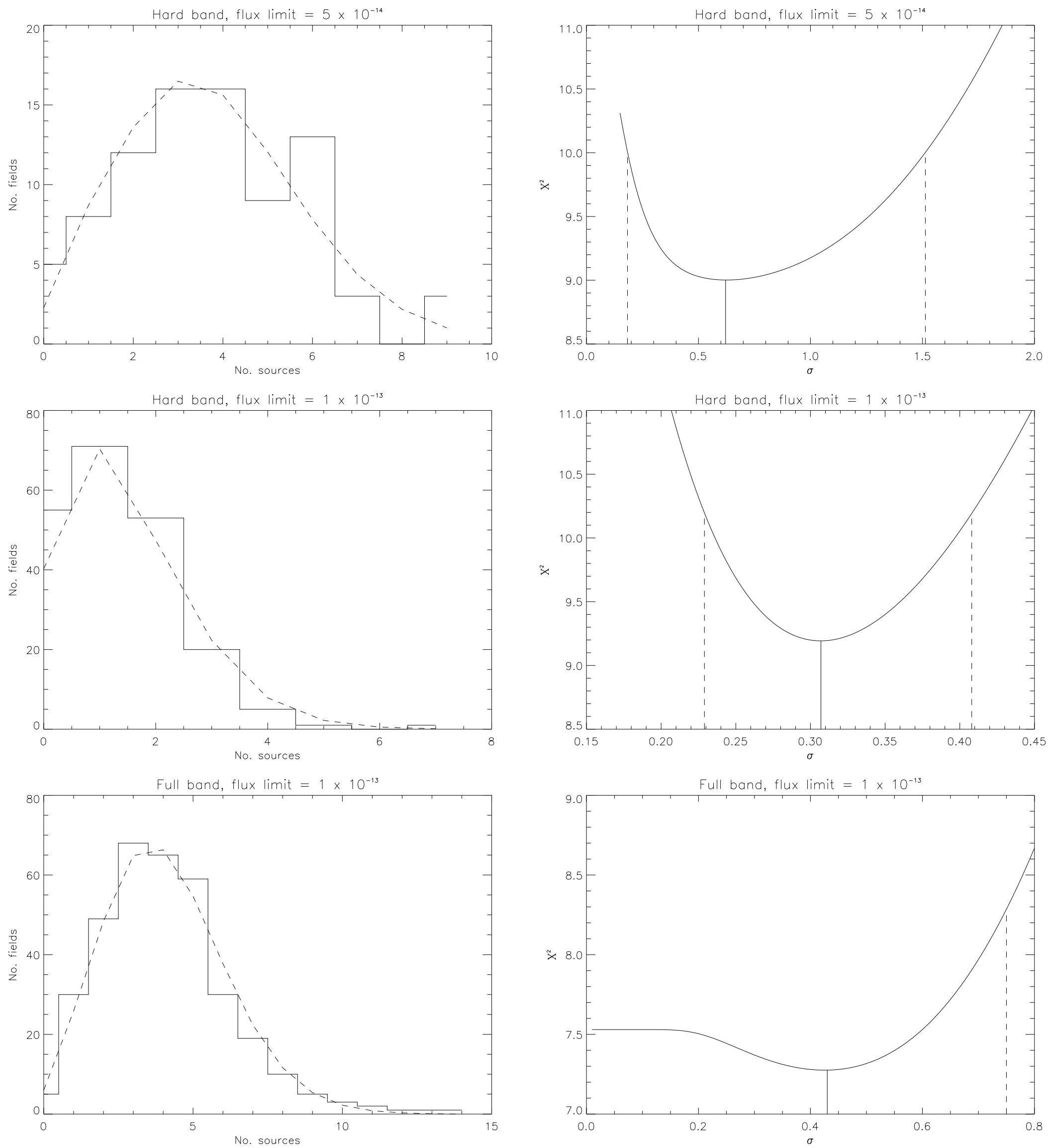


Figure 6.5: Poisson convolution results for the hard band distributions (top: flux limit = $5 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ and middle: flux limit = $1 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) and full band (bottom: flux limit = $1 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$). The left hand plots show the 2XMMp source distribution as a histogram. The dashed line is the convolved curve for the best fit σ_{ex} in each case. The right hand plots show the χ^2 distribution. The solid vertical line indicates σ_{ex}. Dashed lines show the upper and lower error bounds.

Table 6.2: Excess variance

Band	Flux ($\mathrm{erg\,cm^{-2}\,s^{-1}}$)	Mean	σ	σ_{ex}	P/G
(1)	(2)	(3)	(4)	(5)	(6)
Full	1×10^{-14}	$59.4\pm^{2.5}_{2.6}$	$12.0\pm^{3.3}_{2.3}$	$9.2\pm^{4.4}_{3.0}$	G
	2×10^{-14}	29.1 ± 0.9	$6.8\pm^{0.9}_{0.8}$	$4.1\pm^{1.5}_{1.3}$	G
	1×10^{-13}	4.07	2.21	$0.4\pm^{0.3}_{0.4}$	P
Soft	5×10^{-15}	25.6 ± 0.4	$5.6\pm^{0.5}_{0.4}$	2.4 ± 1.0	G
	1×10^{-14}	12.1 ± 0.4	$3.9\pm^{0.5}_{0.4}$	$1.8\pm^{1.1}_{0.9}$	G
Hard	5×10^{-14}	3.71	2.09	$0.6\pm^{0.9}_{0.4}$	P
	1×10^{-13}	1.31	1.14	0.3 ± 0.1	P

NOTES: For bands using the Gaussian method (G), mean and σ are the best fit parameters from the Gaussian fit to the distribution of number of sources per field. For bands using the Poisson method (P), mean and σ are the usual statistical mean and standard deviation

$$a(z) = \frac{1}{1+z} \tag{6.8}$$

$$Z = \int_{a(z)}^{1} \frac{da}{a\sqrt{X}} \tag{6.9}$$

$$\sqrt{X} = \sqrt{\Omega_k + \Omega_\Lambda a^2 + \frac{\Omega_M}{a} + \frac{\Omega_r}{a^2}} \tag{6.10}$$

$$D_A = \frac{cZa(z)}{H_0} \tag{6.11}$$

where z is the redshift at the centre of the slice, H_0 is the Hubble constant ($70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$) and c is the speed of light ($3 \times 10^8\,\mathrm{m\,s^{-1}}$). The Ω values are different components of the density parameter: Ω_k is the curvature density (0.0), Ω_Λ is the vacuum density (0.7), Ω_M is the matter density (0.3) and Ω_r is the radiation density (0.0) (values in line with the *WMAP* concordance cosmology, Spergel *et al.* 2003). The radius of each slice, R, was then calculated using the angular radius of the field ($r = 12'$)

$$R = D_A tan(r) \tag{6.12}$$

The proper distance, D_P was also calculated:

$$D_P = D_A(1+z) \tag{6.13}$$

The predicted mass variance (rms mass fluctuation) in each redshift slice can be calculated as:

$$\sigma_m^2 = \frac{1}{2\pi^2}\int_0^\infty P(k,z)W^2(kR)k^2dk \tag{6.14}$$

where $P(k)$ is the power spectrum of initial inhomogeneities and $W(kR)$ is the window function (this and following equations from Coles and Lucchin 1995). The power spectrum is generally approximated by a scale invariant power law:

$$P(k) = Ak^n \tag{6.15}$$

where A is the scalar fluctuation amplitude and n is the spectral index. Both parameters can be determined from cosmological models using e.g. Cosmic Microwave Background measurements. Tegmark *et al.* (2004) used a combination of WMAP data and data from over 200000 SDSS galaxies to determine the values of A and n ($0.81^{+0.15}_{-0.09}$ and $0.977^{+0.039}_{-0.025}$ respectively). These values are for a specific wavenumber, $k_0 = 0.05$, so the power spectrum becomes:

$$P(k) = A\left(\frac{k}{k_0}\right)^n \tag{6.16}$$

The power spectrum evolves with time according to linear growth law in the absence of other physical effects:

$$P(k, z) = \frac{P(k)}{1 + z} \tag{6.17}$$

A linear approximation to the local power spectrum is calculated and normalised to the value of σ_8 which is the linear rms mass fluctuation in spheres of radius $8\,\mathrm{h}^{-1}\,\mathrm{Mpc}$

($0.917\pm^{0.090}_{0.072}$ as quoted in Tegmark *et al.* 2004).

In line with Tegmark *et al.* (2004), a top hat window function is used:

$$W = \frac{3(sin(kR) - (kR)cos(kR))}{(kR)^3} \tag{6.18}$$

The window function effectively limits the perturbation contributions to the mass variance to a range of wavelength scales greater than the scale corresponding to the projected field size smoothly decreasing the contributions of smaller wavelengths which are effectively averaged over the field. To convert σ_m^2 to the mass variance expected for 2XMMp, a source redshift distribution is required. The redshifts of sources from XWAS (Tedds *et al.* 2006) were used to estimate the number of objects expected in each redshift slice of sky (redshifts kindly provided by Dr Jonathan Tedds). XWAS is a survey aiming to identify serendipitous XMM sources with $F_{0.5-4.5\,\mathrm{keV}} \geq 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$. Over 3000 sources have been targeted in 68 XMM fields at high Galactic latitudes using the AAT (Anglo Australian Telescope) multi fibre spectroscope. As it is a random sample of the fields included in 2XMMp and uses the same energy bands used here, XWAS is appropriate for use as an approximation to the redshift distribution of sources in 2XMMp. Although 3000 sources have been targeted by XWAS, only around 1000 have so far been spectroscopically identified and matched to *XMM-Newton* sources and so the current redshift distribution may not be accurate for the X-ray population of 2XMMp in general. This caveat must be kept in mind when drawing conclusions from the results.

The XWAS sources were split into the same energy bands and flux limits used for the 2XMMp analysis and the distributions in the same redshift bins described above ($\Delta z = 0.05$) were determined. Only extragalactic sources (i.e. Broad Line AGN (BLAGN) and Narrow Emission Line Galaxies (NELG)) were included in the redshift distributions. Each distribution was divided by the total number of sources to get the normalised number of sources per redshift bin (shown in figure 6.6) then each bin was multiplied by the mean number of objects per field from the 2XMMp analysis (see table 6.1) to give the expected number of objects per redshift bin, N, that should be observed in a 2XMMp field.

For each redshift bin, the expected mass variance in a field is:

$$\sigma_z^2 = \frac{N\sigma_m^2}{\frac{\Delta D_P}{2R}} \tag{6.19}$$

where ΔD_P is the difference between the proper distance of the redshift slice in question and the next closest redshift slice.

The statistical variance is $\sigma_s^2 = N$ so total variance is:

$$\sigma_{tot}^2 = \sigma_s^2 + \sigma_z^2 \tag{6.20}$$

$$\sigma_{tot}^2 = N\left(1 + \frac{\sigma_m^2}{\frac{\Delta D_P}{2R}}\right) \tag{6.21}$$

The total contribution from cosmic variance is therefore:

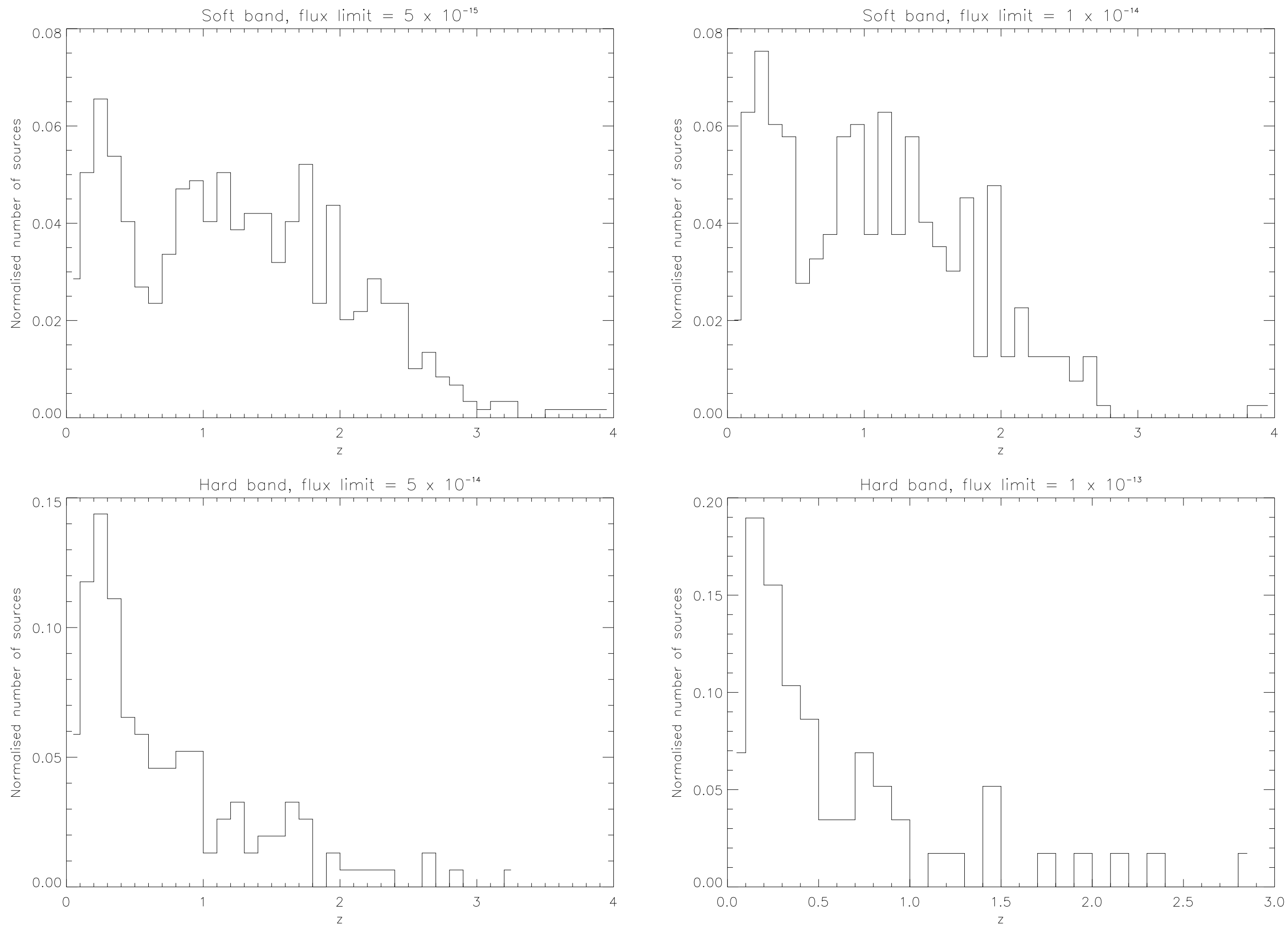


Figure 6.6: Normalised redshift distributions for the different bands and flux limits (as labelled) based on the XWAS survey.

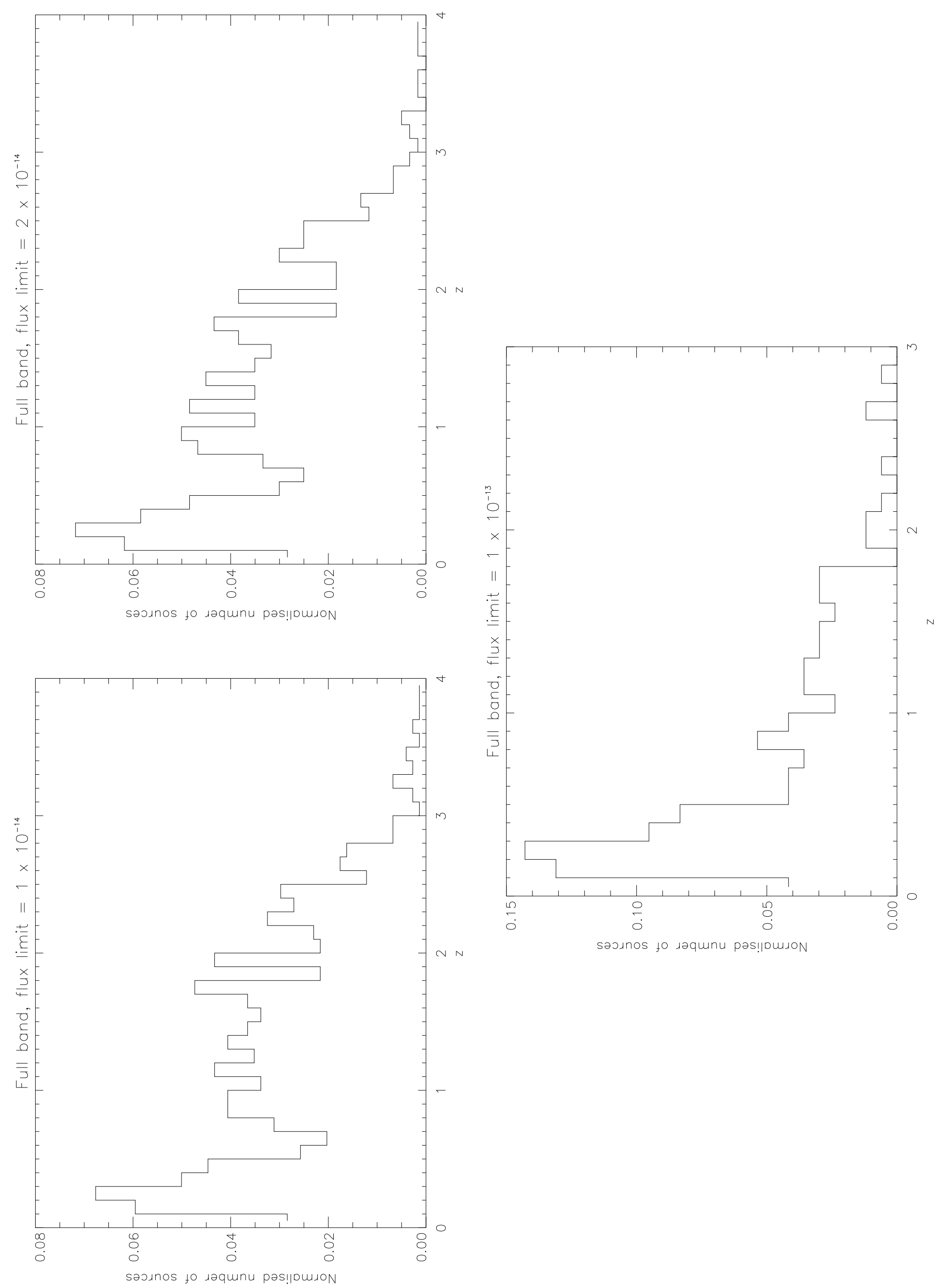


Figure 6.6: continued

$$\sigma_{CV}^2 = \sum_{z_{min}}^{z_{max}} \frac{\sigma_m^2(z)N(z)}{\frac{\Delta D_P}{2R}} \tag{6.22}$$

where z_{min} is the redshift below which the linear growth regime breaks down (i.e. where σ_m^2 becomes greater than 1). This point is evident in figure 6.7. Note that typically only $20-25\%$ of the sources here have redshifts lower than this value and very large power would be needed at the relevant wavelengths to contribute significantly to the variance. At $z < 0.5$, we are probing scales where the power spectrum has, in practice, turned down and so any contribution to the excess variance is expected to be small. Here, z_{min} was taken to be $z = 0.5$ and z_{max} is the highest value for which there is an estimate of σ_m^2 ($z_{max} = 3.9$ for full band (1×10^{-14} and $2 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) and soft band, $z_{max} = 2.85$ for full band ($1 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) and hard band ($1 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) and $z_{max} = 3.25$ for hard band ($5 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$)). Figure 6.7 shows the expected and measured mass variance for the full band (flux = $1 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$). This shows that the sample is most sensitive to detecting cosmic variance at $1 \leq z \leq 2.5$ where the difference between total variance and cosmic variance is most prominent.

The bias parameter is a multiplier between the underlying mass density incorporating all mass (i.e. including dark matter) and the AGN mass density. In this case, σ_{CV} represents the total density and σ_{ex} is the AGN density so the bias parameter is:

$$b = \frac{\sigma_{ex}}{\sigma_{CV}} \tag{6.23}$$

The results of the redshift analysis for each band and flux are given in table 6.3. Also given in table 6.3 are the mean redshift ($\overline{z}$) and scale length (R) for each band.

6.4 Results

The bias values for the various bands and flux levels are displayed in figure 6.8 along with bias values determined by other studies (Miyaji *et al.* 2007, Myers *et al.* 2007, Grazian *et al.* 2004, Yang *et al.* 2006 and Croom *et al.* 2005). Although, in general, the bias values found here are low in comparison to other studies, figure 6.8 shows that the result of the full band analysis (flux limit $= 1 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) is in agreement, within the error bounds, with the very large optical samples of Myers *et al.* (2007) and Croom *et al.* (2005) (300000 and 22655 sources respectively). There is also some tentative evidence for evolution of the bias parameter in the full band with redshift as previously detected in other studies (e.g. Croom *et al.* 2005).

The difference in bias values here compared to the values of optical surveys such as Myers *et al.* (2007) and Croom *et al.* (2005) may be due to the fact that, as an X-ray sample, a more complete sample of AGN is detected here (i.e. including more of the obscured population).

In a study of AXIS fields, Carrera *et al.* (2007) found clustering in the soft and XID bands but not in the hard or ultra-hard bands. Carrera *et al.* (2007) point out that the hard and ultra hard band samples contained fewer sources than the soft and XID bands (under 1000 sources) and that, if sources were randomly removed from the soft band to reduce numbers, the clustering signal disappeared. This implies that the lack of clustering detected in the hard band so far may be due to a lack of sources rather than an actual lack of clustering. The hard band samples used here also have significantly fewer sources than the soft and full band samples which may cause inaccuracies in the results. A repeat of the analysis using the recently released 2XMM catalogue (see section 2.3.2) may yield greater numbers in the hard band and rectify this. Despite the low numbers, there is tentative evidence that the bias is similar for the soft and hard bands in agreement with Gilli *et al.* (2005). Whether this is actually the case will become evident with larger samples.

Using the relation between bias and dark matter halo mass discussed in Sheth and Tormen (1999), the average halo mass of objects in the sample discussed is $10^{12} - 10^{13}\,\mathrm{M_{\odot}\,h^{-1}}$, based on the bias calculated for the full band at flux limit $= 1 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, which is comparable to the finding of Yang *et al.* (2006).

Several caveats must be considered when discussing the findings of this study. The fields in the sample are taken from the 2XMMp catalogue which means they are fields that have been observed by *XMM-Newton* and therefore not a random sample of fields across the sky but rather a sample of fields with interesting objects at the centre. A more important bias is introduced by the use of the XWAS sample to estimate the redshift distribution. XWAS determines redshifts by optical spectroscopy and therefore targets primarily bright sources. This method is therefore biased against both high redshift sources, low luminosity sources and optically obscured AGN. As one of the main aims

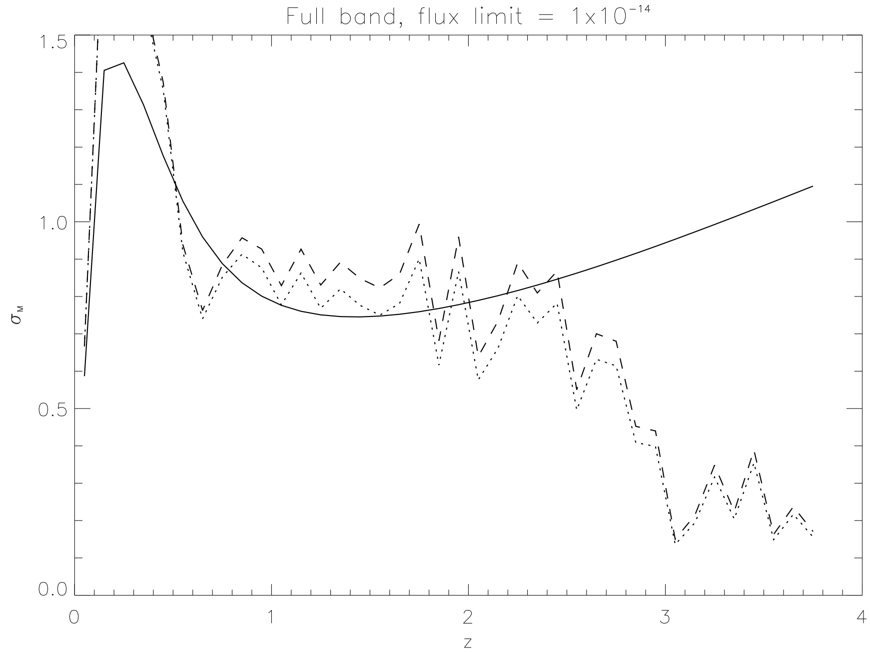


Figure 6.7: Plot of the predicted (solid line) mass variance, the total measured variance (dashed line) and the variance due to cosmic variance (dotted line) in the full band (for flux level $1 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$).

of using X-ray observations to investigate clustering is to target obscured AGN, an improvement to the technique employed here would be to find an alternative method of determining the 2XMMp redshift distribution, e.g. using photometric redshifts.

6.5 Conclusions

Using the serendipitous source catalogue 2XMMp and an estimate of its redshift distribution, estimates of the QSO bias parameter for various X-ray energy bands and flux limits have been determined. The full band $(0.2 - 12\,\mathrm{keV})$ distribution with flux

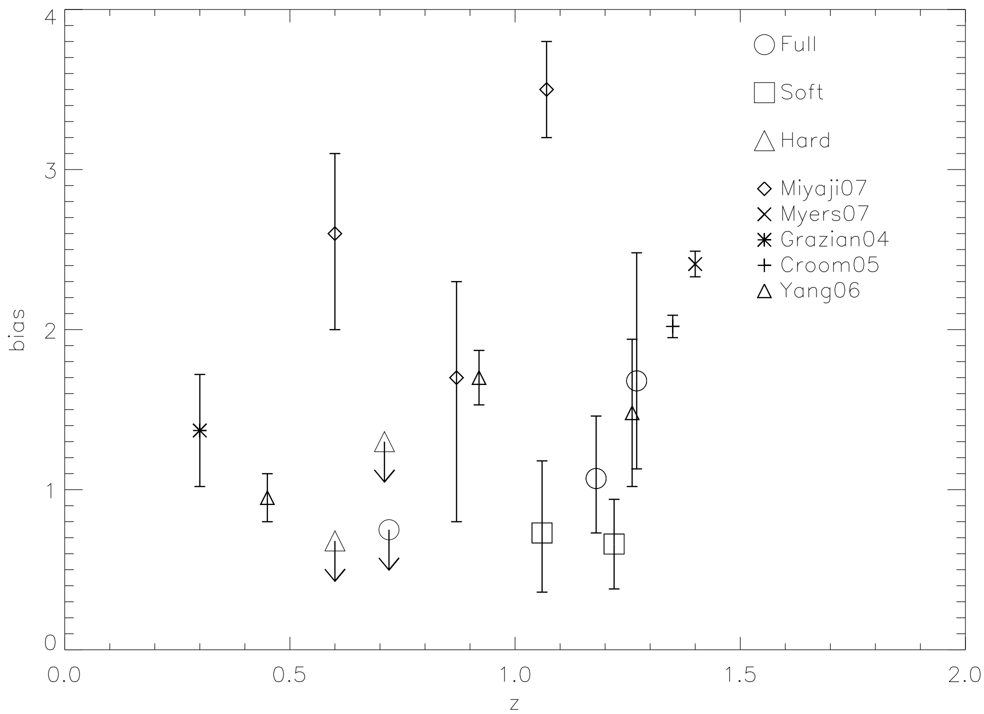


Figure 6.8: The results of the 2XMMp bias analysis compared to other clustering studies. References refer to studies described in section 6.2. The three points from Miyaji *et al.* (2007) are for, from left to right with increasing redshift, the hard ($4.5 - 10\,\mathrm{keV}$), medium ($2 - 4.5\,\mathrm{keV}$) and soft ($0.5 - 2\,\mathrm{keV}$) bands. The three Yang *et al.* (2006) values show the evolution of bias with redshift they detected. Bias values determined here using the Poisson method to determine σ_{ex} (see section 6.3.4) are taken to be upper limits.

Table 6.3: Cosmic variance results

Band	Flux	σ_{CV}	Bias	$\overline{z}$	R
	($\mathrm{erg\,cm^{-2}\,s^{-1}}$)				(Mpc)
(1)	(2)	(3)	(4)	(5)	(6)
Full	1×10^{-14}	5.48	$1.7\pm^{0.8}_{0.5}$	1.27	5.17
	2×10^{-14}	3.82	$1.1\pm^{0.4}_{0.3}$	1.18	5.12
	1×10^{-13}	1.18	< 0.6	0.72	4.47
Soft	5×10^{-15}	3.64	0.7±0.3	1.22	5.15
	1×10^{-14}	2.45	0.7±0.4	1.06	5.02
Hard	5×10^{-14}	1.16	$0.5\pm^{0.8}_{0.4}$	0.71	4.45
	1×10^{-13}	0.60	$0.5\pm^{0.2}_{0.1}$	0.60	4.14

$\geq 1 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ shows a level of bias comparable to large, optically selected QSO samples. The size of the samples used here are still relatively small compared to the large optical samples they are compared to, leading to the large error bars observed.

An area of over 40deg^2 was sampled here using 2XMMp. This study has shown that the technique used can yield promising results but would benefit from still larger samples and better redshift determination, particularly in the soft and hard bands. Both of these requirements should be possible in the near future. The recent release of 2XMM, which includes over 1000 more fields than 2XMMp, will increase the area available to sample for analysis such as this. Annual incremental increases to 2XMM are also planned, further increasing the potential sample size. Using photometric redshifts or a less biased optical catalogue to determine the redshift distribution would also improve results. When these resources are available, the method used here can be used with very large samples to improve on the tentative results found here.

Chapter 7

Conclusions

Throughout this thesis, archival X-ray data has been used to show that a variety of interesting science aims can be investigated without the need for targeted observations.

Techniques for determining photometric redshifts of serendipitous X-ray selected AGN and for cross correlation of X-ray sources with optical samples have been developed and used with small pilot samples to demonstrate the potential use that archive data can have.

The main results found are as follows:

- A method for determining accurate photometric redshifts ($\sigma_{rms} = 0.47$) for X-ray selected AGN with both point like and extended optical counterparts was developed. Modification of a code designed for optically selected QSOs in the SDSS (Weinstein *et al.* 2004) led to a code which uses different training sets, dependent on object properties, and a χ^2 method of error determination to produce photometric redshift estimates with 1σ error estimates. The accuracy of the method surpasses that available from other photometric redshift methods used for X-ray selected AGN.

- Application of the photometric redshift code to 762 sources from 2XMM with matches to optical SDSS counterparts and spectroscopic redshifts showed an impressive 67% of sources assigned photometric redshift that is 'correct' (within the error bounds) and 70% of sources with $|\Delta z| \leq 0.1$.

- Using the techniques for photometric redshift and X-ray/optical source matching developed in Chapter 3, X-ray light curves for the serendipitous sources detected in multiple observations of the 3C 273 field were analysed. The excess variance of the light curves was found to be anti-correlated with the X-ray luminosity as seen in many previous studies.

- Using *XMM-Newton* observations of eight Seyfert 1 galaxies, X-ray and optical light curves were examined for correlations in variability. Variability was detectable in four of the eight objects. In three of these four objects, constraints on the CCF were sufficient to rule out strong X-ray/optical correlation on timescales of less than one day.

- Examination of the 'bigger picture' with archive data was also shown to be possible. Cosmic variance and an estimate of the AGN bias parameter were determined using samples of different flux and energy band from 2XMMp. The bias in the full band ($0.2-12\,\mathrm{keV}$, for flux $\geq 1\times10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) was in agreement, within the error estimates, with bias found in large optically selected samples despite the much smaller sample size.

7.1 Future work

The results described above are all from relatively small samples of AGN when compared to the large numbers (hundreds of thousands) possible with optical samples thanks to large surveys such as SDSS and 2dF. The release of 2XMM in August 2007 has gone some way to redressing this balance and the situation will further improve in the coming years. The *XMM-Newton* Survey Science Centre at Leicester intend to release annual incremental updates to 2XMM with an estimated 30000 new unique sources in each release. With *XMM-Newton* set to be operational until at least 2012 that could mean a serendipitous X-ray resource containing nearly 300000 X-ray sources within four years. Further increases in size of multi-wavelength observations such as SDSS and UKIDSS with which to cross correlate large X-ray samples will continue to improve the results possible with such catalogues.

Despite the expected dearth of new X-ray observatories in the coming decade, this thesis has shown that archival data from *XMM-Newton* and *Chandra* has the potential to be successfully exploited for a variety of key science aims. Unlocking this potential and interpreting the astrophysical implications of the results will keep the X-ray astronomy community busy for many years to come.

BIBLIOGRAPHY

Abramowicz, M. A., Bao, G., Lanza, A., and Zhang, X.-H. (1991). X-ray variability power spectra of active galactic nuclei. *A&A*, **245**, 454–456.

Abrassart, A. and Czerny, B. (2000). Toy model of obscurational variability in active galactic nuclei. *A&A*, **356**, 475–489.

Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., Bailer-Jones, C. A. L., Baldry, I. K., Barentine, J. C., Bassett, B. A., Becker, A. C., Beers, T. C., Bell, E. F., Berlind, A. A., Bernardi, M., Blanton, M. R., Bochanski, J. J., Boroski, W. N., Brinchmann, J., Brinkmann, J., Brunner, R. J., Budavári, T., Carliles, S., Carr, M. A., Castander, F. J., Cinabro, D., Cool, R. J., Covey, K. R., Csabai, I., Cunha, C. E., Davenport, J. R. A., Dilday, B., Doi, M., Eisenstein, D. J., Evans, M. L., Fan, X., Finkbeiner, D. P., Friedman, S. D., Frieman, J. A., Fukugita, M., Gänsicke, B. T., Gates, E., Gillespie, B., Glazebrook, K., Gray, J., Grebel, E. K., Gunn, J. E., Gurbani, V. K., Hall, P. B., Harding, P., Harvanek, M., Hawley, S. L., Hayes, J., Heckman, T. M., Hendry, J. S., Hindsley, R. B., Hirata, C. M., Hogan, C. J., Hogg, D. W., Hyde, J. B., Ichikawa, S.-i., Ivezić, Ž., Jester, S., Johnson, J. A., Jorgensen, A. M., Jurić, M., Kent, S. M., Kessler, R., Kleinman, S. J., Knapp, G. R., Kron, R. G., Krzesinski, J., Kuropatkin, N., Lamb, D. Q., Lampeitl, H., Lebedeva, S., Lee, Y. S., Leger, R. F., Lépine, S., Lima, M., Lin, H., Long, D. C., Loomis, C. P., Loveday, J., Lupton, R. H., Malanushenko, O., Malanushenko, V., Mandelbaum, R., Margon, B., Marriner, J. P., Martínez-Delgado, D., Matsubara, T., McGehee, P. M., McKay, T. A., Meiksin, A., Morrison, H. L., Munn, J. A., Nakajima, R., Neilsen, Jr., E. H., Newberg, H. J., Nichol, R. C., Nicinski, T., Nieto-Santisteban, M., Nitta, A., Okamura, S., Owen, R., Oyaizu, H., Padmanabhan, N., Pan, K., Park, C., Peoples, J. J., Pier, J. R., Pope, A. C., Purger, N., Raddick, M. J., Re Fiorentin, P., Richards, G. T., Richmond, M. W., Riess, A. G., Rix, H.-W., Rockosi, C. M., Sako, M., Schlegel, D. J., Schneider, D. P., Schreiber, M. R., Schwope, A. D., Seljak, U., Sesar, B., Sheldon, E., Shimasaku, K., Sivarani, T., Smith, J. A., Snedden, S. A., Steinmetz, M., Strauss, M. A., SubbaRao, M., Suto, Y., Szalay, A. S., Szapudi, I., Szkody, P., Tegmark, M., Thakar, A. R., Tremonti, C. A., Tucker, D. L., Uomoto, A., Vanden Berk, D. E., Vandenberg, J., Vidrih, S., Vogeley, M. S., Voges, W., Vogt, N. P., Wadadekar, Y., Weinberg, D. H., West, A. A., White, S. D. M., Wilhite, B. C., Yanny, B., Yocum, D. R., York, D. G., Zehavi, I., and Zucker, D. B. (2008). The Sixth Data Release of the Sloan Digital Sky Survey. *ApJS*, **175**, 297–313.

Alexander, T. (1997). Is AGN Variability Correlated with Other AGN Properties? ZDCF Analysis of Small Samples of Sparse Light Curves. In D. Maoz, A. Sternberg, and E. M. Leibowitz, editors, *ASSL Vol. 218: Astronomical Time Series*, pages 163–166.

Almaini, O., Lawrence, A., Shanks, T., Edge, A., Boyle, B. J., Georgantopoulos, I., Gunn, K. F., Stewart, G. C., and Griffiths, R. E. (2000). X-ray variability in a deep, flux-limited sample of QSOs. *MNRAS*, **315**, 325–336.

Antonucci, R. (1993). Unified models for active galactic nuclei and quasars. *ARA&A*, **31**, 473–521.

Antonucci, R. R. J. and Miller, J. S. (1985). Spectropolarimetry and the nature of NGC 1068. *ApJ*, **297**, 621–632.

Arévalo, P., Papadakis, I., Kuhlbrodt, B., and Brinkmann, W. (2005). X-ray to UV variability correlation in MCG-6-30-15. *A&A*, **430**, 435–442.

Barcons, X., Carrera, F. J., Ceballos, M. T., Page, M. J., Bussons-Gordo, J., Corral, A., Ebrero, J., Mateos, S., Tedds, J. A., Watson, M. G., Baskill, D., Birkinshaw, M., Boller, T., Borisov, N., Bremer, M., Bromage, G. E., Brunner, H., Caccianiga, A., Crawford, C. S., Cropper, M. S., Della Ceca, R., Derry, P., Fabian, A. C., Guillout, P., Hashimoto, Y., Hasinger, G., Hassall, B. J. M., Lamer, G., Loaring, N. S., Maccacaro, T., Mason, K. O., McMahon, R. G., Mirioni, L., Mittaz, J. P. D., Motch, C., Negueruela, I., Osborne, J. P., Panessa, F., Pérez-Fournon, I., Pye, J. P., Roberts, T. P., Rosen, S., Schartel, N., Schurch, N., Schwope, A., Severgnini, P., Sharp, R., Stewart, G. C., Szokoly, G., Ullán, A., Ward, M. J., Warwick, R. S., Wheatley, P. J., Webb, N. A., Worrall, D., Yuan, W., and Ziaeepour, H. (2007). The XMM-Newton serendipitous survey. IV. Optical identification of the XMM-Newton medium sensitivity survey (XMS). *A&A*, **476**, 1191–1203.

Barger, A. J., Cowie, L. L., Capak, P., Alexander, D. M., Bauer, F. E., Fernandez, E., Brandt, W. N., Garmire, G. P., and Hornschemeier, A. E. (2003). Optical and Infrared Properties of the 2 Ms Chandra Deep Field North X-Ray Sources. *AJ*, **126**, 632–665.

Barger, A. J., Cowie, L. L., Mushotzky, R. F., Yang, Y., Wang, W.-H., Steffen, A. T., and Capak, P. (2005). The Cosmic Evolution of Hard X-Ray-selected Active Galactic Nuclei. *AJ*, **129**, 578–609.

Barr, P. and Mushotzky, R. F. (1986). Limits of X-ray variability in active galactic nuclei. *Nature*, **320**, 421–423.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge University Press, Cambridge.

Battye, R. A., Robinson, J., and Albrecht, A. (1998). Structure Formation by Cosmic Strings with a Cosmological Constant. *Physical Review Letters*, **80**, 4847–4850.

Bauer, F. E., Alexander, D. M., Brandt, W. N., Schneider, D. P., Treister, E., Hornschemeier, A. E., and Garmire, G. P. (2004). The Fall of Active Galactic Nuclei and the Rise of Star-forming Galaxies: A Close Look at the Chandra Deep Field X-Ray Number Counts. *AJ*, **128**, 2048–2065.

Baum, W. A. (1962). Photoelectric Magnitudes and Red-Shifts. In G. C. McVittie, editor, *Problems of Extra-Galactic Research*, volume 15 of *IAU Symposium*, pages 390–400.

Bekki, K. and Noguchi, M. (1994). Gas fueling to the central 10pc in merging galaxies. *A&A*, **290**, 7–18.

Bendat, J. S. and Piersol, A. G. (1986). *Random Data: Analysis and Measurement Procedures*. Wiley, New York.

Benítez, N. (2000). Bayesian Photometric Redshift Estimation. *ApJ*, **536**, 571–583.

Bian, W. and Zhao, Y. (2003). On X-ray variability in narrow-line and broad-line active galactic nuclei. *MNRAS*, **343**, 164–168.

Bolzonella, M., Miralles, J.-M., and Pelló, R. (2000). Photometric redshifts based on standard SED fitting procedures. *A&A*, **363**, 476–492.

Botte, V., Ciroi, S., di Mille, F., Rafanelli, P., and Romano, A. (2005). Stellar velocity dispersion in narrow-line Seyfert 1 galaxies. *MNRAS*, **356**, 789–793.

Box, G. E. P. and Jenkins, G. M., editors (1976). *Time series analysis. Forecasting and control*.

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting*. Springer, New York.

Budavári, T., Csabai, I., Szalay, A. S., Connolly, A. J., Szokoly, G. P., Vanden Berk, D. E., Richards, G. T., Weinstein, M. A., Schneider, D. P., Benítez, N., Brinkmann, J., Brunner, R., Hall, P. B., Hennessy, G. S., Ivezić, Ž., Kunszt, P. Z., Munn, J. A., Nichol, R. C., Pier, J. R., and York, D. G. (2001). Photometric Redshifts from Reconstructed Quasar Templates. *AJ*, **122**, 1163–1171.

Burrows, D. N., Hill, J. E., Nousek, J. A., Kennea, J. A., Wells, A., Osborne, J. P., Abbey, A. F., Beardmore, A., Mukerjee, K., Short, A. D. T., Chincarini, G., Campana, S., Citterio, O., Moretti, A., Pagani, C., Tagliaferri, G., Giommi, P., Capalbi, M., Tamburelli, F., Angelini, L., Cusumano, G., Bräuninger, H. W., Burkert, W., and Hartner, G. D. (2005). The Swift X-Ray Telescope. *Space Science Reviews*, **120**, 165–195.

Buzzoni, A. (2005). Broad-band colours and overall photometric properties of template galaxy models from stellar population synthesis. *MNRAS*, **361**, 725–742.

Carrera, F. J., Ebrero, J., Mateos, S., Ceballos, M. T., Corral, A., Barcons, X., Page, M. J., Rosen, S. R., Watson, M. G., Tedds, J. A., Della Ceca, R., Maccacaro, T., Brunner, H., Freyberg, M., Lamer, G., Bauer, F. E., and Ueda, Y. (2007). The XMM-Newton serendipitous survey. III. The AXIS X-ray source counts and angular clustering. *A&A*, **469**, 27–46.

Cash, W. (1979). Parameter estimation in astronomy through application of the likelihood ratio. *ApJ*, **228**, 939–947.

Clavel, J., Wamsteker, W., and Glass, I. S. (1989). Hot dust on the outskirts of the broad-line region in Fairall 9. *ApJ*, **337**, 236–250.

Coles, P. and Lucchin, F. (1995). *Cosmology. The origin and evolution of cosmic structure*. Wiley, Chichester.

Collister, A. A. and Lahav, O. (2004). ANNz: Estimating Photometric Redshifts Using

Artificial Neural Networks. *Publications of the Astronomical Society of the Pacific*, **116**, 345–351.

Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., and Munn, J. A. (1995). Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *AJ*, **110**, 2655–2664.

Croom, S. M., Shanks, T., Boyle, B. J., Smith, R. J., Miller, L., Loaring, N. S., and Hoyle, F. (2001). The 2dF QSO Redshift Survey - II. Structure and evolution at high redshift. *MNRAS*, **325**, 483–496.

Croom, S. M., Boyle, B. J., Shanks, T., Smith, R. J., Miller, L., Outram, P. J., Loaring, N. S., Hoyle, F., and da Ângela, J. (2005). The 2dF QSO Redshift Survey - XIV. Structure and evolution from the two-point correlation function. *MNRAS*, **356**, 415–438.

Cruddace, R. G., Hasinger, G. R., and Schmitt, J. H. (1988). The application of a maximum likelihood analysis to detection of sources in the ROSAT data base. In F. Murtagh and A. Heck, editors, *European Southern Observatory Astrophysics Symposia*, volume 28 of *European Southern Observatory Astrophysics Symposia*, pages 177–182.

Csabai, I., Budavári, T., Connolly, A. J., Szalay, A. S., Győry, Z., Benítez, N., Annis, J., Brinkmann, J., Eisenstein, D., Fukugita, M., Gunn, J., Kent, S., Lupton, R., Nichol, R. C., and Stoughton, C. (2003). The Application of Photometric Redshifts to the SDSS Early Data Release. *AJ*, **125**, 580–592.

Davis, M., Faber, S. M., Newman, J., Phillips, A. C., Ellis, R. S., Steidel, C. C., Conselice, C., Coil, A. L., Finkbeiner, D. P., Koo, D. C., Guhathakurta, P., Weiner, B., Schiavon, R., Willmer, C., Kaiser, N., Luppino, G. A., Wirth, G., Connolly, A., Eisenhardt, P., Cooper, M., and Gerke, B. (2003). Science Objectives and Early Results of the DEEP2 Redshift Survey. In P. Guhathakurta, editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II. Edited by Guhathakurta, Puragra. Proceedings of the SPIE, Volume 4834, pp. 161-172 (2003).*, volume 4834 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 161–172.

Davis, M., Guhathakurta, P., Konidaris, N. P., Newman, J. A., Ashby, M. L. N., Biggs, A. D., Barmby, P., Bundy, K., Chapman, S. C., Coil, A. L., Conselice, C. J., Cooper, M. C., Croton, D. J., Eisenhardt, P. R. M., Ellis, R. S., Faber, S. M., Fang, T., Fazio, G. G., Georgakakis, A., Gerke, B. F., Goss, W. M., Gwyn, S., Harker, J., Hopkins, A. M., Huang, J.-S., Ivison, R. J., Kassin, S. A., Kirby, E. N., Koekemoer, A. M., Koo, D. C., Laird, E. S., Le Floc'h, E., Lin, L., Lotz, J. M., Marshall, P. J., Martin, D. C., Metevier, A. J., Moustakas, L. A., Nandra, K., Noeske, K. G., Papovich, C., Phillips, A. C., Rich, R. M., Rieke, G. H., Rigopoulou, D., Salim, S., Schiminovich, D., Simard, L., Smail, I., Small, T. A., Weiner, B. J., Willmer, C. N. A., Willner, S. P., Wilson, G., Wright, E. L., and Yan, R. (2007). The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets. *ApJL*, **660**, L1–L6.

den Herder, J. W., Brinkman, A. C., Kahn, S. M., Branduardi-Raymont, G., Thomsen, K., Aarts, H., Audard, M., Bixler, J. V., den Boggende, A. J., Cottam, J., Decker, T., Dubbeldam, L., Erd, C., Goulooze, H., Güdel, M., Guttridge, P., Hailey, C. J., Janabi, K. A., Kaastra, J. S., de Korte, P. A. J., van Leeuwen, B. J., Mauche, C., McCalden, A. J., Mewe, R., Naber, A., Paerels, F. B., Peterson, J. R., Rasmussen, A. P., Rees, K., Sakelliou, I., Sako, M., Spodek, J., Stern, M., Tamura, T., Tandy, J., de Vries, C. P., Welch, S., and Zehnder, A. (2001). The Reflection Grating Spectrometer on board XMM-Newton. *A&A*, **365**, L7–L17.

Dunlop, J. S., McLure, R. J., Kukula, M. J., Baum, S. A., O'Dea, C. P., and Hughes, D. H. (2003). Quasars, their host galaxies and their central black holes. *MNRAS*, **340**, 1095–1135.

Edelson, R., Koratkar, A., Nandra, K., Goad, M., Peterson, B. M., Collier, S., Krolik, J., Malkan, M., Maoz, D., O'Brien, P., Shull, J. M., Vaughan, S., and Warwick, R. (2000). Intensive HST, RXTE, and ASCA Monitoring of NGC 3516: Evidence against Thermal Reprocessing. *ApJ*, **534**, 180–188.

Edelson, R. A. and Krolik, J. H. (1988). The discrete correlation function - A new method for analyzing unevenly sampled variability data. *ApJ*, **333**, 646–659.

Fan, X., Strauss, M. A., Schneider, D. P., Becker, R. H., White, R. L., Haiman, Z., Gregg, M., Pentericci, L., Grebel, E. K., Narayanan, V. K., Loh, Y.-S., Richards, G. T., Gunn, J. E., Lupton, R. H., Knapp, G. R., Ivezić, Ž., Brandt, W. N., Collinge, M., Hao, L., Harbeck, D., Prada, F., Schaye, J., Strateva, I., Zakamska, N., Anderson, S., Brinkmann, J., Bahcall, N. A., Lamb, D. Q., Okamura, S., Szalay, A., and York, D. G. (2003). A Survey of $z>5.7$ Quasars in the Sloan Digital Sky Survey. II. Discovery of Three Additional Quasars at $z>6$. *AJ*, **125**, 1649–1659.

Fanaroff, B. L. and Riley, J. M. (1974). The morphology of extragalactic radio sources of high and low luminosity. *MNRAS*, **167**, 31P–36P.

Forman, W., Jones, C., Cominsky, L., Julien, P., Murray, S., Peters, G., Tananbaum, H., and Giacconi, R. (1978). The fourth Uhuru catalog of X-ray sources. *ApJS*, **38**, 357–412.

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., and Schneider, D. P. (1996). The Sloan Digital Sky Survey Photometric System. *AJ*, **111**, 1748–1756.

Gandhi, P., Crawford, C. S., Fabian, A. C., and Johnstone, R. M. (2004). Powerful, obscured active galactic nuclei among X-ray hard, optically dim serendipitous Chandra sources. *MNRAS*, **348**, 529–550.

Gandhi, P., Garcet, O., Disseau, L., Pacaud, F., Pierre, M., Gueguen, A., Alloin, D., Chiappetti, L., Gosset, E., Maccagni, D., Surdej, J., and Valtchanov, I. (2006). The XMM large scale structure survey: properties and two-point angular correlations of point-like sources. *A&A*, **457**, 393–404.

Gaskell, C. M. and Peterson, B. M. (1987). The accuracy of cross-correlation estimates of quasar emission-line region sizes. *ApJS*, **65**, 1–11.

Gendreau, K. C., Barcons, X., and Fabian, A. C. (1998). Deep hard X-ray source counts from a fluctuation analysis of ASCA SIS images. *MNRAS*, **297**, 41–48.

Giacconi, R., Gursky, H., Paolini, F. R., and Rossi, B. B. (1962). Evidence for x Rays From Sources Outside the Solar System. *Physical Review Letters*, **9**, 439–443.

Giacconi, R., Bechtold, J., Branduardi, G., Forman, W., Henry, J. P., Jones, C., Kellogg, E., van der Laan, H., Liller, W., Marshall, H., Murray, S. S., Pye, J., Schreier, E., Sargent, W. L. W., Seward, F., and Tananbaum, H. (1979). A high-sensitivity X-ray survey using the Einstein Observatory and the discrete source contribution to the extragalactic X-ray background. *ApJL*, **234**, L1–L7.

Gierliński, M., Middleton, M., Ward, M., and Done, C. (2008). A periodicity of ˜1hour in X-ray emission from the active galaxy RE J1034+396. *Nature*, **455**, 369–371.

Gilli, R., Salvati, M., and Hasinger, G. (2001). Testing current synthesis models of the X-ray background. *A&A*, **366**, 407–417.

Gilli, R., Daddi, E., Zamorani, G., Tozzi, P., Borgani, S., Bergeron, J., Giacconi, R., Hasinger, G., Mainieri, V., Norman, C., Rosati, P., Szokoly, G., and Zheng, W. (2005). The spatial clustering of X-ray selected AGN and galaxies in the Chandra Deep Field South and North. *A&A*, **430**, 811–825.

Gonzalez, A. H. and Maccarone, T. J. (2002). A Test of Photometric Redshifts for X-Ray-selected Sources. *ApJ*, **581**, 155–160.

Grazian, A., Negrello, M., Moscardini, L., Cristiani, S., Haehnelt, M. G., Matarrese, S., Omizzolo, A., and Vanzella, E. (2004). The Asiago-ESO/RASS QSO Survey. III. Clustering Analysis and Theoretical Interpretation. *AJ*, **127**, 592–605.

Green, A. R., McHardy, I. M., and Lehto, H. J. (1993). On the nature of rapid X-ray variability in active galactic nuclei. *MNRAS*, **265**, 664–680.

Guilbert, P. W. and Rees, M. J. (1988). 'Cold' material in non-thermal sources. *MNRAS*, **233**, 475–484.

Gunn, J. E., Siegmund, W. A., Mannery, E. J., Owen, R. E., Hull, C. L., Leger, R. F., Carey, L. N., Knapp, G. R., York, D. G., Boroski, W. N., Kent, S. M., Lupton, R. H., Rockosi, C. M., Evans, M. L., Waddell, P., Anderson, J. E., Annis, J., Barentine, J. C., Bartoszek, L. M., Bastian, S., Bracker, S. B., Brewington, H. J., Briegel, C. I., Brinkmann, J., Brown, Y. J., Carr, M. A., Czarapata, P. C., Drennan, C. C., Dombeck, T., Federwitz, G. R., Gillespie, B. A., Gonzales, C., Hansen, S. U., Harvanek, M., Hayes, J., Jordan, W., Kinney, E., Klaene, M., Kleinman, S. J., Kron, R. G., Kresinski, J., Lee, G., Limmongkol, S., Lindenmeyer, C. W., Long, D. C., Loomis, C. L., McGehee, P. M., Mantsch, P. M., Neilsen, Jr., E. H., Neswold, R. M., Newman, P. R., Nitta, A., Peoples, J. J., Pier, J. R., Prieto, P. S., Prosapio, A., Rivetta, C., Schneider, D. P., Snedden, S., and Wang, S.-i. (2006). The 2.5 m Telescope of the Sloan Digital Sky Survey. *AJ*, **131**, 2332–2359.

Guth, A. H. and Pi, S.-Y. (1982). Fluctuations in the new inflationary universe. *Physical Review Letters*, **49**, 1110–1113.

Haardt, F. and Maraschi, L. (1993). X-ray spectra from two-phase accretion disks. *ApJ*, **413**, 507–517.

Hasinger, G., Burg, R., Giacconi, R., Schmidt, M., Trumper, J., and Zamorani, G. (1998). The ROSAT Deep Survey. I. X-ray sources in the Lockman Field. *A&A*, **329**, 482–494.

Hasinger, G., Miyaji, T., and Schmidt, M. (2005). Luminosity-dependent evolution of soft X-ray selected AGN. New Chandra and XMM-Newton surveys. *A&A*, **441**, 417–434.

Hasinger, G., Cappelluti, N., Brunner, H., Brusa, M., Comastri, A., Elvis, M., Finoguenov, A., Fiore, F., Franceschini, A., Gilli, R., Griffiths, R. E., Lehmann, I., Mainieri, V., Matt, G., Matute, I., Miyaji, T., Molendi, S., Paltani, S., Sanders, D. B., Scoville, N., Tresse, L., Urry, C. M., Vettolani, P., and Zamorani, G. (2007). The XMM-Newton Wide-Field Survey in the COSMOS Field. I. Survey Description. *ApJS*, **172**, 29–37.

Jansen, F., Lumb, D., Altieri, B., Clavel, J., Ehle, M., Erd, C., Gabriel, C., Guainazzi, M., Gondoin, P., Much, R., Munoz, R., Santos, M., Schartel, N., Texier, D., and Vacanti, G. (2001). XMM-Newton observatory. I. The spacecraft and operations. *A&A*, **365**, L1–L6.

Jenkins, G. M. and Watts, D. G. (1968). *Spectral analysis and its applications*. Holden-Day, San Francisco.

Kendall, M. and Ord, J. K. (1955). *Time Series, 3rd ed.* Edward Arnold, London.

Kim, D.-W., Cameron, R. A., Drake, J. J., Evans, N. R., Freeman, P., Gaetz, T. J., Ghosh, H., Green, P. J., Harnden, Jr., F. R., Karovska, M., Kashyap, V., Maksym, P. W., Ratzlaff, P. W., Schlegel, E. M., Silverman, J. D., Tananbaum, H. D., Vikhlinin, A. A., Wilkes, B. J., and Grimes, J. P. (2004a). Chandra Multiwavelength Project. I. First X-Ray Source Catalog. *ApJS*, **150**, 19–41.

Kim, D.-W., Wilkes, B. J., Green, P. J., Cameron, R. A., Drake, J. J., Evans, N. R., Freeman, P., Gaetz, T. J., Ghosh, H., Harnden, Jr., F. R., Karovska, M., Kashyap, V., Maksym, P. W., Ratzlaff, P. W., Schlegel, E. M., Silverman, J. D., Tananbaum, H. D., and Vikhlinin, A. A. (2004b). Chandra Multiwavelength Project. II. First Results of X-Ray Source Properties. *ApJ*, **600**, 59–69.

Kitsionas, S., Hatziminaoglou, E., Georgakakis, A., and Georgantopoulos, I. (2005). On the use of photometric redshifts for X-ray selected AGNs. *A&A*, **434**, 475–482.

Koo, D. C. (1985). Optical multicolors - A poor person's Z machine for galaxies. *AJ*, **90**, 418–440.

Koratkar, A. and Blaes, O. (1999). The Ultraviolet and Optical Continuum Emission in Active Galactic Nuclei: The Status of Accretion Disks. *Publications of the Astronomical Society of the Pacific*, **111**, 1–30.

La Franca, F., Fiore, F., Comastri, A., Perola, G. C., Sacchi, N., Brusa, M., Cocchia, F., Feruglio, C., Matt, G., Vignali, C., Carangelo, N., Ciliegi, P., Lamastra, A., Maiolino, R., Mignoli, M., Molendi, S., and Puccetti, S. (2005). The HELLAS2XMM Survey. VII. The Hard X-Ray Luminosity Function of AGNs up to z = 4: More Absorbed AGNs at Low Luminosities and High Redshifts. *ApJ*, **635**, 864–879.

Landy, S. D. and Szalay, A. S. (1993). Bias and variance of angular correlation functions. *ApJ*, **412**, 64–71.

Lawrence, A. and Papadakis, I. (1993). X-ray variability of active galactic nuclei - A universal power spectrum with luminosity-dependent amplitude. *ApJL*, **414**, L85–L88.

Lawrence, A., Warren, S. J., Almaini, O., Edge, A. C., Hambly, N. C., Jameson, R. F., Lucas, P., Casali, M., Adamson, A., Dye, S., Emerson, J. P., Foucaud, S., Hewett, P., Hirst, P., Hodgkin, S. T., Irwin, M. J., Lodieu, N., McMahon, R. G., Simpson, C., Smail, I., Mortlock, D., and Folger, M. (2007). The UKIRT Infrared Deep Sky Survey (UKIDSS). *MNRAS*, **379**, 1599–1617.

Lehmann, I., Hasinger, G., Schmidt, M., Giacconi, R., Trümper, J., Zamorani, G., Gunn, J. E., Pozzetti, L., Schneider, D. P., Stanke, T., Szokoly, G., Thompson, D., and Wilson, G. (2001). The ROSAT Deep Survey. VI. X-ray sources and Optical identifications of the Ultra Deep Survey. *A&A*, **371**, 833–857.

Limber, D. N. (1953). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *ApJ*, **117**, 134–144.

Lu, Y. and Yu, Q. (2001). The relationship between X-ray variability and the central black hole mass. *MNRAS*, **324**, 653–658.

Lynden-Bell, D. (1969). Galactic Nuclei as Collapsed Old Quasars. *Nature*, **223**, 690–694.

MacKenty, J. W. (1990). Seyfert galaxies. I - Morphologies, magnitudes, and disks. *ApJS*, **72**, 231–244.

Markowitz, A. and Edelson, R. (2004). An Expanded Rossi X-Ray Timing Explorer Survey of X-Ray Variability in Seyfert 1 Galaxies. *ApJ*, **617**, 939–965.

Markowitz, A., Edelson, R., Vaughan, S., Uttley, P., George, I. M., Griffiths, R. E., Kaspi, S., Lawrence, A., McHardy, I., Nandra, K., Pounds, K., Reeves, J., Schurch, N., and Warwick, R. (2003). X-Ray Fluctuation Power Spectral Densities of Seyfert 1 Galaxies. *ApJ*, **593**, 96–114.

Marshall, N., Warwick, R. S., and Pounds, K. A. (1981). The variability of X-ray emission from active galaxies. *MNRAS*, **194**, 987–1002.

Mason, K. O., Breeveld, A., Much, R., Carter, M., Cordova, F. A., Cropper, M. S., Fordham, J., Huckle, H., Ho, C., Kawakami, H., Kennea, J., Kennedy, T., Mittaz, J., Pandel, D., Priedhorsky, W. C., Sasseen, T., Shirey, R., Smith, P., and Vreux, J.-M. (2001). The XMM-Newton optical/UV monitor telescope. *A&A*, **365**, L36–L44.

Mason, K. O., McHardy, I. M., Page, M. J., Uttley, P., Córdova, F. A., Maraschi, L., Priedhorsky, W. C., Puchnarewicz, E. M., and Sasseen, T. (2002). XMM-Newton Observations of a Possible Light Echo in the Seyfert 1 Nucleus of NGC 4051. *ApJL*, **580**, L117–L120.

Mateos, S., Barcons, X., Carrera, F. J., Ceballos, M. T., Hasinger, G., Lehmann, I., Fabian, A. C., and Streblyanska, A. (2005). XMM-Newton observations of the Lockman Hole IV: spectra of the brightest AGN. *A&A*, **444**, 79–99.

Mateos, S., Warwick, R. S., Carrera, F. J., Stewart, G. C., Ebrero, J., Della Ceca, R., Caccianiga, A., Gilli, R., Page, M. J., Treister, E., Tedds, J. A., Watson, M. G., Lamer, G., Saxton, R. D., Brunner, H., and Page, C. G. (2008). High precision X-ray log N - log S distributions: implications for the obscured AGN population. *A&A*, **492**, 51–69.

McHardy, I. M., Gunn, K. F., Uttley, P., and Goad, M. R. (2005). MCG-6-30-15: long time-scale X-ray variability, black hole mass and active galactic nuclei high states. *MNRAS*, **359**, 1469–1480.

Miyaji, T., Zamorani, G., Cappelluti, N., Gilli, R., Griffiths, R. E., Comastri, A., Hasinger, G., Brusa, M., Fiore, F., Puccetti, S., Guzzo, L., and Finoguenov, A. (2007). The XMM-Newton Wide-Field Survey in the COSMOS Field. V. Angular Clustering of the X-Ray Point Sources. *ApJS*, **172**, 396–405.

Myers, A. D., Brunner, R. J., Nichol, R. C., Richards, G. T., Schneider, D. P., and Bahcall, N. A. (2007). Clustering Analyses of 300,000 Photometrically Classified Quasars. I. Luminosity and Redshift Evolution in Quasar Bias. *ApJ*, **658**, 85–98.

Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., and Yaqoob, T. (1997). ASCA Observations of Seyfert 1 Galaxies. I. Data Analysis, Imaging, and Timing. *ApJ*, **476**, 70–82.

Nandra, K., Clavel, J., Edelson, R. A., George, I. M., Malkan, M. A., Mushotzky, R. F., Peterson, B. M., and Turner, T. J. (1998). New Constraints on the Continuum Emission Mechanism of Active Galactic Nuclei: Intensive Monitoring of NGC 7469 in the X-Ray and Ultraviolet. *ApJ*, **505**, 594–606.

Nandra, K., Le, T., George, I. M., Edelson, R. A., Mushotzky, R. F., Peterson, B. M., and Turner, T. J. (2000). The Origin of the X-Ray and Ultraviolet Emission in NGC 7469. *ApJ*, **544**, 734–746.

Norman, C., Hasinger, G., Giacconi, R., Gilli, R., Kewley, L., Nonino, M., Rosati, P., Szokoly, G., Tozzi, P., Wang, J., Zheng, W., Zirm, A., Bergeron, J., Gilmozzi, R., Grogin, N., Koekemoer, A., and Schreier, E. (2002). A Classic Type 2 QSO. *ApJ*, **571**, 218–225.

Osterbrock, D. E. (1981). Seyfert galaxies with weak broad H alpha emission lines. *ApJ*, **249**, 462–470.

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., Frieman, J., and Sheldon, E. S. (2008). A Galaxy Photometric Redshift Catalog for the Sloan Digital Sky Survey Data Release 6. *ApJ*, **674**, 768–783.

Papadakis, I. E., Brinkmann, W., Negoro, H., Detsis, E., Papamastorakis, I., and Gliozzi, M. (2000). Optical and X-ray monitoring of the NLS1 galaxy Ark 564. *ArXiv e-prints 0012317*.

Peterson, B. M. (1993). Reverberation mapping of active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, **105**, 247–268.

Peterson, B. M. (1997). *An Introduction to Active Galactic Nuclei*. Cambridge University Press, Cambridge.

Peterson, B. M., McHardy, I. M., Wilkes, B. J., Berlind, P., Bertram, R., Calkins, M., Collier, S. J., Huchra, J. P., Mathur, S., Papadakis, I., Peters, J., Pogge, R. W., Romano, P., Tokarz, S., Uttley, P., Vestergaard, M., and Wagner, R. M. (2000). X-Ray and Optical Variability in NGC 4051 and the Nature of Narrow-Line Seyfert 1 Galaxies. *ApJ*, **542**, 161–174.

Peterson, B. M., Ferrarese, L., Gilbert, K. M., Kaspi, S., Malkan, M. A., Maoz, D., Merritt, D., Netzer, H., Onken, C. A., Pogge, R. W., Vestergaard, M., and Wandel, A. (2004). Central Masses and Broad-Line Region Sizes of Active Galactic Nuclei. II. A Homogeneous Analysis of a Large Reverberation-Mapping Database. *ApJ*, **613**, 682–699.

Pierre, M., Valtchanov, I., Altieri, B., and et al. (2004). The XMM-LSS survey. Survey design and first results. *Journal of Cosmology and Astro-Particle Physics*, **9**, 11.

Priestley, M. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

Ratcliffe, A., Shanks, T., Parker, Q. A., and Fong, R. (1998). The Durham/UKST Galaxy Redshift Survey - III. Large-scale structure via the two-point correlation function. *MNRAS*, **296**, 173–190.

Reynolds, C. S. (1998). Compton reflection and iron fluorescence in AGN and GBHCs. *ArXiv e-prints 9810018*.

Richards, G. T., Weinstein, M. A., Schneider, D. P., Fan, X., Strauss, M. A., Vanden Berk, D. E., Annis, J., Burles, S., Laubacher, E. M., York, D. G., Frieman, J. A., Johnston, D., Scranton, R., Gunn, J. E., Ivezić, Ž., Nichol, R. C., Budavári, T., Csabai, I., Szalay, A. S., Connolly, A. J., Szokoly, G. P., Bahcall, N. A., Benítez, N., Brinkmann, J., Brunner, R., Fukugita, M., Hall, P. B., Hennessy, G. S., Knapp, G. R., Kunszt, P. Z., Lamb, D. Q., Munn, J. A., Newberg, H. J., and Stoughton, C. (2001). Photometric Redshifts of Quasars. *AJ*, **122**, 1151–1162.

Richards, G. T., Strauss, M. A., Fan, X., Hall, P. B., Jester, S., Schneider, D. P., Vanden

Berk, D. E., Stoughton, C., Anderson, S. F., Brunner, R. J., Gray, J., Gunn, J. E., Ivezić, Ž., Kirkland, M. K., Knapp, G. R., Loveday, J., Meiksin, A., Pope, A., Szalay, A. S., Thakar, A. R., Yanny, B., York, D. G., Barentine, J. C., Brewington, H. J., Brinkmann, J., Fukugita, M., Harvanek, M., Kent, S. M., Kleinman, S. J., Krzesiński, J., Long, D. C., Lupton, R. H., Nash, T., Neilsen, Jr., E. H., Nitta, A., Schlegel, D. J., and Snedden, S. A. (2006). The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3. *AJ*, **131**, 2766–2787.

Roming, P. W. A., Kennedy, T. E., Mason, K. O., Nousek, J. A., Ahr, L., Bingham, R. E., Broos, P. S., Carter, M. J., Hancock, B. K., Huckle, H. E., Hunsberger, S. D., Kawakami, H., Killough, R., Koch, T. S., McLelland, M. K., Smith, K., Smith, P. J., Soto, J. C., Boyd, P. T., Breeveld, A. A., Holland, S. T., Ivanushkina, M., Pryzby, M. S., Still, M. D., and Stock, J. (2005). The Swift Ultra-Violet/Optical Telescope. *Space Science Reviews*, **120**, 95–142.

Rutledge, R. E., Brunner, R. J., Prince, T. A., and Lonsdale, C. (2000). XID: Cross-Association of ROSAT/Bright Source Catalog X-Ray Sources with USNO A-2 Optical Point Sources. *ApJS*, **131**, 335–353.

Schmidt, M. (1963). 3C 273 : A Star-Like Object with Large Red-Shift. *Nature*, **197**, 1040.

Schmidt, M., Hasinger, G., Gunn, J., Schneider, D., Burg, R., Giacconi, R., Lehmann, I., MacKenty, J., Trumper, J., and Zamorani, G. (1998). The ROSAT deep survey. II. Optical identification, photometry and spectra of X-ray sources in the Lockman field. *A&A*, **329**, 495–503.

Schneider, D. P., Hall, P. B., Richards, G. T., Strauss, M. A., Vanden Berk, D. E., Anderson, S. F., Brandt, W. N., Fan, X., Jester, S., Gray, J., Gunn, J. E., SubbaRao, M. U., Thakar, A. R., Stoughton, C., Szalay, A. S., Yanny, B., York, D. G., Bahcall, N. A., Barentine, J., Blanton, M. R., Brewington, H., Brinkmann, J., Brunner, R. J., Castander, F. J., Csabai, I., Frieman, J. A., Fukugita, M., Harvanek, M., Hogg, D. W., Ivezić, Ž., Kent, S. M., Kleinman, S. J., Knapp, G. R., Kron, R. G., Krzesiński, J., Long, D. C., Lupton, R. H., Nitta, A., Pier, J. R., Saxe, D. H., Shen, Y., Snedden, S. A., Weinberg, D. H., and Wu, J. (2007). The Sloan Digital Sky Survey Quasar Catalog. IV. Fifth Data Release. *AJ*, **134**, 102–117.

Seyfert, C. K. (1943). Nuclear Emission in Spiral Nebulae. *ApJ*, **97**, 28–40.

Shemmer, O., Romano, P., Bertram, R., Brinkmann, W., Collier, S., Crowley, K. A., Detsis, E., Filippenko, A. V., Gaskell, C. M., George, T. A., Gliozzi, M., Hiller, M. E., Jewell, T. L., Kaspi, S., Klimek, E. S., Lannon, M. H., Li, W., Martini, P., Mathur, S., Negoro, H., Netzer, H., Papadakis, I., Papamastorakis, I., Peterson, B. M., Peterson, B. W., Pogge, R. W., Pronik, V. I., Rumstay, K. S., Sergeev, S. G., Sergeeva, E. A., Stirpe, G. M., Taylor, C. J., Treffers, R. R., Turner, T. J., Uttley, P., Vestergaard, M., von Braun, K., Wagner, R. M., and Zheng, Z. (2001). Multiwavelength Monitoring of the Narrow-Line Seyfert 1 Galaxy Arakelian 564. III. Optical Observations and the Optical-UV-X-Ray Connection. *ApJ*, **561**, 162–170.

Shemmer, O., Uttley, P., Netzer, H., and McHardy, I. M. (2003). Complex optical-X-ray correlations in the narrow-line Seyfert 1 galaxy NGC 4051. *MNRAS*, **343**, 1341–1347.

Sheth, R. K. and Tormen, G. (1999). Large-scale bias and the peak background split. *MNRAS*, **308**, 119–126.

Sheth, R. K., Mo, H. J., and Tormen, G. (2001). Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, **323**, 1–12.

Simkin, S. M., Su, H. J., and Schwarz, M. P. (1980). Nearby Seyfert galaxies. *ApJ*, **237**, 404–413.

Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *ApJS*, **148**, 175–194.

Stern, D., Moran, E. C., Coil, A. L., Connolly, A., Davis, M., Dawson, S., Dey, A., Eisenhardt, P., Elston, R., Graham, J. R., Harrison, F., Helfand, D. J., Holden, B., Mao, P., Rosati, P., Spinrad, H., Stanford, S. A., Tozzi, P., and Wu, K. L. (2002). Chandra Detection of a Type II Quasar at z = 3.288. *ApJ*, **568**, 71–81.

Strüder, L., Briel, U., Dennerl, K., Hartmann, R., Kendziorra, E., Meidinger, N., Pfeffermann, E., Reppin, C., Aschenbach, B., Bornemann, W., Bräuninger, H., Burkert, W., Elender, M., Freyberg, M., Haberl, F., Hartner, G., Heuschmann, F., Hippmann, H., Kastelic, E., Kemmer, S., Kettenring, G., Kink, W., Krause, N., Müller, S., Oppitz, A., Pietsch, W., Popp, M., Predehl, P., Read, A., Stephan, K. H., Stötter, D., Trümper, J., Holl, P., Kemmer, J., Soltau, H., Stötter, R., Weber, U., Weichert, U., von Zanthier, C., Carathanassis, D., Lutz, G., Richter, R. H., Solc, P., Böttcher, H., Kuster, M., Staubert, R., Abbey, A., Holland, A., Turner, M., Balasini, M., Bignami, G. F., La Palombara, N., Villa, G., Buttler, W., Gianini, F., Lainé, R., Lumb, D., and Dhez, P. (2001). The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera. *A&A*, **365**, L18–L26.

Sunyaev, R. A. and Titarchuk, L. G. (1980). Comptonization of X-rays in plasma clouds - Typical radiation spectra. *A&A*, **86**, 121–138.

Tedds, J. A., Page, M. J., and Survey Science Centre, X.-N. (2006). 2dF-XMM Wide Angle Serendipitous Survey. In A. Wilson, editor, *The X-ray Universe 2005*, volume 604 of *ESA Special Publication*, pages 843–844.

Tegmark, M., Strauss, M. A., Blanton, M. R., Abazajian, K., Dodelson, S., Sandvik, H., Wang, X., Weinberg, D. H., Zehavi, I., Bahcall, N. A., Hoyle, F., Schlegel, D., Scoccimarro, R., Vogeley, M. S., Berlind, A., Budavari, T., Connolly, A., Eisenstein, D. J., Finkbeiner, D., Frieman, J. A., Gunn, J. E., Hui, L., Jain, B., Johnston, D., Kent, S., Lin, H., Nakajima, R., Nichol, R. C., Ostriker, J. P., Pope, A., Scranton, R., Seljak, U., Sheth, R. K., Stebbins, A., Szalay, A. S., Szapudi, I., Xu, Y., Annis, J., Brinkmann, J., Burles, S., Castander, F. J., Csabai, I., Loveday, J., Doi, M., Fukugita, M., Gillespie, B., Hennessy, G., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., Lee, B. C., Lupton, R. H., McKay, T. A., Kunszt, P., Munn, J. A., O'Connell, L., Peoples, J., Pier, J. R., Richmond, M., Rockosi, C., Schneider, D. P., Stoughton, C., Tucker, D. L., vanden Berk, D. E., Yanny, B., and York, D. G. (2004). Cosmological parameters from SDSS and WMAP. *Physics Review D*, **69**(10), 103501.

Tucker, D. L., Oemler, Jr., A., Kirshner, R. P., Lin, H., Shectman, S. A., Landy, S. D., Schechter, P. L., Muller, V., Gottlober, S., and Einasto, J. (1997). The Las Campanas Redshift Survey galaxy-galaxy autocorrelation function. *MNRAS*, **285**, L5–L9.

Turner, M. J. L., Abbey, A., Arnaud, M., Balasini, M., Barbera, M., Belsole, E., Bennie, P. J., Bernard, J. P., Bignami, G. F., Boer, M., Briel, U., Butler, I., Cara, C., Chabaud, C., Cole, R., Collura, A., Conte, M., Cros, A., Denby, M., Dhez, P., Di Coco, G., Dowson, J., Ferrando, P., Ghizzardi, S., Gianotti, F., Goodall, C. V., Gretton, L., Griffiths, R. G., Hainaut, O., Hochedez, J. F., Holland, A. D., Jourdain, E., Kendziorra, E., Lagostina, A., Laine, R., La Palombara, N., Lortholary, M., Lumb, D., Marty, P., Molendi, S., Pigot, C., Poindron, E., Pounds, K. A., Reeves, J. N., Reppin, C., Rothenflug, R., Salvetat, P., Sauvageot, J. L., Schmitt, D., Sembay, S., Short, A. D. T., Spragg, J., Stephen, J., Strüder, L., Tiengo, A., Trifoglio, M., Trümper, J., Vercellone, S., Vigroux, L., Villa, G., Ward, M. J., Whitehead, S., and Zonca, E. (2001). The European Photon Imaging Camera on XMM-Newton: The MOS cameras : The MOS cameras. *A&A*, **365**, L27–L35.

Turner, T. J., George, I. M., Nandra, K., and Turcan, D. (1999). On X-Ray Variability in Seyfert Galaxies. *ApJ*, **524**, 667–673.

Urry, C. M. and Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, **107**, 803–845.

Uttley, P. (2005). The relation between optical and X-ray variability in Seyfert galaxies. *ArXiv e-prints 0501157*.

Uttley, P., Edelson, R., McHardy, I. M., Peterson, B. M., and Markowitz, A. (2003). Correlated Long-Term Optical and X-Ray Variations in NGC 5548. *ApJL*, **584**, L53–L56.

Vaughan, S., Edelson, R., Warwick, R. S., and Uttley, P. (2003a). On characterizing the variability properties of X-ray light curves from active galaxies. *MNRAS*, **345**, 1271–1284.

Vaughan, S., Fabian, A. C., and Nandra, K. (2003b). X-ray continuum variability of MCG-6-30-15. *MNRAS*, **339**, 1237–1255.

Wake, D. A., Miller, C. J., Di Matteo, T., Nichol, R. C., Pope, A., Szalay, A. S., Gray, A., Schneider, D. P., and York, D. G. (2004). The Clustering of Active Galactic Nuclei in the Sloan Digital Sky Survey. *ApJL*, **610**, L85–L88.

Wang, D., Zhang, Y. X., Liu, C., and Zhao, Y. H. (2007). Kernel regression for determining photometric redshifts from Sloan broad-band photometry. *MNRAS*, **382**, 1601–1606.

Watson, M. G., Schröder, A. C., Fyfe, D., Page, C. G., Lamer, G., Mateos, S., Pye, J., Sakano, M., Rosen, S., Ballet, J., Barcons, X., Barret, D., Boller, T., Brunner, H., Brusa, M., Caccianiga, A., Carrera, F. J., Ceballos, M., Della Ceca, R., Denby, M., Denkinson, G., Dupuy, S., Farrell, S., Fraschetti, F., Freyberg, M. J., Guillout, P., Hambaryan, V., Maccacaro, T., Mathiesen, B., McMahon, R., Michel, L., Motch, C., Osborne, J. P., Page, M., Pakull, M. W., Pietsch, W., Saxton, R., Schwope, A., Severgnini, P., Simpson, M., Sironi, G., Stewart, G., Stewart, I. M., Stobbart, A.-M., Tedds, J., Warwick, R., Webb, N., West, R., Worrall, D., and Yuan, W. (2009). The XMM-Newton serendipitous survey. V. The Second XMM-Newton serendipitous source catalogue. *A&A*, **493**, 339–373.

Weinstein, M. A., Richards, G. T., Schneider, D. P., Younger, J. D., Strauss, M. A., Hall, P. B., Budavári, T., Gunn, J. E., York, D. G., and Brinkmann, J. (2004). An Empirical Algorithm for Broadband Photometric Redshifts of Quasars from the Sloan Digital Sky Survey. *ApJS*, **155**, 243–256.

Welsh, W. F. (1999). On the Reliability of Cross-Correlation Function Lag Determinations in Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, **111**, 1347–1366.

Wu, X.-B., Zhang, W., and Zhou, X. (2004). Color-Redshift Relations and Photometric Redshift Estimations of Quasars in Large Sky Surveys. *Chinese Journal of Astronomy and Astrophysics*, **4**, 17–27.

Yang, Y., Mushotzky, R. F., Barger, A. J., Cowie, L. L., Sanders, D. B., and Steffen, A. T. (2003). Imaging Large-Scale Structure in the X-Ray Sky. *ApJL*, **585**, L85–L88.

Yang, Y., Mushotzky, R. F., Barger, A. J., and Cowie, L. L. (2006). Spatial Correlation Function of the Chandra-selected Active Galactic Nuclei. *ApJ*, **645**, 68–82.

Zheng, W., Mikles, V. J., Mainieri, V., Hasinger, G., Rosati, P., Wolf, C., Norman, C., Szokoly, G., Gilli, R., Tozzi, P., Wang, J. X., Zirm, A., and Giacconi, R. (2004).

Photometric Redshift of X-Ray Sources in the Chandra Deep Field-South. *ApJS*, **155**, 73–87.

Zhou, X.-L. and Wang, J.-M. (2005). Narrow Iron Kα Lines in Active Galactic Nuclei: Evolving Populations? *ApJL*, **618**, L83–L86.