Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **423,** 2633–2641 (2012) doi:10.1111/j.1365-2966.2012.21069.x

Detectability of low-energy X-ray spectral components in type 1 active galactic nuclei

A. E. Scott,[1⋆] G. C. Stewart[1] and S. Mateos[1,2]

[1]*Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH*
[2]*Instituto de Fisica de Cantabria (CSIC-UC), Avenida de los Castros, 39005 Santander, Spain*

Accepted 2012 April 5. Received 2012 April 4; in original form 2011 November 18

ABSTRACT

In this paper we examine the percentage of type 1 active galactic nuclei (AGN) which require the inclusion of a soft excess component and/or significant cold absorption in the modelling of their X-ray spectra obtained by *XMM–Newton*. We do this by simulating spectra which mimic typical spectral shapes in order to find the maximum detectability expected at different count levels. We then apply a correction to the observed percentages found for the Scott et al. sample of 761 sources. We estimate the true percentage of AGN with a soft excess component to be 75 ± 23 per cent, suggesting that soft excesses are ubiquitous in the X-ray spectra of type 1 AGN. By carrying out joint fits on groups of low count spectra in narrow z bins in which additional spectral components were not originally detected, we show that the soft excess feature is recovered with a mean temperature kT and blackbody-to-power-law normalization ratio consistent with those of components detected in individual high count spectra. Cold absorption with $N_{\rm H}$ values broadly consistent with those reported in individual spectra is also recovered. We suggest such intrinsic cold absorption is found in a minimum of ∼5 per cent of type 1 AGN and may be present in up to ∼10 per cent.

Key words: galaxies: active – quasars: general – X-rays: galaxies.

1 INTRODUCTION

The X-ray spectral properties of active galactic nuclei (AGN) classified optically as type 1 have been recently extensively studied (e.g. Young, Elvis & Risaliti 2009; Mateos et al. 2010; Corral et al. 2011; Scott et al. 2011, hereafter S11). The underlying spectrum in the 0.5–12.0 keV band consists of a power law of photon index $\Gamma \sim 2$ thought to be produced by the inverse-Compton scattering of low-energy disc photons by a corona of relativistic electrons (Haardt & Maraschi 1993). In higher quality spectra, a soft excess component is detected, rising above the power law at rest-frame energies $\lesssim$1 keV (Arnaud et al. 1985). This was originally interpreted as the hard tail of the 'big blue bump', accretion disc emission seen in the ultraviolet (UV), but is now thought perhaps to be an artefact of ionized absorption (Gierliński & Done 2004) or ionized reflection (Ross & Fabian 2005; Crummy et al. 2006). Lower energies are also affected by photoelectric absorption, although the standard orientation-based unified model (Antonucci 1993) does not predict any intrinsic X-ray absorption to be present in objects which have been classified as type 1 due to the presence of broad lines in their UV/optical spectra.

There have been many studies in which some type 1 objects have shown evidence for intrinsic X-ray absorption. The typical percentage of such objects is ∼10 per cent, with many studies quoting this similar value (Page et al. 2003; Perola et al. 2004; Mateos et al. 2005b; Piconcelli et al. 2005; Garcet et al. 2007; Mainieri et al. 2007; Young et al. 2009; Mateos et al. 2010; Corral et al. 2011).

There are a large range of values quoted for the percentage of type 1 sources which include a soft excess. The earliest study with *EXOSAT* suggested 30–50 per cent of objects included the component (Turner & Pounds 1989) and a study with *ASCA* found ∼40 per cent (Reeves & Turner 2000). It has also been suggested that the soft excess may be a ubiquitous feature in optically selected PG quasars (Porquet et al. 2004; Piconcelli et al. 2005); however, these samples are biased towards bright and low-redshift sources and Winter et al. (2012) find the detection rate increases from ∼40% to ∼90% when only the highest quality spectra of low z sources are considered. The quoted percentage of soft excesses can be very different depending upon the redshift range being considered. For example, Mateos et al. (2010) find a percentage of only 8 per cent when considering their entire sample, but this is increased to 36 per cent when only sources at $z < 0.5$ are considered. Similarly the CAIXA sample of *XMM–Newton* target sources finds a high percentage of ∼80 per cent (Bianchi et al. 2009). This could be because the sample is biased towards low-redshift objects and/or good-quality spectra in which detecting the spectral component is easier. Clearly, in order to determine whether the soft excess is present in all sources, the influences of redshift and spectral quality need to be taken into account, using a sample which covers a large range in these properties.

⋆E-mail: aes25@star.le.ac.uk

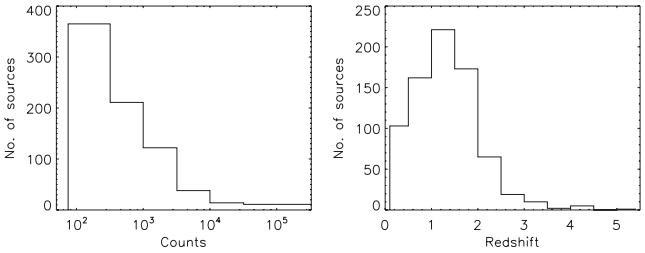

Figure 1. Histograms showing the net counts and redshift distributions of the sample used here, and described further in S11.

In a recent study of the X-ray spectral properties of 761 type 1 AGN, S11 find ∼8 per cent of their sample require a soft excess component and ∼4 per cent require intrinsic cold absorption in the modelling of their X-ray spectra. It was noted that these values represent lower limits on the intrinsic percentage of sources which include such components, since their detectability is limited by the quality of the spectra. In this paper we follow on from this analysis and aim to deduce how common these spectral features really are. In the case of the soft excess we do this by simulating typical spectra at different count levels in order to determine the maximum detection rate expected for such an additional spectral feature. This can then be compared to the observed results in order to determine the intrinsic percentage. The original data sample is described in Section 2. The soft excess simulations and a joint fitting of multiple low count spectra in an attempt to recover the soft excess are described in Section 3. Section 4 considers the presence of absorption components. We discuss our results in Section 5 and summarize our conclusions in Section 6.

2 DATA

The S11 sample was created by cross-correlating the Sloan Digital Sky Survey Data Release 5 quasar catalogue (Schneider et al. 2007) and the serendipitous X-ray source catalogue, 2XMMi (Watson et al. 2009). The X-ray spectra were extracted using standard SAS[1] tasks and fit using XSPEC v11.3.2 (Arnaud 1996) over the energy range 0.5–12.0 keV. Histograms showing the redshift and net (i.e. background subtracted) count distributions of the sources can be found in Fig. 1. All sources have >75 combined MOS+pn counts, allowing both a simple power law and an absorbed power law to be fitted, with the power-law slope, Γ, allowed to vary freely. For 680 sources with >100 counts, models including a blackbody component were also considered in order to model any soft excess. The best-fitting model was assumed to be a simple power law unless the F-test, used at the 99 per cent significance level, determined that additional components were statistically required. A summary of the different models considered can be found in Table 1 along with the number of sources best fitted with each. Soft excesses are found in ∼8 per cent of the sources in the sample and ∼4 per cent require intrinsic absorption. An additional Galactic absorption component was included in each model, fixed at the $N_{\rm H}$ value determined from the H I map of Dickey & Lockman (1990).

Fig. 2 shows the detected percentages of the soft excess (thick, blue) and absorption (red) components as a function of the number of counts in the spectra. The detected percentage of the additional components is much lower in spectra with low counts where the statistics are poorer and the features are not detected with enough significance. It was suggested in S11 that since at the highest count levels we might expect to be able to detect all soft excess components if they are present, the intrinsic percentage could be as high as the ∼80 per cent found in the highest count bin, making soft excesses common in the X-ray spectra of type 1 AGN. Fig. 2 also shows intrinsic absorption detected in up to ∼25 per cent of sources in the higher count bins. Since the sample is drawn from a population of type 1 AGN, such a component is not expected to be required in the modelling of their X-ray spectra. The F-test was used at 99 per cent significance when choosing the best-fitting model for a particular source; therefore 1 per cent of the detections of a specific spectral component can be considered spurious. This 1 per cent level is shown by the dashed line in Fig. 2.

The sample contains 62 sources which were the target of an *XMM–Newton* observation and therefore generally contain more counts in their X-ray spectra than the serendipitously detected

[1] The description and documentation are available online at http://xmm.esac.esa.int/sas/.

Table 1. The different models used to fit the X-ray spectra of the sources in the S11 sample.

Model	Spectral components	Sources
po	Power law	672
apo	PL + absorption	29
po+bb	PL + soft excess	55
apo+bb	PL + absorption + soft excess	5
		761

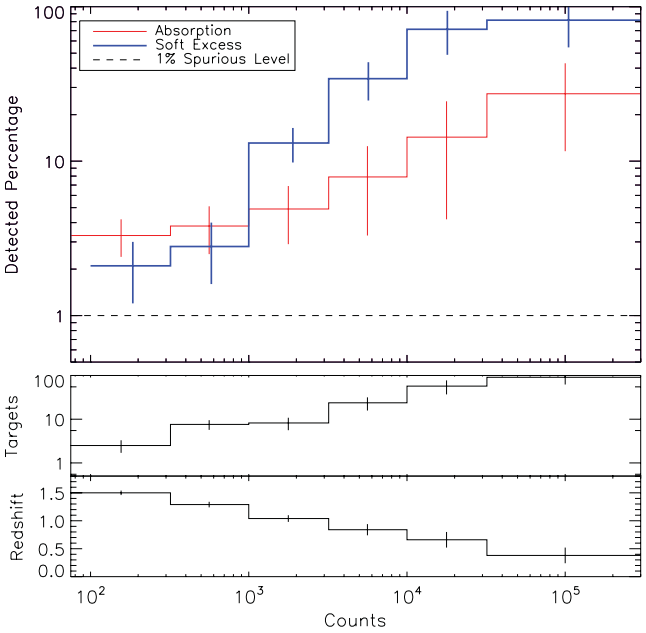

Figure 2. The top plot shows how the percentage of sources which require an absorption or soft excess component varies depending on the number of counts (MOS+pn) that are available in the spectra. The lowest count bin includes sources with between 75–320 total counts for absorption and 100–320 total counts for the soft excess, since these were the minimum numbers of counts required for fitting that particular spectral component. The sample includes 62 sources which were the target of an *XMM–Newton* observation, rather than a serendipitous detection. The middle plot shows the percentage of sources within each count bin which are target sources. The errors in the top two plots have been calculated from Poissonian statistics. The bottom plot shows how the average redshift of the sources in each count bin varies and includes standard errors on the mean.

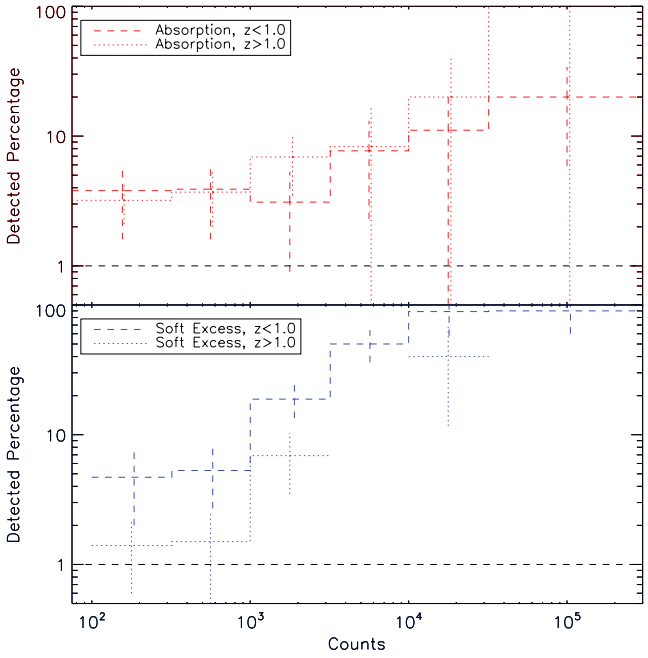

Figure 3. These figures show how the detected fraction of absorption (top) and soft excess (bottom) components varies with the numbers of counts in the spectra, considered in broad redshift bins of $z < 1$ and $z > 1$. There is a notable difference between the two curves in the bottom panel, indicating a strong redshift dependence on the detectability of soft excess components.

sources. The percentage of targets in each count bin is shown in Fig. 2 and increases towards the higher count bins as expected, with the top bin including almost only target sources. These sources could bias the detection percentages if they were selected for observation due to a previously known soft excess or intrinsic absorption. We therefore exclude the target sources from our subsequent analysis, leaving 699 sources in the sample.[2]

For sources at increasing redshifts, the contribution of a soft excess component or absorption in the spectra will gradually decrease as a larger contribution is redshifted outside of the *XMM–Newton* EPIC instrument bandpass (Strüder et al. 2001; Turner 2001). Therefore, the detected percentages of these components are expected to be higher in bins containing mostly low-redshift sources, which tend to be the bins with higher numbers of counts; hence the higher detection rate may be due to this redshift bias. The average redshift of the sources in each count bin is plotted in Fig. 2 and does decrease with increasing counts as expected. However, we note that within each count bin the sources do cover a large range in redshifts.

To further investigate the redshift issue, the sources are split into broad redshift bins of $z < 1$ and $z > 1$. Separate detection curves are created and are shown in Fig. 3. In the case of the absorbed sources (top), the detected percentages for the low- and high-redshift sources appear to be consistent within the error ranges. This is likely due to the large range of $N_{\rm H}$ values found in the sample, which means that we are able to detect absorption in sources at a range of redshifts. The percentage of sources detected with a soft excess component is higher in the low-redshift sample than in the high-redshift sample as expected. The curve for the $z > 1$ sources shows gaps where the detected fraction falls to zero, due to the low numbers of sources; although soft excesses are detected in the sample up to $z = 2$, the majority (82 per cent) are found in sources with $z \lesssim 1$, as expected.

[2] 699 sources have >75 counts and 619 have >100 counts which we fit with models including a soft excess.

3 SOFT EXCESS COMPONENTS

3.1 Intrinsic percentage

In order to determine the intrinsic percentage of sources with a soft excess component, we carry out multiple sets of simulations in order to quantify their detectability. This is done by finding the maximum percentage of components that are expected to be detected in spectra with different numbers of counts.

At each of five redshifts and between seven and 11 different count levels, we simulate 1000 spectra which include a soft excess. Each spectrum is fitted over the energy range 0.5–12.0 keV with the 'po' and 'po+bb' models, and the F-test is used at 99 per cent significance to determine whether the component is statistically required. The percentage of sources in which we significantly detect the component is then determined. By repeating this procedure with sets of spectra at different count levels and redshifts we construct synthetic detectability curves from which we can determine the maximum detection percentage at any count level.

We create our spectra, using the `fakeit` command in XSPEC which distributes a given number of counts, controlled by varying the exposure time, around a defined model with statistical fluctuations and assigns them Poissonian errors. We define the model such that it mimics the shape of a typical source in the S11 sample, both in terms of the shape of the components, i.e. the Γ and kT values and the size of the components, particularly the ratio of the blackbody normalization to that of the power-law normalization since this will also determine how easy the blackbody is to detect over the power-law continuum. 55 sources required the 'po+bb' model in the original S11 analysis. The distribution of power-law slopes and kT values for these sources were each fitted with a Gaussian, and the best-fitting mean values with dispersions were $\Gamma = 1.79 \pm 0.46$ and $kT = 0.17 \pm 0.08$ keV. We therefore simulate sources with $\Gamma = 1.8$ and $kT = 0.2$ keV and fix the normalization ratio to the median value of 0.04. These values are intended to represent a 'typical' source rather than the full range of values, although the distributions of kT and normalization ratio are reasonably narrow. As shown in fig. 22 of S11, there is a very tight correlation between the luminosities of the blackbody and power-law components. The median normalization ratio of 0.04 used here corresponds to a luminosity ratio of ∼0.2.

As has been previously discussed, the detectability of soft excess components is strongly dependent upon the redshift of the source. Therefore we run sets of simulations at a range of redshifts ($z = 0.4$, 0.75, 1.0, 1.3 and 1.5) and create five separate detectability curves shown by the different lines in Fig. 4. For interpolation purposes they are fitted with a quadratic function which describes four of the curves well. The curve for $z = 1.3$ has a slightly different curvature from the asymptotic behaviour expected and therefore in this case we use a linear function to fit the simulated points. The curves indicate the detection percentage that we would expect to see for a particular count level if all sources contained this component. If the observed percentages were consistent with this line, it would indicate that the spectral features are present in all of the sources in our sample and it is merely the quality of the spectra that limits our ability to detect them. For sources at $z \leq 1$ with $\gtrsim$10 000 counts, the detectability curves lie at 100 per cent. This suggests that the intrinsic percentage could be as high as ∼70 per cent, the percentage found in the highest count bin (after removal of the target sources). We can determine the true soft excess detection

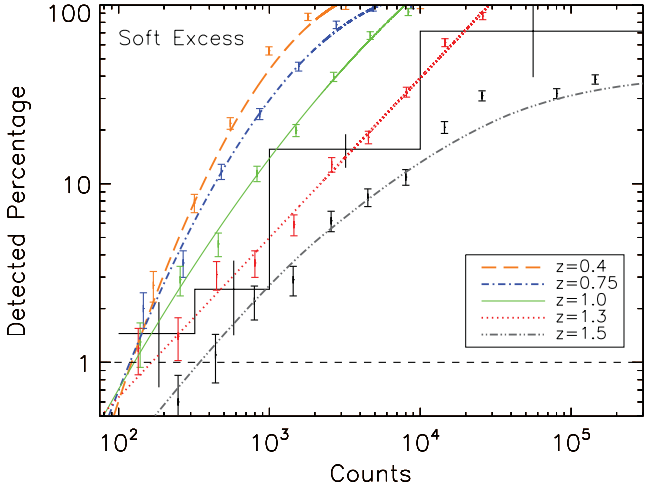

Figure 4. The solid black line with error bars shows the observed percentage of soft excess components as a function of counts for the S11 sample after the target sources have been removed, which required a rebinning of the sources. The coloured curves show the detectability level for a soft excess component with a typical shape and size. Sets of simulations are carried out at different redshifts and count levels as shown by the coloured error bars. The curves are created from a quadratic fit to at least seven such results (a linear fit is used for $z = 1.3$).

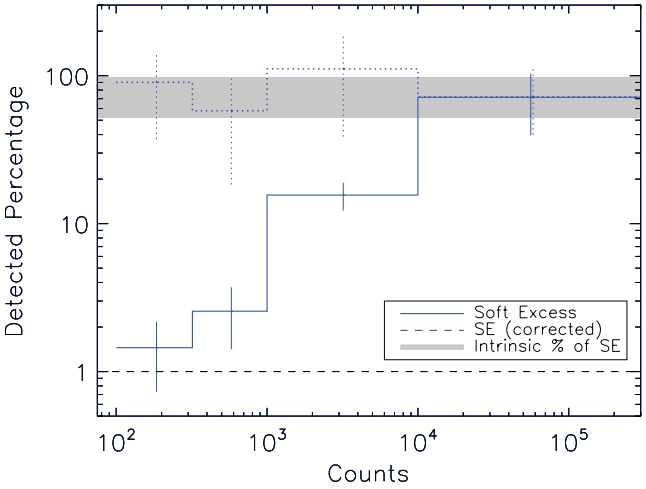

Figure 5. The blue solid line shows the observed percentage of soft excess components in the S11 sample (after the target sources have been removed). The dashed line shows the detected percentage determined after the correction for spectral quality. The grey bar indicates the location of the intrinsic percentage of sources with soft excesses, where the width of the bar represents the 1σ error boundary.

percentage (corrected percentage) for each count bin in our data by dividing the observed detection percentage by the maximum percentage obtained from our simulated curves which is fixed to lie between the spurious level, 1 per cent, and 100 per cent.

As each bin includes sources at a range of redshifts, we calculate a corrected percentage using all five of the detectability curves and determine a weighted average value according to the equation

$$\text{Result} = \sum (\text{weight} \times \text{corrected percentage}), \quad (1)$$

where the 'weight' is the fraction of sources for which each particular redshift curve is appropriate for within that count bin. For sources at $z > 2$, the maximum percentage is fixed at 1 per cent, since no real soft excesses are expected, but spurious detections may occur. Since the count bins are broad and the sources are not distributed evenly within them, they are further divided into sub-bins. The redshift-corrected percentage is calculated for each as outlined above, and an overall corrected percentage is reconstructed for the full count bin using equation (1), where 'weight' in this case is the fraction of sources within the particular count sub-bin.

The dashed blue line in Fig. 5 shows the corrected percentages.[3] The values are roughly constant after the effect of spectral quality has been removed. In order to determine a value for the intrinsic percentage, χ^2 is calculated for constant percentages between 0 and 100 per cent. The minimum χ^2 occurs for a constant percentage of 75 ± 23 per cent, for which the fit has a null hypothesis probability of $p = 91$ per cent. This method produces a corrected percentage greater than 100 per cent in the third bin. Whilst this has no physical meaning, its large error bar makes it entirely consistent with the corrected percentages obtained for the other bins. Capping this bin at 100 per cent reduces our intrinsic percentage estimate by only 1 per cent.

To test whether a 1 per cent spurious level is appropriate we simulated spectra which did not include a soft excess, and determined in how many we incorrectly detected this component with >99 per cent significance. For the majority of count levels this percentage was consistent with or lower than 1 per cent. However, for the lowest and highest count levels ($\lesssim$100 and ~50 000), the spurious level was somewhat higher (up to ~5 per cent), suggesting that both the usual statistical problems affecting simulated fits and some small systematic calibration errors exist. The count bin most affected by the spurious level is the lowest, as the simulated curve for $z = 1.5$ sources falls below 1 per cent. If the maximum percentage from the simulated curve is fixed at 1 per cent, the spurious level, the correction factor is underestimated. Excluding this bin from the intrinsic percentage determination gives 71^{+26}_{-24} per cent, which is consistent with the value obtained when the bin is included.

3.2 Joint spectral fitting

We have shown that after correcting for spectral quality, a soft excess component may be ubiquitous in the S11 sample. We now re-analyse spectra with low numbers of counts in a joint fitting to see if the soft excess feature can be recovered in a combination of spectra where it was previously undetected.

The S11 sample includes 436 sources which were originally fitted with a simple power-law model, i.e. the soft excess feature was not detected in their spectra with >99 per cent significance, and which have <500 counts. This limit is imposed so that the joint fitting is not dominated by a single object with high numbers of counts. Samples of ~30 sources (~50 MOS and/or pn spectra) are created which cover a narrow range in redshift. The resulting groups of spectra include a total of ~7000 counts, a level at which a soft excess in a $z < 1$ source is expected to be detected 100 per cent of the time if it is present. This is shown in Fig. 4 where the detectability curves for $z = 0.4$ and 0.75 sources are at 100 per cent at a 7000 count level.

The groups of spectra are jointly fitted with the simple power-law model, 'po', in which Γ is free to vary. In each case we find a

[3] We note that percentages >100 per cent are a possible consequence of using this correction factor method. The count bins have been redefined in order to reduce this statistical effect. However, should the intrinsic percentage be exactly 100 per cent, we would expect some bins to show correction percentages lower than 100 per cent, and others greater than 100 per cent.

Table 2. The samples used for the joint fitting and the results from fitting the spectra with the simple power-law model 'po' in which Γ is allowed to vary freely and the 'po+bb' model in which Γ is both fixed at 1.8 and free to vary. 'po+bb fix A' refers to the model in which the blackbody-to-power-law normalization ratio is kept the same for each spectrum and 'po+bb fix B' is the model in which both the blackbody and power-law normalizations are free to vary. In the case of model A, the normalization ratio quoted is the best-fitting value with a 68 per cent error and for model B it is the median value. The errors on the kT and Γ parameters are 90 per cent.

z range	No. of sources (spectra)[a]	Total counts	Model	χ^2/ν	Γ	kT (keV)	Normalization ratio	Percentage with soft excess
0.21–0.55	34 (51)	8078	po	859/636 (1.351)	$2.10^{+0.04}_{-0.05}$			
			po+bb	664/584 (1.137)	$1.58^{+0.08}_{-0.06}$	$0.23^{+0.01}_{-0.03}$	0.07	78 ± 17
			po+bb fix A	873/635 (1.375)	1.8 (fixed)	$0.23^{+0.03}_{-0.03}$	0.024 ± 0.002	
			po+bb fix B	675/585 (1.154)	1.8 (fixed)	$0.19^{+0.02}_{-0.02}$	0.04	74 ± 16
0.56–0.77	30 (50)	7661	po	861/615 (1.400)	$2.07^{+0.05}_{-0.05}$			
			po+bb	682/564 (1.209)	$1.62^{+0.08}_{-0.13}$	$0.24^{+0.03}_{-0.02}$	0.07	84 ± 18
			po+bb fix A	854/614 (1.391)	1.8 (fixed)	$0.19^{+0.04}_{-0.04}$	0.040 ± 0.010	
			po+bb fix B	693/565 (1.227)	1.8 (fixed)	$0.20^{+0.03}_{-0.02}$	0.05	72 ± 16
0.77–0.98	33 (51)	7367	po	683/611 (1.118)	$2.03^{+0.05}_{-0.04}$			
			po+bb	624/559 (1.116)	$1.98^{+0.04}_{-0.05}$	$0.044^{+0.002}_{-0.001}$		
			po+bb fix A	681/610 (1.116)	1.8 (fixed)	$0.27^{+0.05}_{-0.05}$	0.036 ± 0.005	
			po+bb fix B	579/560 (1.034)	1.8 (fixed)	$0.28^{+0.03}_{-0.04}$	0.06	63 ± 14
0.98–1.12	31 (51)	7583	po	632/629 (1.005)	$2.02^{+0.05}_{-0.04}$			
			po+bb	576/577 (0.998)	$1.97^{+0.04}_{-0.06}$	$0.053^{+0.003}_{-0.002}$		
			po+bb fix A	616/628 (0.981)	1.8 (fixed)	$0.26^{+0.04}_{-0.04}$	0.048 ± 0.008	
			po+bb fix B	629/578 (1.088)	1.8 (fixed)	0.02[b]		
1.12–1.19	30 (50)	6749	po	604/578 (1.045)	$2.03^{+0.04}_{-0.05}$			
			po+bb	571/527 (1.083)	$1.98^{+0.04}_{-0.06}$	$0.06^{+0.01}_{-0.03}$		
			po+bb fix A	597/577 (1.035)	1.8 (fixed)	$0.35^{+0.05}_{-0.05}$	0.045 ± 0.005	
			po+bb fix B	606/528 (1.148)	1.8 (fixed)	$0.065^{+0.005}_{-0.005}$		
1.19–1.28	29 (50)	5614	po	485/509 (0.953)	$1.95^{+0.05}_{-0.05}$			
			po+bb	450/458 (0.983)	$1.89^{+0.05}_{-0.06}$	$0.062^{+0.002}_{-0.002}$		
			po+bb fix A	471/508 (0.927)	1.8 (fixed)	$0.24^{+0.06}_{-0.06}$	0.050 ± 0.030	
			po+bb fix B	416/459 (0.906)	1.8 (fixed)	$0.34^{+0.07}_{-0.06}$	0.04	66 ± 15

[a]Not all sources have both a MOS spectrum and a pn spectrum available.
[b]For this sample we are unable to estimate reliable errors from the fit.

best-fitting result of Γ ∼ 2. In the case of the lowest two redshift bins, the fit is significantly improved by using the 'po+bb' model, both in terms of a lower χ^2/ν, as listed in Table 2, and an F-test comparing the two models which equals 100 per cent in both cases. However the Γ values are lower than 1.8, the value used in the simulations of the previous section. At lower redshifts, we expect to be more sensitive to cold absorption (see Section 4) and we note that if a fixed absorption component of $N_{\rm H} = 10^{21}{\rm cm}^{-2}$ is included in the fit, steeper values of Γ ∼ 1.70 ± 0.05 are obtained. This does not fully correct for the effect as the resulting continuum from stacking spectra with different levels of absorption is known to be flatter (Mateos et al. 2010).

We also fit each group of spectra with the 'po+bb' model in which Γ is fixed at 1.8 and kT is tied to the same value for each of the spectra. We determine a single, best-fitting normalization ratio common to all the spectra by fixing the power-law normalizations at 10^{-5} and tying the blackbody normalizations for each spectrum to a common value. A freely varying constant is added to the model to allow each data set to vary independently. The results can be found in Table 2 listed as the model 'po+bb fix A'. For all except the first bin we find average normalization ratios which are consistent with the value of 0.04 used in the simulations of the previous section. For all except the fifth bin, the kT values are consistent with those found in individual sources with high count spectra.

Each group of spectra is also fitted with the 'po+bb' model in which Γ is fixed at 1.8, kT is tied to a common value for each of the spectra and the normalizations of both the blackbody and power-law components are left free to vary. These results are listed in Table 2 as the model 'po+bb fix B', and include the median value of the blackbody-to-power-law normalization ratios and the percentage of spectra for which the ratio is >0.01, which is interpreted as a blackbody component being present. For the three lower redshift bins ($z \lesssim 1$), the inclusion of a blackbody component in the spectral fit results in a better fit and the blackbody temperatures are consistent with the previous values obtained from the individual high count spectra. The normalization ratios of the blackbody to power-law components are also consistent, both in terms of the median value and a Kolmogorov–Smirnov (KS) test which finds the full distribution of ratios to be not significantly different from those found in single object spectra in S11. The median normalization ratio of ∼0.04 is consistent with the value used in the simulations of the previous section and corresponds to a luminosity ratio of ∼0.2. In addition, the percentage of spectra which do include a blackbody component (since the fit allows a blackbody normalization of zero

Table 3. The samples created based on Eddington ratio values and the results of the joint spectral fitting.

$\lambda_{\rm Edd}$ range[a]	No. of sources (spectra)	Total counts	Model	χ^2/ν	Γ	kT (keV)	Normalization ratio	Percentage with soft excess
-2.3 to -1.3	28 (47)	5613	po+bb fix A	685/498 (1.376)	1.8 (fixed)	$0.15^{+0.16}_{-0.10}$	0.008 ± 0.7	
			po+bb fix B	570/452 (1.261)	1.8 (fixed)	$0.18^{+0.04}_{-0.02}$	0.005	45 ± 12
-1.3 to -1.0	32 (52)	8383	po+bb fix A	821/660 (1.244)	1.8 (fixed)	$0.25^{+0.03}_{-0.02}$	0.048 ± 0.004	
			po+bb fix B	657/609 (1.079)	1.8 (fixed)	$0.23^{+0.02}_{-0.02}$	0.04	75 ± 16
-1.0 to $+0.2$	26 (36)	6536	po+bb fix A	546/500 (1.092)	1.8 (fixed)	$0.18^{+0.04}_{-0.03}$	0.05 ± 0.01	
			po+bb fix B	472/465 (1.015)	1.8 (fixed)	$0.16^{+0.04}_{-0.03}$	0.084	83 ± 21

[a] $\lambda_{\rm Edd} = \log(L_{\rm bol}/L_{\rm Edd})$.

for individual spectra) is consistent with the intrinsic percentage calculated in Section 3.1. This model also gives a better fit for the highest of the redshift bins. However, the kT value is too high to be consistent with the temperatures observed in individual spectra.

It has been suggested that the soft excess feature is ubiquitous in high accretion rate AGN and that it is this parameter which may determine its presence in the spectra, or the size of the component (e.g. Done et al. 2012). We use the Eddington ratio as a proxy for mass accretion rate defined as $\lambda_{\rm Edd} = \log(L_{\rm bol}/L_{\rm Edd})$. The bolometric luminosity was estimated from the observed X-ray luminosity in the 2–10 keV band, measured from the best-fitting spectral model, by applying the luminosity-dependent correction of Marconi et al. (2004). The Eddington luminosity was estimated using virially determined black hole mass estimates from the Shen et al. (2008) catalogue. As before, we use only sources with <500 total counts and those best fit with the model 'po'. In addition, we also further restrict the sources to those at $z < 1$, and we split the remaining sources into three subsamples containing roughly equal numbers of sources. The samples are fitted with both the 'po+bb fix A' and 'po+bb fix B' models, the results of which are listed in Table 3 along with the properties of each sample. The 'po+bb fix A' model finds that a best-fitting normalization ratio is consistent with the 0.04 found in single spectra for the top two bins. In the case of the lowest $\lambda_{\rm Edd}$ bin, this value is considerably lower, although it has a large associated error. A KS test finds the distributions of the normalization ratios determined from the 'po+bb fix B' model to be consistent; however, the median values do vary from a lower value than expected in the low $\lambda_{\rm Edd}$ bin to a higher value than expected in the high $\lambda_{\rm Edd}$ bin. In addition, the percentage of spectra for which the normalization ratio is >0.01 is consistent with the intrinsic percentage in the case of the top two bins, but is lower, although still within errors, for the lowest $\lambda_{\rm Edd}$ bin. Whilst the evidence is not strong, this may suggest that soft excesses are smaller and less common in sources with lower accretion rates and we do not rule this out.

4 ABSORPTION COMPONENTS

Intrinsic cold absorption may also be present in type 1 AGN, suppressing the lower energy emission. The percentage of sources with detected absorption is shown in Fig. 2 and appears to be limited by the spectral quality in a similar way to that of the soft excess. However, the effect is not as strong, with the detected percentage decreasing by approximately 20 per cent from the higher to lower count bins rather than ∼80 per cent in the case of the soft excess. It was suggested in S11 that the true percentage of absorbed sources could be as high as the ∼25 per cent found in the highest count range, where we might expect the spectra to be of good enough quality to detect any significant absorption if present. However, the highest count bins are heavily contaminated by target sources, resulting in a lower detected percentage of absorption components once they are removed from consideration. Fig. 6 shows the detected percentage after target removal. The highest count bin in this plot now suggests that 5.6 ± 3.9 per cent of type 1 AGN may include an intrinsic absorption component. In the lower count bins (<3200), the detected percentage of absorbed sources does not vary significantly, suggesting that the detectability is not as heavily dependent on spectral quality as it is for the soft excess feature.

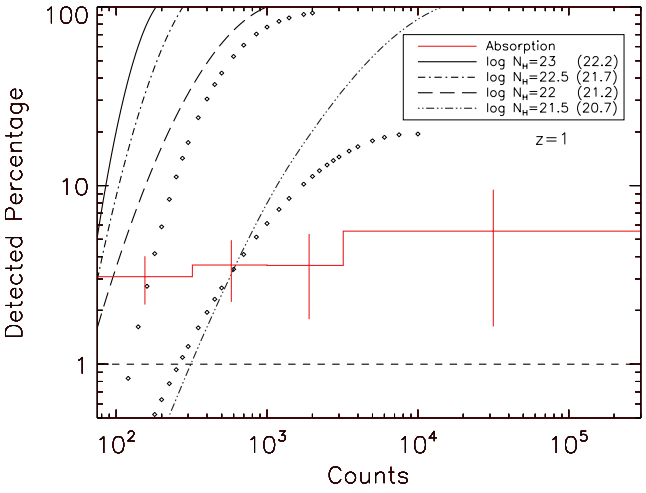

Figure 6. This figure shows how the detected percentage of an intrinsic cold absorption component varies with the number of counts in the X-ray spectra (red solid line). It reproduces Fig. 2, but the sources which were the target of an *XMM–Newton* observation have been removed. As a result of the reduced number of sources this leaves, the original top three bins are combined. The percentage appears to remain constant at ∼3 per cent for spectra with 75–32 000 counts, suggesting the detectability is not as heavily dependent on the spectral quality as is the case for the soft excess. Detectability curves for absorption components with different column densities are shown with different line styles in simulated spectra at $z = 1$. The equivalent $N_{\rm H}$ value at $z = 0$ is quoted in brackets on the figure. The open squares show the detectability of absorption of $\log N_{\rm H} = 22$ and $\log N_{\rm H} = 21.5$, when a soft excess of typical shape and size is also included in the spectra.

The detectability of absorption in the spectra is highly dependent on the $N_{\rm H}$ value, and the range of rest-frame column densities found in the absorbed sources is very broad, 10^{21}–10^{23} cm^{-2}. Therefore, we cannot choose a single model to simulate which is representative of all the absorbed sources we detect, unlike in the case of the soft excess, and we do not have enough statistics in order to weight detectability curves by both $N_{\rm H}$ and z. We do attempt to quantify

the detectability of absorption components with different column densities by simulating absorbed spectra at $z = 1$. These detectability curves are shown in Fig. 6 by the curves of different line style. They show that a column of 10^{23} cm^{-2} (equivalent to 2×10^{22} cm^{-2} at $z = 0$) would be detected in most spectra with >200 counts, whereas a column of $\sim 3 \times 10^{21}$ cm^{-2} (equivalent to 5×10^{20} cm^{-2} at $z = 0$) is not strong enough to be detected in spectra of this quality at $z = 1$. For the highest count bin ($\gtrsim$3200 counts), the detectability curves for all but the lowest level of $N_{\rm H}$ shown in Fig. 6 are at 100 per cent. This means that we expect to be sensitive to all reasonable levels of $N_{\rm H}$ and therefore our intrinsic percentage estimate is robust. The fraction of sources with particular $N_{\rm H}$ levels is roughly constant in both different z and count bins (the two properties being correlated), and hence the slight decrease in the detected percentage between the top bin where we expect to detect all levels of $N_{\rm H}$ and the bottom bins is what is expected when objects with lower $N_{\rm H}$ and/or higher redshifts are no longer detectable.

These simulations do not include a soft excess component which could also reduce the detectability of any absorption present. We investigate this by simulating sources with both the standard soft excess parameters and two values of $N_{\rm H}$ (shown by the lines of open squares in Fig. 6). For $\log N_{\rm H} = 22$ our sensitivity drops by approximately 10 per cent at low count levels (∼200), increasing to ∼25 per cent at higher count levels (∼1000), making little difference to our conclusions. In the case of $\log N_{\rm H} = 21.5$, at low count levels ($\lesssim$1000) we are mostly insensitive to the absorption anyway, such that the inclusion of a soft excess makes little difference to its detectability. At higher count levels (∼10 000) where the statistics are better, including the soft excess can reduce our sensitivity to the absorption component by ∼65 per cent.

In Section 3.2, a joint fitting was carried out on groups of low count spectra to see if the soft excess feature could be recovered. Similarly, we fit the 'apo+bb' model to the same samples to see if an absorption component can be recovered in addition to the soft excess already found to be present. We implement the model with Γ fixed to 1.8 and kT fixed at 0.2 keV and leave both the power-law and blackbody normalizations free to vary. The best-fitting $N_{\rm H}$ values are listed in Table 4. We find that an absorption component can be recovered in the low count spectra, but only in the lower redshift bins is this component constrained. Although the $N_{\rm H}$ values are of the order of those seen in single object fits in the S11 sample, the range in column densities means that the values we obtain here merely represent an 'average' $N_{\rm H}$ value, the exact value of which should be treated with caution.

Table 4. Results from fitting the 'apo+bb' model to the groups of low count spectra described in Table 2. In each case, Γ and kT are fixed at 1.8 and 0.2 keV, respectively. 90 per cent errors are quoted on the $N_{\rm H}$ value.

z range	$N_{\rm H}$ ($\times 10^{22}$ cm^{-2})
0.21–0.55	$0.57^{+0.27}_{-0.15}$
0.56–0.77	$0.83^{+0.22}_{-0.19}$
0.77–0.98	<0.43
0.98–1.12	<0.17
1.12–1.19	<0.53
1.19–1.28	<0.61

5 DISCUSSION

The origin of the soft excess emission is still a matter of debate. Since this feature has been shown to be ubiquitous in the X-ray spectra of type 1 AGN, any models to explain its origin must be applicable to all sources. Some theories of the soft excess describe it as an 'apparent' feature, rather than an additional component in the spectrum. A strong jump in opacity at ∼0.7 keV is created by lines and edges of ionized O VII and O VIII which are smeared by the high velocities or gravitational redshifts found close to a black hole. This can appear from absorption through optically thin material in the line of sight (Gierliński & Done 2004) or via reflection from optically thick material out of the line of sight (Ross & Fabian 2005; Crummy et al. 2006). More recently, Done et al. (2012) have suggested that the soft excess may be intrinsic emission from the disc, which is shifted into soft X-ray energies due to the required colour-temperature correction, and further Compton upscattering produces the full components observed. However, this only applies to the lowest mass/highest accretion rate AGN and since we find the soft excess to be present not only in high accretion rate sources, this suggests two separate interpretations for the soft excess are required. An alternative theory is that part of the soft X-ray emission may be due to cooling outflows which are now thought to be reasonably common in AGN (Tombesi et al. 2010). During Eddington accretion episodes, high-velocity ($v \sim 0.1c$) and highly ionized ($\xi \sim 10^4$) winds are produced and when these interact with the interstellar medium of the host galaxy, the gas is strongly shocked (King 2010). Subsequent Compton cooling of this gas may be observable as a soft X-ray component, as suggested in the case of NGC 4051 (Pounds & Vaughan 2011).

The ubiquity of the soft excess means that any X-ray spectral fitting of type 1 AGN must take this feature into account. It has been shown that leaving this component unmodelled can lead to a Γ value ∼0.1 too steep. In addition, any attempt to constrain an intrinsic $N_{\rm H}$ value must also include the blackbody component in the fit since they appear in the spectra at a similar energy range. It was found in S11 that the average Γ values for sources fitted with the 'po+bb' model were significantly flatter than those fitted with the 'po' model ($\Gamma_{\rm po} = 1.98 \pm 0.01$, $\Gamma_{\rm po+bb} = 1.87 \pm 0.05$ and KS significance = 0.0003), suggesting that the underlying power-law slope in sources with a soft excess is different. However, our modelling uses a blackbody at low energies to model the soft excess component and this is purely phenomenological – it provides a good representation of the feature seen in spectra of our quality. If the soft excess is actually a broad spectral feature as suggested by reflection models, our modelling may not be fully accounting for the spectral complexity and the power-law slope at higher energies could still contain some of this component.

This work and many in the literature have confirmed the presence of a population of AGN which are classified as type 1 due to the presence of broad emission lines in their UV/optical spectra, but also show significant X-ray absorption. Conversely, there are also objects optically classified as type 2 which are unabsorbed in X-rays (e.g. Panessa & Bassani 2002; Mateos et al. 2005a,b). These objects are not reconciled by the standard orientation-based unified model which invokes an obscuring torus to both block the line of sight to the broad-line region and give X-ray absorption in the case of type 2 AGN, but not type 1. Constraining the fraction of absorbed type 1 sources may aid in interpreting these objects in terms of a correction to the unified model. Such a correction may include invoking the 'clumpy torus' model (Nenkova et al. 2008a,b) in which the torus consists of individual clouds. Observations of AGN which showed

large amplitude and rapid variability of the X-ray column density were interpreted as occultations by clouds in the broad-line region, suggesting clumpy absorbers may be present at a range of different scales. In this model, observing a given source as absorbed depends upon the covering factor of the clouds and is a probability issue rather than one of just orientation (Risaliti, Elvis & Nicastro 2002). It has also been suggested that obscuration of AGN could occur due to the presence of a warped accretion disc (Greenhill et al. 2003; Nayakshin 2005; Lawrence & Elvis 2010). In this scenario there is no need for an obscuring torus; the type 2 objects tend to be the ones with larger misalignments, giving larger covering factors.

Whilst it would be interesting to compare how the detection of the low-energy spectral components varies with radio loudness,[4] we lack the statistics to do this. However, we do note that 6/75 radio-loud quasars (RLQs) have a detected soft excess and a joint fit of 29 sources with a total of ∼7000 counts gives similar soft excess parameters to those found for the radio-quiet quasars (RQQs). The prevalence and magnitude of soft excesses thus appear very similar for RQQs and RLQs, contrary to previous suggestions that RLQs do not include this component (e.g. Sulentic, Zamfir & Marziani 2010).

6 CONCLUSIONS

In this paper we have simulated X-ray spectra consisting of a power law and a soft excess, in which the shape of the latter component represents a typical example of those found in the S11 sample. The spectra were fitted in XSPEC over the energy range 0.5–12.0 keV and an F-test was used at 99 per cent significance to test whether the component was statistically required in the fit. By repeating this procedure for spectra with different numbers of counts, maximum detection curves were generated and compared to the observed results from the real data sample presented in S11. The effect of redshift on the detectability was also taken into account. Despite the raw percentage of sources with a soft excess being ∼8 per cent, we showed that after correcting for the spectral quality, the intrinsic percentage is 75 ± 23 per cent. This suggests that within the S11 sample, almost all of the sources could include a soft excess component with a shape and size typical of that seen in the highest count spectra, and it is merely the quality of the spectra that is limiting our ability to detect them. The detectability of a soft excess component is dependent on both the blackbody temperature and the size of its normalization with respect to the underlying power law used in the modelling, but we find that using slightly different values for either does not change our overall conclusion.

If soft excesses are ubiquitous, then the feature should be recovered in a combination of low count spectra. Groups of ∼50 spectra (∼7000 counts) were created in narrow z bins, including spectra with <500 counts and which had no previous evidence for additional spectral features. The groups at $z < 1$ were shown to be better fitted with a model including a soft excess, and the temperature and normalization with respect to the underlying power law of the components required were consistent with those found in individual high count spectra.

We are unable to conduct a simulation procedure in order to determine the percentage of type 1 AGN which require an intrinsic cold absorption component. However, we suggest that its detectability may not be as dependent upon spectral quality as the soft excess. We stress that a non-negligible percentage, ∼5 per cent, of type 1 AGN may include such an absorption feature and therefore any spectral modelling must take the possibility of this feature into account.

[4] The determination of the radio loudness of the sources is described in S11.

ACKNOWLEDGMENTS

We thank the referee for comments which helped improve this paper. AES acknowledges support from an STFC studentship and SM acknowledges financial support from the Ministerio de Ciencia e Innovación under projects AYA2009-08059 and AYA2010-21490-C02-01. This work was based on observations obtained with *XMM–Newton*, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA.

REFERENCES

Antonucci R., 1993, ARA&A, 31, 473
Arnaud K. A., 1996, in Jacoby G. H., Barnes J., eds, ASP Conf. Ser. Vol. 101, Astronomical Data Analysis Software and Systems V. Astron. Soc. Pac., San Francisco, p. 17
Arnaud K. A. et al., 1985, MNRAS, 217, 105
Bianchi S., Guainazzi M., Matt G., Fonseca Bonilla N., Ponti G., 2009, A&A, 495, 421
Corral A., Della Ceca R., Caccianiga A., Severgnini P., Brunner H., Carrera F. J., Page M. J., Schwope A. D., 2011, A&A, 530, A42
Crummy J., Fabian A. C., Gallo L., Ross R. R., 2006, MNRAS, 365, 1067
Dickey J. M., Lockman F. J., 1990, ARA&A, 28, 215
Done C., Davis S. W., Jin C., Blaes O., Ward M., 2012, MNRAS, 420, 1848
Garcet O. et al., 2007, A&A, 474, 473
Gierliński M., Done C., 2004, MNRAS, 349, L7
Greenhill L. J. et al., 2003, ApJ, 590, 162
Haardt F., Maraschi L., 1993, ApJ, 413, 507
King A. R., 2010, MNRAS, 402, 1516
Lawrence A., Elvis M., 2010, ApJ, 714, 561
Mainieri V. et al., 2007, ApJS, 172, 368
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Mateos S. et al., 2005a, A&A, 433, 855
Mateos S., Barcons X., Carrera F. J., Ceballos M. T., Hasinger G., Lehmann I., Fabian A. C., Streblyanska A., 2005b, A&A, 444, 79
Mateos S. et al., 2010, A&A, 510, A35
Nayakshin S., 2005, MNRAS, 359, 545
Nenkova M., Sirocky M. M., Ivezić Ž., Elitzur M., 2008a, ApJ, 685, 147
Nenkova M., Sirocky M. M., Nikutta R., Ivezić Ž., Elitzur M., 2008b, ApJ, 685, 160
Page M. J. et al., 2003, Astron. Nachr., 324, 101
Panessa F., Bassani L., 2002, A&A, 394, 435
Perola G. C. et al., 2004, A&A, 421, 491
Piconcelli E., Jimenez-Bailón E., Guainazzi M., Schartel N., Rodríguez-Pascual P. M., Santos-Lleó M., 2005, A&A, 432, 15
Porquet D., Reeves J. N., O'Brien P., Brinkmann W., 2004, A&A, 422, 85
Pounds K. A., Vaughan S., 2011, MNRAS, 413, 1251
Reeves J. N., Turner M. J. L., 2000, MNRAS, 316, 234
Risaliti G., Elvis M., Nicastro F., 2002, ApJ, 571, 234
Ross R. R., Fabian A. C., 2005, MNRAS, 358, 211
Schneider D. P., Hall P. B., Richards G. T., Strauss M. A., Vanden Berk D. E., 2007, AJ, 134, 102
Scott A. E., Stewart G. C., Mateos S., Alexander D. M., Hutton S., Ward M. J., 2011, MNRAS, 417, 992, (S11)
Shen Y., Greene J. E., Strauss M. A., Richards G. T., Schneider D. P., 2008, ApJ, 680, 169
Strüder L., Briel U., Dennerl K., Hartmann R., Kendziorra E., Meidinger N., Pfeffermann E., Reppin C., 2001, A&A, 365, L18

Sulentic J., Zamfir S., Marziani P., 2010, in Maraschi L., Ghisellini G., Della Ceca R., Tavecchio F., eds, ASP Conf. Ser. Vol. 427, Accretion and Ejection in AGN: A Global View. Astron. Soc. Pac., San Francisco, p. 371
Tombesi F., Cappi M., Reeves J. N., Palumbo G. G. C., Yaqoob T., Braito V., Dadina M., 2010, A&A, 521, A57
Turner M. J. L., 2001, A&A, 365, L27
Turner T. J., Pounds K. A., 1989, MNRAS, 240, 833
Watson M. G. et al., 2009, A&A, 493, 339
Winter L. M., Veilleux S., McKernan B., Kallman T. R., 2012, ApJ, 745, 107
Young M., Elvis M., Risaliti G., 2009, ApJS, 183, 17

This paper has been typeset from a TEX/LATEX file prepared by the author.