Probing large-scale structure with large samples of X-ray selected AGN: I Baryonic acoustic oscillations

Gert Hütsi[1,2], Marat Gilfanov[1,3], Alexander Kolodzig[1], Rashid Sunyaev[1,3]

[1] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
    e-mail: `gert@mpa-garching.mpg.de`
[2] Tartu Observatory, Tõravere 61602, Estonia
[3] Space Research Institute of Russian Academy of Sciences, Profsoyuznaya 84/32, 117997 Moscow, Russia

ABSTRACT

We investigate the potential of large X-ray selected AGN samples for detecting baryonic acoustic oscillations (BAO). Though AGN selection in X-ray band is very clean and efficient, it does not provide us redshift information, and thus needs to be complemented with an optical follow-up. The main focus of this study is: (i) to find necessary requirements to the quality of the optical follow-up and (ii) to formulate the optimal strategy of the X-ray survey, in order to detect the BAO. We demonstrate that redshift accuracy of $\sigma_0 = 10^{-2}$ at $z = 1$ and the catastrophic failure rate of $f_{\rm fail} \lesssim 30\%$ are sufficient for a reliable detection of BAO in future X-ray surveys. Spectroscopic quality redshifts ($\sigma_0 = 10^{-3}$ and $f_{\rm fail} \sim 0$) will boost the confidence level of the BAO detection by a factor of ~ 2. For the meaningful detection of BAO, X-ray surveys of moderate depth of $F_{\rm lim} \sim$ few 10^{-15} erg/s/cm^2 covering sky area from a ∼few hundred to ∼ten thousand square degrees are required. The optimal strategy for the BAO detection does not necessarily require full sky coverage. For example, in a 1000 days long survey by an eROSITA type telescope, an optimal strategy would be to survey a sky area of ~ 9000 deg^2, yielding a $\sim 16\sigma$ BAO detection. A similar detection will be achieved by ATHENA+ or WFXT class telescopes in a survey with a duration of 100 days, covering similar sky area. XMM-Newton can achieve a marginal BAO detection in a 100 days survey covering ~ 400 deg^2. These surveys would demand a moderate to high cost in terms the optical follow-ups, requiring determination of redshifts of $\sim 10^5$ (XMM-Newton) to $\sim 3 \times 10^6$ objects (eROSITA, ATHENA+ and WFXT) in the above mentioned sky areas.

Key words. Galaxies: active – X-rays: galaxies – Cosmology: theory – large-scale structure of Universe

1. Introduction

Mapping the large-scale structure (LSS) in the low redshift Universe with modern redshift surveys, together with cosmic microwave background (CMB) (de Bernardis et al. 2000; Hanany et al. 2000; Hinshaw et al. 2013; Planck Collaboration et al. 2013a) and type Ia supernova (Riess et al. 1998; Perlmutter et al. 1999) measurements, has helped to establish the current standard cosmological model – the ΛCDM model. The initial discovery of the baryonic acoustic oscillations (BAO) (Sunyaev & Zeldovich 1970; Peebles & Yu 1970) in the two-point clustering statistics of the blue-band 2dFGRS[1] and r-band selected SDSS[2] luminous red galaxy (LRG) samples (redshifts $z \sim 0.2$ and $z \sim 0.35$, respectively) (Cole et al. 2005; Eisenstein et al. 2005) has initiated significant efforts to try to measure this signal, which provides us a theoretically well-understood standard ruler, at ever larger redshifts, this way mapping out the low redshift expansion history of the Universe in great detail.[3] Detecting BAO demands major observational efforts as this relatively weak signal, $\sim 5\%$ modulation in the matter power spectrum, can only be detected by combining large survey volumes with high enough sampling density, to overcome cosmic variance and shot noise, respec-

tively. This leads to a typical number of redshifts obtained in these surveys ranging from $\sim 10^5$ up to $\sim 10^6$. The measurements of BAO in the galaxy two-point function have been extended to redshifts of $z \sim 0.55$ and $z \sim 0.7$ by BOSS[4] and WiggleZ[5] surveys, using LRGs and emission line galaxies (ELGs), respectively (Anderson et al. 2012; Blake et al. 2011). By exploiting the correlations in quasar Lyman-α forest fluctuations the BAO detection has been pushed up to $z \sim 2.3$ (Busca et al. 2013; Slosar et al. 2013).

Even though at the moment there is no BAO detections in the redshift range $z \sim 1 - 2$, it will be covered by the upcoming eBOSS[6] survey. The measurement at even higher redshifts, $z \sim 3$, will be achieved by HETDEX[7] survey.

The eBOSS will use ELGs up to redshifts $z \sim 1$ and beyond that $\sim 750,000$ quasars (QSOs) will be used to sample cosmic density field over ~ 7500 deg^2, i.e., with sampling rate ~ 100 deg^{-2}. This is a factor of ~ 2 higher than a predicted number density of the X-ray active galactic nuclei (AGN) with $z > 1$ from the upcoming eROSITA[8] [9]

---

[1] `www2.aao.gov.au/2dfgrs/`
[2] `www.sdss.org/`
[3] For a recent review see Weinberg et al. (2013).

[4] `http://www.sdss3.org/surveys/boss.php`
[5] `http://wigglez.swin.edu.au/`
[6] `http://www.sdss3.org/future/eboss.php`
[7] `http://hetdex.org/`
[8] `http://www.mpe.mpg.de/erosita/`
[9] `http://hea.iki.rssi.ru/SRG/`

all-sky-survey (eRASS) (e.g., Merloni et al. 2012; Kolodzig et al. 2013b).

Since the emission of X-rays is a generic feature accompanying AGN activity the AGN selection in X-ray band is very effective (e.g., Brandt & Hasinger 2005) and certainly extremely clean compared to the selection in other, lower energy wavebands. As X-rays do not penetrate Earth's atmosphere these observations have to be carried out in space, which severely limits the size of the achievable collective area. Despite that, as we know for example from the Chandra[10] deep field measurements (Alexander et al. 2003; Xue et al. 2011), the amazing AGN densities of $\sim 10^4$ deg^{-2} are in principle achievable. However, to sample efficiently cosmic LSS such high densities are completely unnecessary and even modest AGN number densities, as achievable with eRASS, in combination with large survey volumes, competitive measurements of two-point clustering statistics is possible.

It is clear that in order to obtain distance information X-ray surveys have to be complemented with optical spectroscopic follow-up, i.e. just as in case of the optical redshift surveys, only imaging part is replaced with imaging in the X-ray band.

In this paper we investigate in detail the potential of the X-ray-selected AGN for probing the cosmic LSS. To have a clear benchmark of achievable quality we focus on the ability to detect the BAO in the clustering power spectrum. As the BAO represent a $\sim 5\%$ modulation on top of the smooth broad-band spectral component, the smooth part itself can be detected with an order of magnitude larger signal-to-noise.

One might wonder, why to attempt to use X-ray AGN for measuring the BAO? After all, almost all that matters to get a BAO measurement is just a large uniformly covered survey volume combined with large number of measured redshifts. Beyond the primary requirement for the redshifts to be easily obtainable, the particular type of objects used is only of secondary importance. Indeed, this is the way most of the upcoming BAO surveys are optimized, e.g. in case of ELGs one only needs to detect the location of a few emission lines without the need to go down to the continuum level. These type of low quality spectra are fine for the BAO measurements but are quite possibly not that useful for other astrophysical studies. Instead, it would be nice if the cosmological BAO surveys could also be useful for the study of the astrophysics of the target objects. And this, we argue, could easily be the case with X-ray-selected AGN. Indeed, the clean sample of AGN detected in X-ray band would certainly facilitate the study of accreting supermassive black holes (SMBHs) in the centers of galaxies, arguably one of the most remarkable discoveries of modern astrophysics, throughout cosmic time. Also, additional synergetic effect might be expected from the fact that imaging and necessary spectroscopic follow-up are done in completely separate parts of the electromagnetic spectrum, this way helping to probe the physics of the AGN in somewhat broader context. It is also important to realize that due to the actuality of the topic of evolution of SMBHs the optical follow-up of X-ray AGN samples detected by upcoming surveys like eROSITA (Predehl et al. 2010) will be done in one way or the other, and so using these AGN samples as

---

[10] http://chandra.harvard.edu/
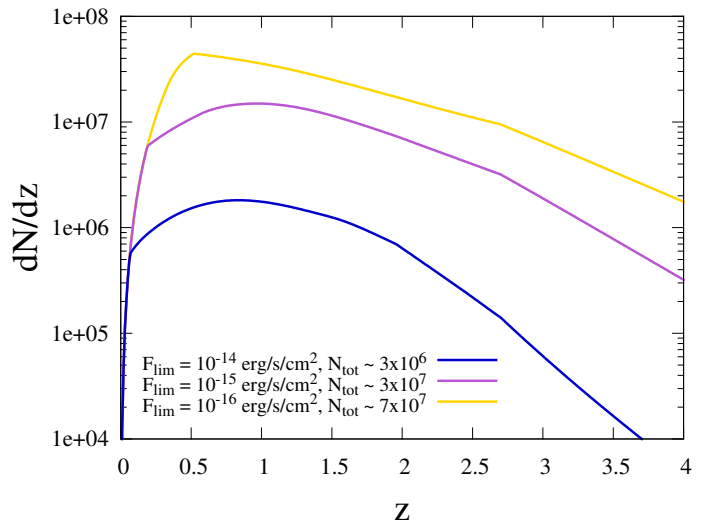
Fig. 1. Redshift distributions of X-ray-selected AGN, assuming the whole sky coverage and soft-band limiting fluxes as shown in the legend. Total numbers of AGN are also given there.

probes of the LSS can be considered at least as an extra auxiliary science possible.

Due to the importance of the optical follow-up, for turning the X-ray selected AGN samples into genuine LSS probes, in this paper we aim to derive the necessary criteria for the quality of the spectroscopic/photometric redshifts. As stated above, we focus on the ability to detect the BAO, i.e., the corresponding broad-band clustering signal is then detectable with an order of magnitude larger signal-to-noise.

Our paper is organized as follows. In Section 2 we present an initial feasibility study, Section 3 provides a short description of the modeling details, followed by our main results in Section 4. Our summary and conclusions are given in Section 5.

Throughout this paper we assume flat ΛCDM cosmology with $\Omega_m = 0.3$, $\Omega_b = 0.05$, $h = 0.7$ and $\sigma_8 = 0.8$.

2. Initial feasibility study

To assess the potential usefulness of X-ray-selected AGN for probing LSS we start here with a brief feasibility study. As the X-ray instruments are typically most efficient in the soft X-ray band we focus on X-ray AGN selection in $0.5 - 2$ keV energy range. To calculate the AGN redshift distributions for various limiting X-ray fluxes $F_{\rm lim}$, we use the soft-band AGN luminosity function as determined by Hasinger et al. (2005) along with corrections for $z > 2.7$ as proposed in Brusa et al. (2009) (see also Kolodzig et al. (2013b)). Note that the AGN luminosity function of Hasinger et al. (2005) includes Type I AGN only. Therefore it under-predicts somewhat the total number of objects, the discrepancy with observed source counts increasing towards the low flux end. As long as the correlation properties of Type I and Type II AGN are similar, this is not critically important for the main purpose of our calculations, but it will lead to an underestimation of the confidence level of the BAO detection, predicted below.

In Fig. 1 we show the resulting AGN redshift distributions dN/dz for three $F_{\rm lim}$ values: 10^{-14}, 10^{-15}, and

10^{-16} erg/s/cm^2. Here the full sky coverage is assumed. The first of the above limiting flux values is typical for the upcoming eRASS (e.g., Kolodzig et al. 2013b). The second, 10^{-15} erg/s/cm^2, is approximately the soft-band limiting point source flux for the existing ~ 2 deg^2 XMM-COSMOS field (Hasinger et al. 2007), and, depending on the final survey strategy, may also be almost reachable (albeit close to the confusion limit) in 'pole regions' of the eRASS where over ~ 100 deg^2 the exposure will be significantly deeper compared to the sky average (e.g., Kolodzig et al. 2013b). The smallest limiting fluxes we consider correspond to 10^{-16} erg/s/cm^2, which is about an order of magnitude larger than typical for the current deepest X-ray fields, i.e., Chandra Deep Field South (Xue et al. 2011) and Chandra Deep Field North (Alexander et al. 2003).

In Fig. 1 we see that, quite irrespective of the $F_{\rm lim}$ value, the X-ray AGN distribution peaks around $z \sim 1$. The expected total number of AGN over the full sky computed with the AGN X-ray luminosity function of Hasinger et al. (2005) is $\sim 3 \times 10^6$, $\sim 3 \times 10^7$, and $\sim 7 \times 10^7$ for $F_{\rm lim} = 10^{-14}, 10^{-15}$, and 10^{-16}, respectively. In the redshift range $z = 0.5 - 2.5$, which is used extensively throughout the rest of the paper, the corresponding numbers reduce to $\sim 2 \times 10^6$, $\sim 2 \times 10^7$, and $\sim 5 \times 10^7$, with corresponding sky densities of ~ 60, ~ 500, and ~ 1300 deg^{-2}, respectively. It is interesting to note that once the redshift interval $z = 0.5 - 2.5$ is divided into three subintervals $z = 0.5 - 1.0$, $z = 1.0 - 1.5$, and $z = 1.5 - 2.5$ then, quite independently of the value of $F_{\rm lim}$, each of these contains approximately one third of the objects.

To investigate if the above numbers are large enough to efficiently probe the LSS we have to include some knowledge of the X-ray AGN clustering strength. Several recent studies have shown (e.g., Allevato et al. 2011; Krumpe et al. 2012; Mountrichas et al. 2013) that the clustering of X-ray-selected AGN is well accounted for if they populate group-size dark matter (DM) halos. Throughout this paper we assume a single effective host DM halo mass of $M_{\rm eff} = 2 \times 10^{13}$ M$_\odot$, which is well compatible with clustering bias measurements as given in Allevato et al. (2011) up to redshift of ~ 3.

In Fig. 2 we show the achievable signal-to-noise per Fourier mode at redshift $z = 1$ for the same limiting fluxes as in Fig. 1. To calculate the matter power spectrum we use the approximate fitting forms as given in Eisenstein & Hu (1998). The clustering bias parameter is taken from Sheth et al. (2001) for the effective DM halo mass of 2×10^{13} M$_\odot$, as stated above. The discreteness noise due to finite sampling density is given as usual as $1/n$, where n is the comoving number density of AGN. We see that in case of limiting fluxes of 10^{-15} and 10^{-16} erg/s/cm^2 the achievable signal-to-noise is significantly larger than one for a broad range of wavenumbers (note that shown in Fig. 2 is the signal-to-noise per a Fourier mode, i.e. for the unbinned power spectra). In fact, one does not need so high sampling density,[11] and in terms of optimizing the observational strategy it would be wiser to move to a new field, instead of integrating too long at the same position to achieve lower $F_{\rm lim}$.
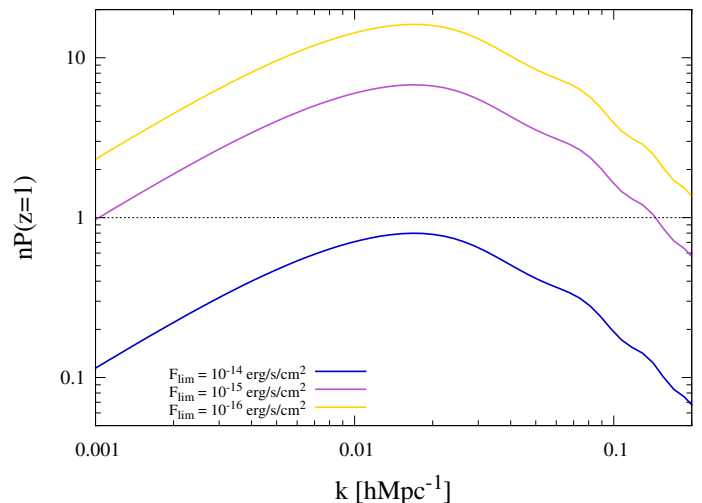
Fig. 2. Signal-to-noise per Fourier mode at redshift $z = 1$ for flux-limited samples of X-ray AGN. The limiting flux values are shown in the legend. The linear clustering bias is assumed to correspond to DM halos with mass $M_{\rm eff} = 2 \times 10^{13}$ M$_\odot$.

With eRASS type of sensitivities, i.e. $F_{\rm lim} \simeq 10^{-14}$, one is typically undersampling the large-scale density field. Even then, the above mild undersampling can be hugely compensated by a large survey volume, which in the end can still lead to a tight measurement of the two-point clustering signal (Kolodzig et al. 2013a).

Thus, we conclude that the with the limiting flux in the $\sim 10^{-16} - 10^{-14}$ erg/s/cm^2 range, typical for modern narrow and large field surveys, there is enough X-ray AGN to efficiently probe the LSS at redshifts $z \sim 0 - 3$, with the tightest clustering measurements possible around redshift $z \sim 1$.

3. Modeling details

Compared to our earlier study in Kolodzig et al. (2013a) we have implemented several improvements in our two-point function calculations: (i) in addition to the auto-spectra we also include the information encoded in cross-spectra, (ii) we include linear redshift-space distortions, (iii) BAO damping is treated by using the resummed Lagrangian perturbation theory (LPT) approach by Matsubara (2008). Most importantly, as the main focus of this study is to try to determine necessary criteria for the quality of the optical follow-up, we have significantly more elaborate treatment for the photometric redshift (photo-z) errors.[12]

In order to have a flexible treatment for the smoothing caused by the photometric redshifts we divide the full survey volume into narrow photo-z bins and calculate projected power spectra within bins, i.e. auto-spectra, and also cross-spectra between the bins.

Even if full spectroscopic redshifts are available it might be beneficial to carry out clustering analysis by dividing survey volume into narrow redshift bins and calculate projected auto and cross power spectra. From these spectra full three dimensional power spectrum for the whole volume can

---

[11] In practice, for two-point clustering measurements an effective signal-to-noise ratio of $nP \sim 3$ is often more than enough, with significantly larger values leading to a wasteful oversampling of the density field.

[12] In Kolodzig et al. (2013a), to approximate the impact of photo-z errors we simply adjusted the effective width of the redshift bins accordingly.

3

be recovered up to some wavenumber k_{max} that depends on how narrow redshift binning one has decided to use. This is quite in contrast to how currently spectroscopic galaxy surveys are analyzed. There one typically makes a single three dimensional power spectrum measurement using the whole survey volume. But this way one looses sensitivity to evolutionary trends of the clustering properties with redshift. Also, by calculating projected auto- and cross-spectra in narrow bins one does not need to assume any fiducial cosmological model in order to convert redshifts to comoving distances, and thus data analysis can be done just once and for all. Even though the consequent model fitting is computationally more demanding, especially when the number of redshift bins gets large, e.g. ~ 100, this in reality is not an obstacle for the modern computational technology.

3.1. Photometric redshifts

We assume that the conditional probability distribution for photometric redshift z_p given spectroscopic redshift z can be decomposed as

$$P(z_p|z) = (1 - f_{fail})P_1(z_p|z) + f_{fail}P_2(z_p|z). (1)$$

Here P_1 is assumed to be a truncated Gaussian with the requirement that $P_1(z_p < 0|z) = 0$, and thus the probability distribution function (pdf) with appropriate normalization is given as

$$P_1(z_p|z) = \sqrt{\frac{2}{\pi}} \frac{\exp\left[-\frac{(z-z_p)^2}{2\sigma^2(z)}\right]}{\sigma(z)\left[1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}\sigma(z)}\right)\right]}. (2)$$

For $\sigma(z)$ we assume a commonly used form

$$\sigma(z) = \sigma_0(1 + z). (3)$$

The quantity f_{fail} in Eq. (1) represents the fraction of catastrophic errors. For simplicity we assume that in the case of a catastrophic failure the measured redshift value can be represented by a random number uniformly distributed between 0 and some z_p^{max}, i.e.

$$P_2(z_p|z) = \frac{1}{z_p^{max}}. (4)$$

As a default value we use $z_p^{max} = 5$ throughout this paper.[13]

Thus, we have assumed photo-z errors to be independent of luminosity and the fraction of catastrophic errors to be independent of redshift. In realistic situation this is certainly not true (see Salvato et al. (2011) for current performance of photometric redshifts for X-ray-selected AGN), but since in our study the AGN clustering is taken to be independent of luminosity and we consider only a cumulative clustering signal over broad redshift range, one can always use effective luminosity- and redshift-averaged values for σ_0 and f_{fail}.

The pdf of true redshifts corresponding to the objects selected in the i-th photo-z bin $z_{p_1}^{(i)} - z_{p_2}^{(i)}$ is given as (see also, e.g., Budavári et al. 2003)

$$f^{(i)}(z) = f(z)\int_{z_{p_1}^{(i)}}^{z_{p_2}^{(i)}} P(z_p|z)\,\mathrm{d}z_p. (5)$$

Here $f(z)$ is the true underlying full redshift distribution function, i.e.

$$f(z) = \frac{\frac{\mathrm{d}N}{\mathrm{d}z}}{\int \frac{\mathrm{d}N}{\mathrm{d}z}\,\mathrm{dz}} = \frac{1}{N_{tot}}\frac{\mathrm{d}N}{\mathrm{d}z}. (6)$$

Taking $P(z_p|z)$ from Eq. (1) we arrive at the result

$$f^{(i)}(z) = f(z)\left[(1 - f_{fail})\frac{\mathrm{erf}\left(\frac{z_{p_2}^{(i)}-z}{\sqrt{2}\sigma(z)}\right) - \mathrm{erf}\left(\frac{z_{p_1}^{(i)}-z}{\sqrt{2}\sigma(z)}\right)}{1 + \mathrm{erf}\left(\frac{z}{\sqrt{2}\sigma(z)}\right)} + \right.$$
$$\left. + f_{fail}\frac{z_{p_2}^{(i)} - z_{p_1}^{(i)}}{z_p^{max}}\right], (7)$$

with $\sigma(z)$ given by Eq. (3).

Thus, apart from the photo-z bin boundary values $z_{p_1}^{(i)}$, $z_{p_2}^{(i)}$, and underlying true redshift distribution $f(z)$ (see Fig. 1), in our model the radial selection functions $f^{(i)}(z)$ are fully determined by two parameters only: σ_0 and f_{fail}. In this paper we assume these parameters to stay within ranges $\sigma_0 = 10^{-3} - 10^{-1}$ and $f_{fail} = 0 - 0.5$. It should be clear that for the large-scale clustering analysis the highest accuracy photo-z choice $\sigma_0 = 10^{-3}$ is in practice almost equivalent to the availability of the full spectroscopic measurements.

In Fig. 3 with solid lines we show some examples of the resulting $\mathrm{d}N^{(i)}/\mathrm{d}z \equiv N_{tot}f^{(i)}(z)$ distributions. Here we assume eight equal size photo-z bins between $z_p = 0.5 - 2.5$ and limiting flux $F_{lim} = 10^{-14}$ erg/s/cm^2. Three panels correspond to various choices of σ_0 and f_{fail} as indicated in the upper right-hand corners. The dashed line corresponds to the full underlying $\mathrm{d}N/\mathrm{d}z$ distribution.

3.2. Choice of the redshift bins width
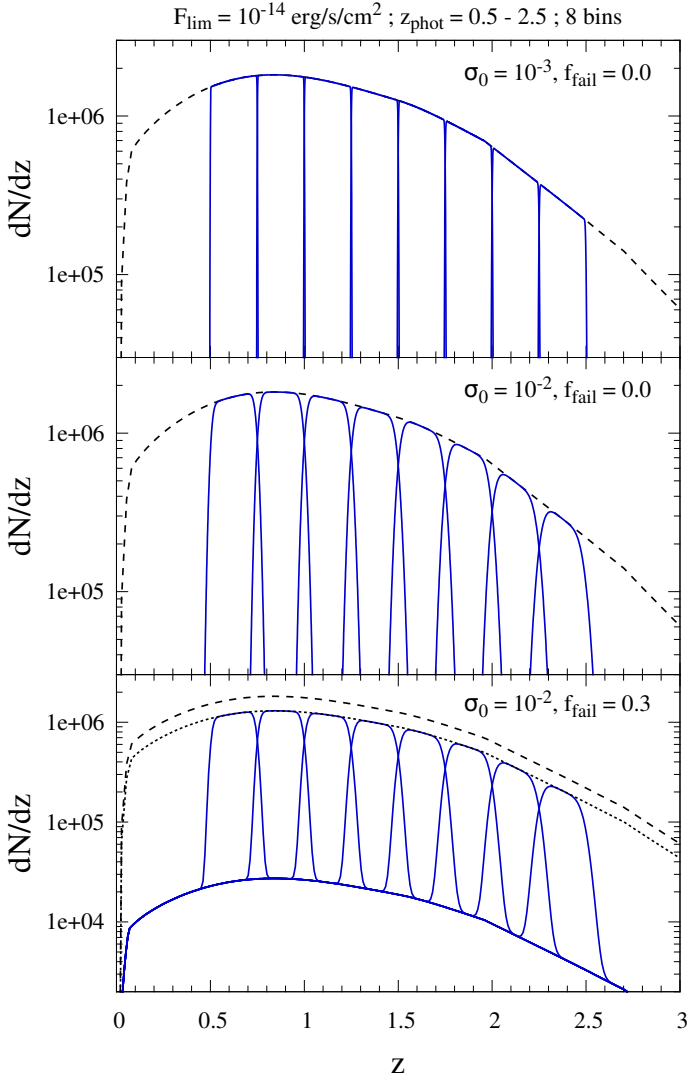
How many photo-z bins should one use?
In reality there is no need for using redshift bins much smaller than $\sigma(z)$, since photometric redshift errors in that case would highly correlate neighboring bins and by using more of them we just make our calculations computationally more demanding, without gaining any extra information.

There is also another bound for the redshift bin size one has to consider. Even in case photo-z becomes very accurate, e.g., by approaching the accuracy of spectroscopic redshifts, we still should not use extremely narrow redshift bins, since in that case we would effectively include nonlinear radial modes into our clustering analysis. To avoid these complications one typically applies some wavenumber cutoff to eliminate nonlinear Fourier modes.

The variance of the underlying DM density field at redshift z can be expressed as

$$\sigma_{DM}^2(z) = \frac{g^2(z)}{2\pi^2}\int k^2 P(k|z=0)\,\mathrm{d}k, (8)$$

where $g(z)$ is the linear growth factor and $P(k|z=0)$ is the matter power spectrum at $z = 0$. One typically cuts this integral at some k_{max}, such that the resulting $\sigma_{DM}^2(z)$ stays safely below one. At $z = 0$ one usually assumes k_{max} to be in the range $\sim 0.1 - 0.2\,h\,\mathrm{Mpc}^{-1}$, while at higher redshifts one is able to accommodate larger values for k_{max}. Here we

---

[13] As it turns out, our results have little dependence by changing z_p^{max}, by e.g. ± 2.

Fig. 3. Radial selection functions for eight equal size photo-z bins between $z_p = 0.5 - 2.5$. Three different combinations of σ_0 and f_{fail} are used and the limiting flux is assumed to be $F_{lim} = 10^{-14}$ erg/s/cm^2. The dashed lines show the full underlying spectroscopic redshift distribution.

obtain k_{max} as a function of redshift by demanding that the above integral for $\sigma_{DM}^2(z)$ with appropriately adjusted $k_{max}(z)$ stays the same as in $z = 0$ case, i.e.

$$g^2(z) \int_0^{k_{max}(z)} k^2 P(k|z=0) \mathrm{d}k = g^2(0) \int_0^{k_{max}(0)} k^2 P(k|z=0) \mathrm{d}k. \tag{9}$$

The numerical solution $k_{max}(z)$ for the above equation for various values of $k_{max}(0)$ are shown in the lower panel of Fig. 4.

The maximal wavenumber $k_{max}(z)$ corresponds to the minimal 'allowed' comoving spatial scale $r_{min}(z) = 2\pi/k_{max}(z)$. In case this spatial interval is placed at redshift z and oriented along the line of sight, it corresponds to the redshift interval $\Delta z(z)$ as plotted in the upper panel of Fig. 4. We see that for the conservative choice of $k_{max}(0) = 0.1 \, h\,\mathrm{Mpc}^{-1}$ redshift interval Δz stays remarkably constant with a value close to 0.02. In the following we adopt $\Delta z = 0.02$ as the minimal allowed size for the redshift bin.
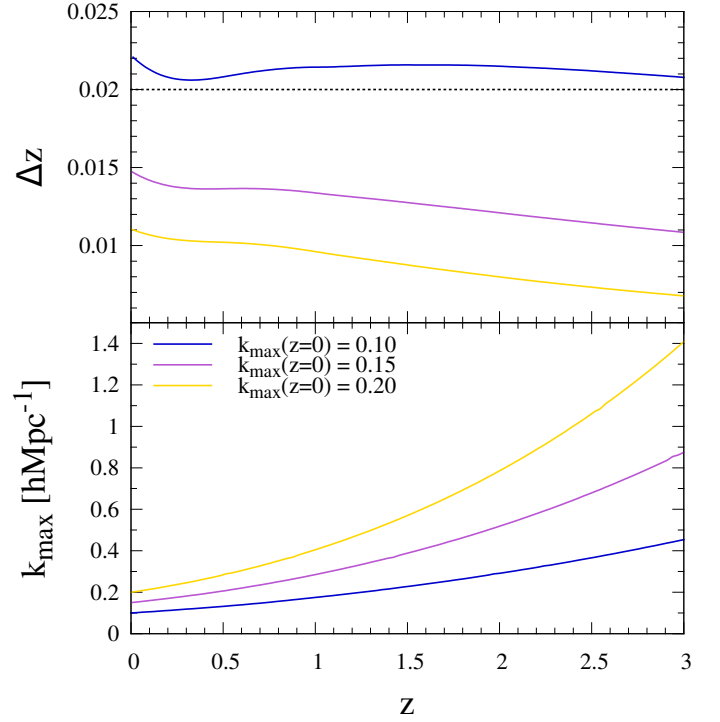
Fig. 4. Maximum wavenumber as a function of redshift, $k_{max}(z)$, calculated from Eq. (9) for three different values of $k_{max}(0)$ (lower panel) and corresponding redshift intervals (upper panel).

As noted above, due to photo-z smoothing along radial direction there is no benefit of considering redshift bins significantly smaller than $\sigma(z)$. It turns out that by choosing $\Delta z = \max(\sigma_0, 0.02)$ one already obtains well converged results.

To fully eliminate small scale nonlinear modes one has to impose also some cutoff ℓ_{max} for the angular multipole number. Throughout this paper we have used $\ell_{max} = 500$. The justification is the following. The AGN distribution peaks at $z \sim 1$, which corresponds to a comoving distance of $\sim 2300 \, h^{-1}\,\mathrm{Mpc}$. The cutoff $k_{max}(0) = 0.1 \, h\,\mathrm{Mpc}^{-1}$, which was also applied for the radial modes, corresponds to $\sim 0.2 \, h\,\mathrm{Mpc}^{-1}$ at $z \sim 1$ (see Fig. 4), which results in $\ell_{max} \sim 0.2 \times 2300 \sim 500$. For simplicity, as in this study we look at cumulative signal from broad redshift range ($z = 0.5 - 2.5$), we keep this value fixed, independent of redshift.

3.3. Auto- and cross-spectra

In the linear regime the angular clustering spectra can be given as (Fisher et al. 1994; Huterer et al. 2001; Padmanabhan et al. 2007; Asorey et al. 2012)

$$C_\ell^{(ij)} = \frac{2}{\pi} \int W_\ell^{(i)}(k) W_\ell^{(j)}(k) P(k|z=0) k^2 \, \mathrm{d}k. \tag{10}$$

If $i = j$ we obtain auto-spectra, $i \neq j$ gives us cross-spectra between different redshift bins. The projection kernel for the i-th redshift bin, $W_\ell^{(i)}(k)$, is given as a sum of two parts

$$W_\ell^{(i)}(k) = W_\ell^{(i),1}(k) + W_\ell^{(i),2}(k). \tag{11}$$

Here the first part takes care of the projection in real space while the second one includes the effect of redshift-space

distortions. They are given as

$$W_\ell^{(i),1}(k) = \int j_\ell(kr) f^{(i)}(r) g(r) b(r) \, dr \,, \tag{12}$$

$$W_\ell^{(i),2}(k) = \int \left[\frac{2\ell^2 + 2\ell - 1}{(2\ell + 3)(2\ell - 1)} j_\ell(kr) - \right.$$
$$- \frac{\ell(\ell - 1)}{(2\ell - 1)(2\ell + 1)} j_{\ell-2}(kr) -$$
$$\left. - \frac{(\ell + 1)(\ell + 2)}{(2\ell + 1)(2\ell + 3)} j_{\ell+2}(kr) \right] f^{(i)}(r) g(r) f(r) \, dr \,. \tag{13}$$

Here the integrals are over comoving distance r, j_ℓ is the ℓ-th order spherical Bessel function, $f^{(i)}(r)$ is the normalized distribution of comoving distances of LSS tracers (AGN) in the i-th photo-z bin, i.e. $f^{(i)}(r) = f^{(i)}(z) \frac{dz}{dr}$ with $f^{(i)}(z)$ given in Eq. (7), g and b are the linear growth factor and clustering bias parameter, respectively. In Eq. (13) $f(r) \equiv \frac{d \ln g}{d \ln a}(r)$, with a the scale factor, is the linear growth rate at comoving distance r.

3.4. BAO damping

As the focus of the AGN clustering analysis performed in this paper is the ability to detect the BAO, to be more accurate we have to include the damping of linear BAO due to nonlinear evolution of the cosmic density field. Here we adopt the two-point function results for DM halos derived from the resummed LPT by Matsubara (2008). Even with BAO damping included we still use the simple form of Eq. (10) where the k-independent linear growth factor separates out. Namely, we make the following approximate replacement in Eq. (10)

$$P(k|z = 0) \longrightarrow \sqrt{\frac{P^{\text{LPT}}(k, z_i) P^{\text{LPT}}(k, z_j)}{g(z_i) g(z_j)}} \,, \tag{14}$$

where P^{LPT} is the power spectrum calculated with LPT, and z_i, z_j are central redshifts in the i-th and j-th bin, respectively.

In Fig. 5 we illustrate the damping of BAO at various redshifts ($z = 0, 1, 2$) for DM halos with mass 2×10^{13} $M_\odot$ and similarly for the full DM distribution. Linear theory BAO is also shown for comparison. Thinner lines correspond to higher redshifts. As expected, for the DM the damping behavior as a function of redshift is simply monotonic, with the BAO steadily approaching linear theory prediction at higher redshifts. The low redshift behavior for the DM halos mimics the trend seen in the full DM distribution, albeit with somewhat reduced amplitude, but at higher redshifts the evolutionary behavior has reversed, with the BAO amplitude becoming more suppressed. This is due to the fact that halos with masses 2×10^{13} $M_\odot$ at $z \sim 2$ are significantly rarer than similar mass halos at lower redshifts, and so in relative terms involve significantly higher level of 'nonlinear processing'.

3.5. Covariance matrix. Confidence level for BAO detection

From the CMB measurements we know that the initial density fluctuations can be described as a Gaussian random
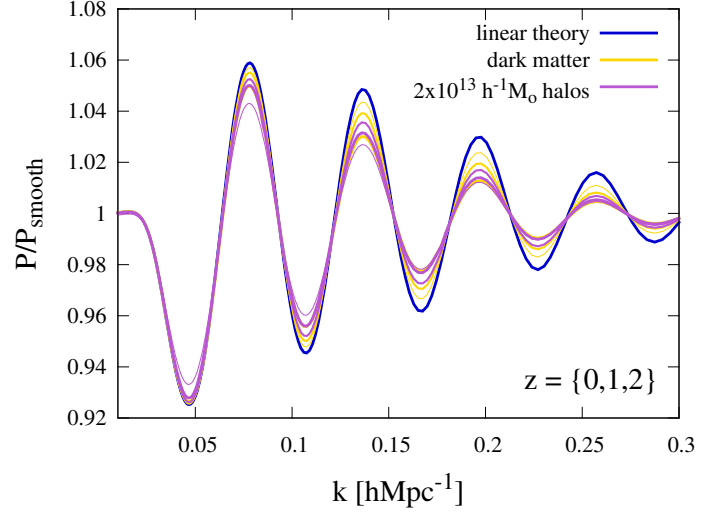
Fig. 5. The damping of BAO at redshifts $z = 0, 1, 2$ for DM and for halos with mass 2×10^{13} $M_\odot$ compared to the linear BAO. Thinner lines correspond to higher redshifts.

field with very good accuracy (Planck Collaboration et al. 2013b). Since the linear evolution does not change this statistical property one can write (using Wick's theorem) the covariance matrix for the angular power spectra (which contain the full information content of the Gaussian field) as follows

$$\mathcal{C}_\ell \equiv \text{cov}\left(C_\ell^{(ij)} C_\ell^{(mn)} \right) =$$
$$= \frac{1}{(2\ell + 1) f_{\text{sky}}} \left[\widetilde{C}_\ell^{(im)} \widetilde{C}_\ell^{(jn)} + \widetilde{C}_\ell^{(in)} \widetilde{C}_\ell^{(jm)} \right] \,, \tag{15}$$

i.e., it can be factorized as a product of power spectra (e.g., Feldman et al. 1994; Tegmark et al. 1997; Meiksin & White 1999; Scoccimarro et al. 1999; Smith 2009; Asorey et al. 2012). Here the number of modes with multipole number ℓ for the full sky, i.e. $2\ell + 1$, is reduced by a factor f_{sky} –the fraction of sky covered by the survey. However, there is also a slight complication due to discrete sampling of the density field. Namely, the auto-spectra that enter the product in Eq. (15) get extra Poisson noise contribution, i.e.,

$$\widetilde{C}_\ell^{(ij)} \equiv C_\ell^{(ij)} + \frac{1}{\mathcal{N}^{(i)}} \delta_{ij} \,. \tag{16}$$

Here δ_{ij} is the Kronecker delta and $\mathcal{N}^{(i)}$ is the number of objects per steradian in the i-th redshift bin. For in-depth discussion on covariance of auto- vs cross-spectra see Smith (2009).

Note that Eq. (15) assumes that covariance matrix for the angular power spectra can be written separately for each multipole ℓ, i.e., it assumes that different multipoles are not coupled, which is only strictly true for the unmasked Gaussian random field. However, in case of reasonably large survey areas the resulting mode coupling is fairly tightly localized in multipole space, and Eq. (15) is a very good approximation in practice. With N redshift bins each matrix $\mathcal{C}_\ell$ has a dimension of $\frac{N}{2}(N + 1) \times \frac{N}{2}(N + 1)$.

Having specified the covariance matrix we can proceed to ask the question how well the BAO can be detected. First we have to specify a smooth template spectrum without acoustic features with respect to what the

BAO is measured. Instead of the standard 'no wiggle' form of Eisenstein & Hu (1998) we use a slightly different form, which we believe has somewhat better behavior. The details are presented in Appendix A. In reality, due to the applied wavenumber cuts these small deviations in the 'no wiggle' spectral form have negligible impact on our results.

To estimate how accurately the BAO can be measured we look at the parametric class of models with power spectra given as

$$P^{A_{\mathrm{BAO}}}(k) = (1 - A_{\mathrm{BAO}})P^{\mathrm{NW}}(k) + A_{\mathrm{BAO}}P(k) \,, \quad (17)$$

i.e., models interpolating between the 'no wiggle' spectral form and the form with a full ΛCDM BAO. Thus A_{BAO} can take the values between zero and one, with the fiducial value $A_{\mathrm{BAO}} = 1$.

The angular spectra can then be given similarly, i.e.,

$$C_\ell^{(ij),A_{\mathrm{BAO}}} = (1 - A_{\mathrm{BAO}})C_\ell^{(ij),\mathrm{NW}} + A_{\mathrm{BAO}}C_\ell^{(ij)} \,. \quad (18)$$

Now one can write down the Fisher information[14] for the parameter A_{BAO} from the spectra at fixed multipole ℓ

$$
\begin{aligned}
F_\ell^{A_{\mathrm{BAO}}} &= \left(\frac{\partial \mathbf{C}_\ell}{\partial A_{\mathrm{BAO}}}\right)^T \mathcal{C}_\ell^{-1} \left(\frac{\partial \mathbf{C}_\ell}{\partial A_{\mathrm{BAO}}}\right) = \\
&= \left(\mathbf{C}_\ell - \mathbf{C}_\ell^{\mathrm{NW}}\right)^T \mathcal{C}_\ell^{-1} \left(\mathbf{C}_\ell - \mathbf{C}_\ell^{\mathrm{NW}}\right) \,,
\end{aligned} \quad (19)
$$

where with $\mathbf{C}_\ell$ ($\mathbf{C}_\ell^{\mathrm{NW}}$) we have denoted the vector with components $C_\ell^{(ij)}$ ($C_\ell^{(ij),\mathrm{NW}}$), i.e., in case of N redshift bins an $\frac{N}{2}(N+1)$-dimensional vector. Since all the multipoles are assumed to be independent, the full Fisher information for A_{BAO} and its standard error are given as

$$F^{A_{\mathrm{BAO}}} = \sum_\ell F_\ell^{A_{\mathrm{BAO}}} \,, \quad (20)$$

$$\sigma_{A_{\mathrm{BAO}}} = \frac{1}{\sqrt{F^{A_{\mathrm{BAO}}}}} \,. \quad (21)$$

Since our fiducial $A_{\mathrm{BAO}} = 1$ the confidence level (CL) for the BAO detection can be expressed as[15]

$$\mathrm{CL} = \frac{1}{\sigma_{A_{\mathrm{BAO}}}} = \sqrt{F^{A_{\mathrm{BAO}}}} \,. \quad (22)$$

In case we have N redshift bins, the number of spectra one has to calculate, to find CL for the BAO detection as presented above, is $\frac{N}{2}(N+1)$. In case of a survey from $z = 0.5$ up to $z = 2.5$ with our minimal $\Delta z = 0.02$, i.e. $N = 100$, this leads to a total of 5050 spectra. However, it is clear that cross-spectra between far away bins are quite small in amplitude, and thus in practice can be neglected.

To speed up the computation we calculate CL of Eq. (22) in an iterative manner: (i) we start with auto-spectra $C_\ell^{(ii)}$ only, i.e. the initial covariance matrix dimension is $N \times N$; (ii) in the next step we include to the previously calculated auto-spectra cross-spectra between the

---

[14] For a detailed presentation of Fisher information in Gaussian cases see Tegmark et al. (1997).

[15] One could have arrived to this result also by performing model comparison through a likelihood ratio test. For the Gaussian case, if the dimensionalities of the model parameter spaces differ by a single unit, the resulting $\Delta\chi^2$ is distributed as a χ^2-distribution with one degree of freedom, and thus one model is preferred over the other with $\mathrm{CL} = \sqrt{\Delta\chi^2}$, which is equivalent to the result in Eq. (22).
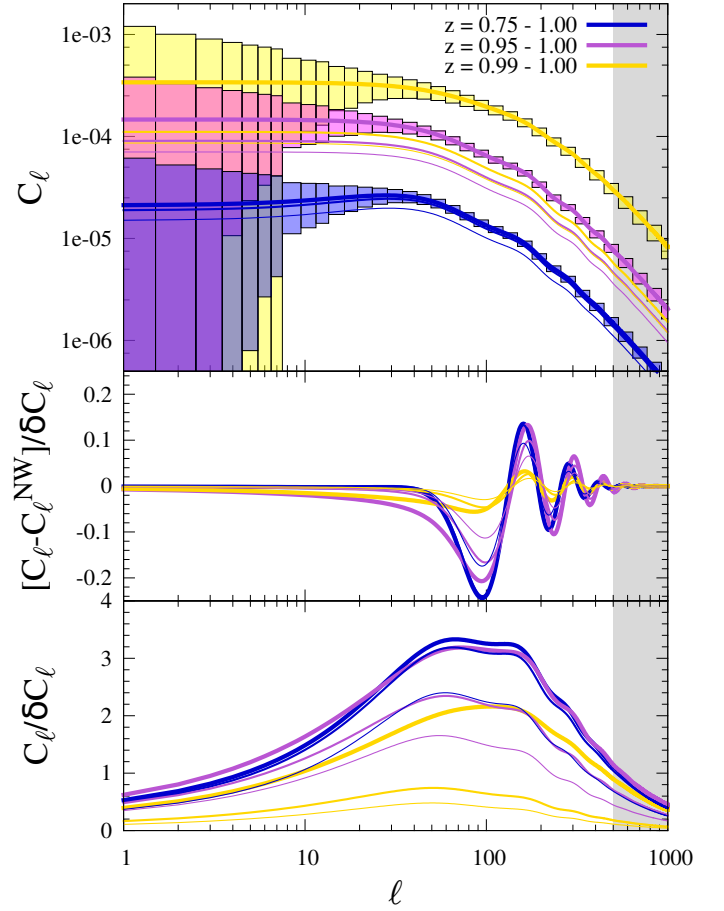

Fig. 6. Example angular auto-spectra (upper panel) and unbinned signal-to-noise ratios for the BAO (middle panel) and for the whole spectra (lower panel). Three different choices for the size of the photo-z bin are shown. The lines with three different thickness levels, starting from the thickest, assume the optical follow-up parameters (i) $\sigma_0 = 10^{-3}$, $f_{\mathrm{fail}} = 0.0$, (ii) $\sigma_0 = 10^{-2}$, $f_{\mathrm{fail}} = 0.0$, and (iii) $\sigma_0 = 10^{-2}$, $f_{\mathrm{fail}} = 0.3$, respectively. In the upper panel for the first of the above three cases binned 1σ error boxes are also shown. Light gray vertical stripe marks the small-scale modes excluded from our analysis.

neighboring bins, i.e. $C_\ell^{(i,i+1)}$; (iii) step by step we keep on including $C_\ell^{(i,i+2)}$, $C_\ell^{(i,i+3)}$, etc, each time calculating CL from Eq. (22); (iv) once the value for CL does not change by more than 1% we consider the obtained CL to be converged and stop our calculation.

3.6. Example spectra

In the upper panel of Fig. 6 we show angular auto-spectra for three different choices for the width of the photo-z bin as specified in the upper right-hand corner. Here we have assumed a full sky coverage down to soft-band limiting flux of 10^{-14} erg/s/cm^2. Thick, middle-thick and thin lines correspond to the optical follow-up with (i) $\sigma_0 = 10^{-3}$, $f_{\mathrm{fail}} = 0.0$, (ii) $\sigma_0 = 10^{-2}$, $f_{\mathrm{fail}} = 0.0$, and (iii) $\sigma_0 = 10^{-2}$, $f_{\mathrm{fail}} = 0.3$, respectively. Only for the first of the above three cases we show binned 1σ error boxes. The small-scale modes with $\ell > \ell_{\mathrm{max}} = 500$, which are not included in our analysis, are marked with a gray shaded band.
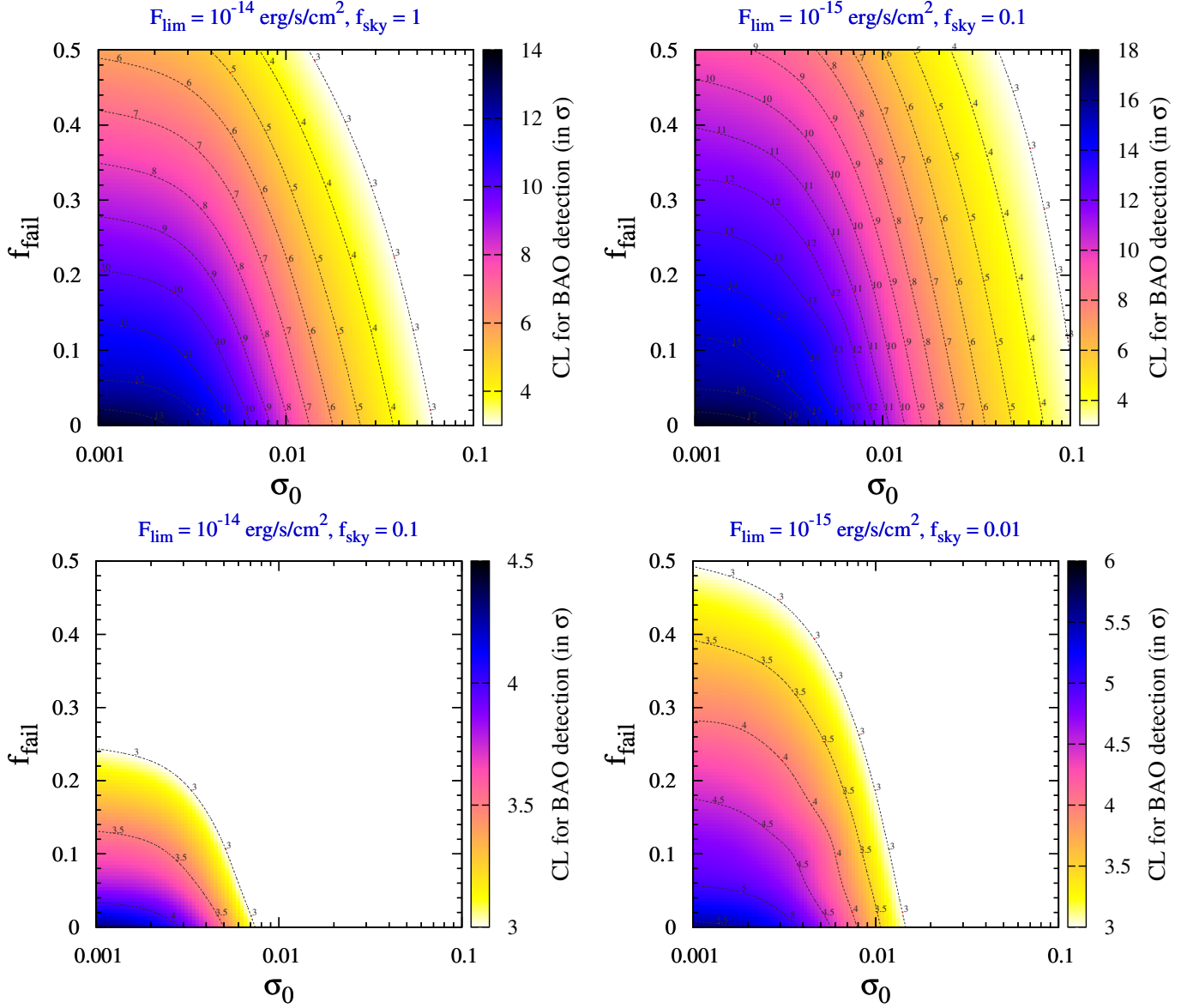
Fig. 7. Confidence levels for the BAO detection as a function of the quality of the optical follow-up, specified by σ_0 and $f_{\rm fail}$, for limiting soft-band X-ray fluxes $F_{\rm lim} = 10^{-14}$ (left-hand panels) and 10^{-15} erg/s/cm^2 (right-hand panels). For $F_{\rm lim} = 10^{-14}$ (10^{-15}) cases with a complete and 10% (10% and 1%) sky coverage are shown. Redshift range $z = 0.5 - 2.5$ is assumed.

One can see how the angular clustering strength increases once reducing the size of the photo-z bin, since the radial selection function gets correspondingly more sharply peaked. This behavior is qualitatively understandable, because smaller radial projection length leads to less smoothing and thus keeps higher level of coherence in the projected field. More quantitatively, from Eq. (10) one obtains for a narrow selection function (i.e., can be replaced with a delta function) and for sufficiently high ℓ, so that spherical Bessel function can be assumed to be sharply peaked at $\ell = kr$ (also for high ℓ redshift distortion part vanishes), $C_\ell \propto \frac{P}{r\Delta r}$, with r and Δr the comoving distance and its interval, corresponding to z and Δz, respectively. Effective Δr is increased by choosing wider photo-z bin and also by having larger values for σ_0 and $f_{\rm fail}$, i.e., poorer quality photo-z.

Even though for the narrowest photo-z bin shown in Fig. 6 the signal has the highest amplitude, the noise also starts to increase considerably due to lower number of AGN

available in narrow redshift bin. This is more clearly visible in the middle and lower panels of Fig. 6 where we show the unbinned signal-to-noise for the BAO part and for the whole spectrum, respectively. One can see that the signal-to-noise for the full spectrum is approximately an order of magnitude larger than for the BAO only, since the BAO represents only a $\sim 5\%$ fluctuation on top of the smooth broad-band spectrum.

Although in case of narrow redshift bins an increased Poisson noise leads to a reduced signal-to-noise per redshift bin, by joining the information from the larger total number of available bins, one is able to achieve globally higher signal-to-noise, as will be shown in the next section.

4. Main results

The main results of this paper are shown in Figs. 7 and 8. In Fig. 7 the limiting soft-band flux and survey area,
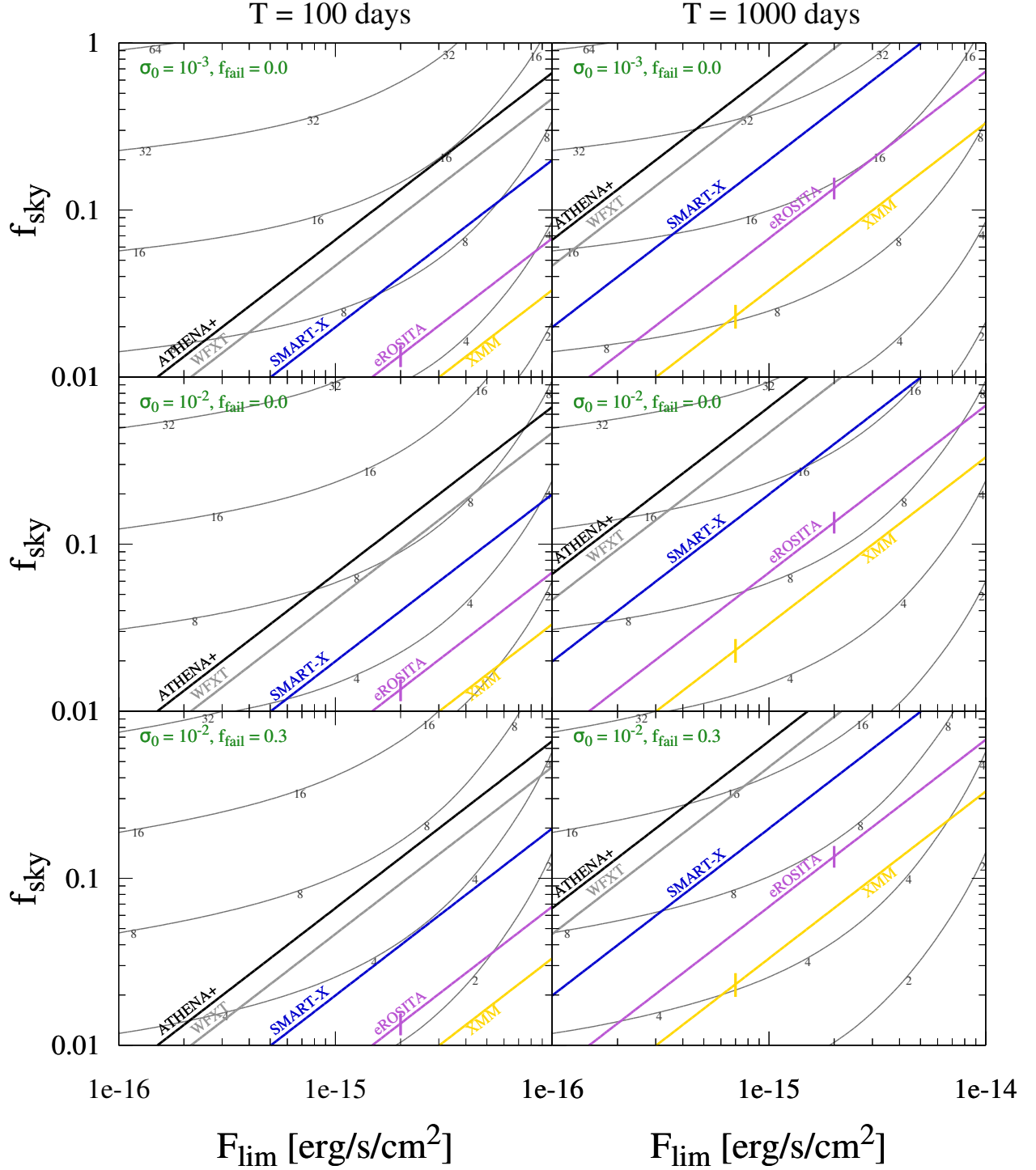
Fig. 8. Achievable signal-to-noise for the BAO detection as a function of limiting flux and sky coverage, shown with numbered contour lines. From top to bottom three different choices for the quality of optical follow-up are considered. The diagonal solid lines show possible survey size-depth relations for several existing and proposed X-ray instruments assuming a total observation time of 100 days (left-hand panels) and 1000 days (right-hand panels). Redshift range $z = 0.5 - 2.5$ is assumed. In case of eROSITA and XMM-Newton the small vertical lines at $F_{\rm lim} \simeq 2 \times 10^{-15}$ and 7×10^{-16} erg/s/cm^2 mark approximate confusion noise limits for these instruments. For the other telescopes confusion noise is not achieved in the considered flux range.

i.e. the parameters describing X-ray survey, are kept fixed, while the parameters describing optical follow-up σ_0 and $f_{\rm fail}$ are allowed to vary within ranges $10^{-3} - 10^{-1}$ and $0.0 - 0.5$, respectively. The range of redshifts is constrained to be between 0.5 and 2.5. In Fig. 8 the situation is reversed, now

the follow-up parameters are kept fixed and X-ray survey parameters, $F_{\rm lim}$ and $f_{\rm sky}$, vary.

In Fig. 7 we have considered two values for $F_{\rm lim}$: 10^{-14} and 10^{-15} erg/s/cm^2. The first of these values is achievable by the eRASS for the entire extragalactic sky, while the

second value is more characteristic of the deeper surveys, like XMM-COSMOS; a comparable sensitivity may be also achieved in the polar regions of eRASS covering about ~ 100 deg^2. For $F_{\rm lim} = 10^{-14}$ and 10^{-15} we have presented cases with $f_{\rm sky} = \{1, 0.1\}$ and $\{0.1, 0.01\}$, respectively.

For example, for a survey with $F_{\rm lim} = 10^{-14}$ and full sky coverage, $\gtrsim 5\sigma$ detection of the BAO is achievable once $\sigma_0 \sim 0.01$ and $f_{\rm fail} \lesssim 0.3$. A similar quality optical follow-up in combination with $F_{\rm lim} = 10^{-15}$ and $f_{\rm sky} = 0.1$ would lead to $\gtrsim 8\sigma$ detection. With $F_{\rm lim} = 10^{-14}$ and 10% sky coverage the BAO can only be detected provided very high accuracy photometric redshifts or a full spectroscopy is available. If 1% of the sky is covered down to $F_{\rm lim} = 10^{-15}$ photometric redshifts with $\sigma_0 \sim 0.01$ and $f_{\rm fail} \sim 0$ should lead to a marginal BAO detection with $\sim 3\sigma$.

In Fig. 8 we investigate dependence of the CL for the BAO detection as a function of the survey parameters. The contour lines labelled with numbers show achievable CL as a function of $F_{\rm lim}$ and $f_{\rm sky}$, while keeping the follow-up parameters σ_0 and $f_{\rm fail}$ fixed to the values as denoted at the top left-hand corners of the panels. In addition to the above contour lines, which are the same for respective left- and right-hand panels, we show the loci of possible surveys with the XMM-Newton[16] (the only existing X-ray instrument with reasonable survey capabilities) and several upcoming/proposed X-ray instruments, assuming total survey duration of 100 days (left-hand panels) and 1000 days (right-hand panels).

These lines are calculated as follows. The fraction of sky covered by a survey can be expressed as

$$f_{\rm sky} = \frac{T}{t} f_{\rm FoV} \,, \qquad (23)$$

where T is the total survey duration, t is the time spent per pointing, and $f_{\rm FoV}$ the fraction of sky covered by the instrument's field of view (FoV). We will assuming that limiting flux for the AGN detection scales as

$$F_{\rm lim} \propto (A_{\rm eff} \times t)^{-1} \,, \qquad (24)$$

where $A_{\rm eff}$ is the soft-band effective area of the instrument. In reality, limiting flux for a point source detection also depends on the angular resolution of the instrument (via background) and parameters of its orbit around the Earth. For the purpose of this illustrative calculation we will ignore these complications.

As for most of the future X-ray instruments considered in this work, in particular for SMART-X, WFXT, ATHENA+, the angular resolution is sufficiently high, confusion is not an issue for the fluxes considered in this study. However, observations with eROSITA and XMM-Newton, having similar angular resolution of $\sim 15''$ 'on axis', become confusion noise dominated at fluxes below $F_{\rm lim} \sim 1 - 2 \times 10^{-15}$ erg/s/cm^2. In Fig. 8 the approximate confusion noise limits for eROSITA and XMM-Newton are marked by small vertical lines on the corresponding curves.

With equations (23) and (24), we can write

$$f_{\rm sky} \propto T F_{\rm lim} A_{\rm eff} f_{\rm FoV} \,. \qquad (25)$$

Here the combination $A_{\rm eff} f_{\rm FoV}$ is the grasp of the instrument. The above scaling is normalized such that for

[16] http://xmm.esac.esa.int/

Table 1. Assumed values of the effective area and size of the field of view for the X-ray instruments of Fig. 8.

	$A_{\rm eff}$ [cm^2]	FoV [deg^2]
XMM-Newton	2000	0.25
eROSITA	1200 [(*)]	0.85
SMART-X	20000	0.15
WFXT	7000	1.0
ATHENA+	20000	0.5

[(*)] FoV-averaged

eRASS, with $T = 4$ yr and with grasp taken from instrument's webpage http://www.mpe.mpg.de/erosita/, gives $F_{\rm lim} = 1.1 \times 10^{-14}$ erg/s/cm^2 for $f_{\rm sky} = 1$ (see e.g., Kolodzig et al. 2013b). The values of grasp in the soft band for ATHENA+[17], WFXT[18], SMART-X[19], and XMM-Newton were computed from the parameters given in the instruments' manuals and websites. The particular numbers for $A_{\rm eff}$ and FoV used in our calculations are given in Table 1.

Only for eROSITA we have used $A_{\rm eff}$ appropriately averaged over the instrument's FoV and over the energy range of $0.5-2$ keV, with the additional assumption that a typical AGN has a photon index $\Gamma = 1.9$[20] The value of $A_{\rm eff}$ calculated this way for eROSITA is used to obtain appropriate normalization factors in Eqs. (25) and (27). For the other instruments the values of $A_{\rm eff}$ as given in Table 1 are somewhat overestimated, since averaging over the FoV would certainly reduce them somewhat. Since for the future instruments we have no information available regarding the behavior of the 'off axis' effective area, for simplicity in these cases we have used the 'on axis' $A_{\rm eff}$ at $E \sim 1$ keV, instead.

From Fig.8, one can see that, for example, for eROSITA type of instrument, an all-sky survey would not be the best strategy to study BAO. In particular, for a $T = 1000$ days survey, the optimal BAO detection of $\sim 16\sigma$ will be achieved when covering $f_{\rm sky} \sim 0.2$ of the sky, i.e. the sky area of ~ 9000 deg^2. The optimal sky area does not strongly depend on the parameters of the optical followup within considered range, however, the BAO detection significance does. For the actual eRASS parameters ($T = 4$ yr ~ 1500 days, $f_{\rm sky} \sim 1$), BAO should be detected at the significance levels of $\sim 14\sigma$, $\sim 8\sigma$, and $\sim 5\sigma$, for the three sets of the optical followup parameters considered in the upper, middle and bottom panels in Fig. 8, respectively. Note that these numbers are somewhat larger than the ones found in Kolodzig et al. (2013a), mostly due to the inclusion of information carried by the cross-spectra. In addition to cross-spectra the other differences compared to Kolodzig et al. (2013a) analysis were described at the beginning of section 3.

From Fig. 8 we can also see that for the next generation of X-ray telescopes like ATHENA+ or WFXT only 100

[17] http://the-athena-x-ray-observatory.eu/
[18] http://wfxt.pha.jhu.edu/
[19] http://smart-x.cfa.harvard.edu/
[20] Thus, the effective area is calculated as

$$A_{\rm eff} = \frac{\int {\rm d}E\, E^{-\Gamma} \int_{\rm FoV} {\rm d}\Omega\, A_{\rm eff}(E, \hat{\bf n})}{\int {\rm d}E\, E^{-\Gamma} \int_{\rm FoV} {\rm d}\Omega} \,,$$

where $\hat{\bf n}$ is a unit vector specifying position inside the FoV and ${\rm d}\Omega$ is a solid angle element.

days long survey covering $\sim 10 - 20\%$ of the sky could lead to $6\sigma - 13\sigma$ BAO detection, assuming the optical follow-up with quality parameters shown there. Also, $T = 100$ days survey with XMM over less than $\sim 1\%$ of the sky plus a high accuracy photometric or full spectroscopic follow-up could lead to a marginal $\sim 3\sigma$ detection, which in reality is quite comparable to most of the existing BAO measurements in the optical band, but probes the BAO at $z \sim 1$, which is not covered by any survey thus far.

It is interesting to estimate the cost of these surveys in terms of the optical follow-up, i.e. in terms of the number of objects requiring accurate redshift determination. For an optimal eROSITA survey with duration $T = 1000$ days, the sensitivity of $\sim 3 \times 10^{-15}$ erg/s/cm^2 needs to be reached over an area of ~ 9000 deg^2. At this flux limit, the AGN density is ~ 350 deg^{-2}, giving in total about $\sim 3 \times 10^6$ objects for a follow-up. This number is comparable to the total number of AGN to be detected in the planned $T = 4$ yr all-sky survey by eROSITA, but is located in a ~ 5 times smaller sky area, which will simplify the optical follow-up. A similar number of objects for the optical follow-up over similar sky area will be produced by a "BAO-optimized" survey by ATHENA+ with duration $T = 100$ days. On the other hand, a $T = 100$ days BAO survey with XMM-Newton down to the same optimal limiting flux will cover ~ 440 deg^2, and provide in total $\sim 1.5 \times 10^5$ objects for the optical follow-up. This is entirely feasible task, given the progress in the instrumentation for multi-object spectroscopy. The optical follow-up requirements for a ATHENA+ survey (about $\sim 3 \times 10^6$ objects over ~ 9000 deg^2) also do not seem to be entirely beyond the reach, being similar in the sky area and exceeding by a factor of just a few in the object density the recent BOSS survey.

Optimal survey strategy

In the following we provide in a compact form the results for the optimal survey strategy.

In Fig. 8 the the optimal surveys correspond to the points where the instrumental lines are tangent to the signal-to-noise curves. As one can see from this figure, up to a good approximation, the optimal limiting flux turns out not to depend on a particular instrument used to perform observations. It also weakly depends on the parameters of the optical follow-up, within the considered range. Once the optimal limiting flux is reached, instead of integrating any longer at the same pointing direction, one should move to the new field.

In case of follow-up parameters $\sigma_0 = 10^{-3}$ & $f_{\rm fail} = 0.0$ the optimal limiting flux is

$$F_{\rm lim}^{\rm optim} \simeq 3.2 \times 10^{-15} \, {\rm erg/s/cm}^2 \,. \quad (26)$$

Having fixed $F_{\rm lim}$ to the above value, the survey area can now be expressed as

$$A[{\rm deg}^2] \simeq 220 \left[\frac{T}{1000 \, {\rm days}} \right] \left[\frac{\rm FoV}{0.25 \, {\rm deg}^2} \right] \left[\frac{A_{\rm eff}}{1000 \, {\rm cm}^2} \right], \quad (27)$$

and the CL for the BAO detection as

$$\mathrm{CL} \simeq 5.4 \sqrt{\frac{A}{1000 \, {\rm deg}^2}} \quad {\rm sigma}. \quad (28)$$

The extragalactic source density at $F_{\rm lim} \simeq 3.2 \times 10^{-15}$ is $\simeq 350$ deg^{-2}, i.e., the number of sources above this limiting flux can be estimated as (here we assume soft-band number counts taken from Kim et al. (2007)).

$$N \simeq 350 \times A[{\rm deg}^2] \,. \quad (29)$$

For the other two sets of follow-up parameters discussed in this paper, i.e. $\sigma_0 = 10^{-2}$ & $f_{\rm fail} = 0.0$ and $\sigma_0 = 10^{-2}$ & $f_{\rm fail} = 0.3$, the optimal limiting fluxes are $F_{\rm lim} \simeq 2.9 \times 10^{-15}$ and 2.4×10^{-15} erg/s/cm^2, respectively. In the first case this leads to the replacements $220 \to 200$, $5.4 \to 3.5$, and $350 \to 380$ in Eqs. (27), (28), and (29), respectively. In case of $\sigma_0 = 10^{-2}$ & $f_{\rm fail} = 0.3$ the corresponding replacements are: $220 \to 170$, $5.4 \to 2.8$, $350 \to 440$.

5. Summary and conclusions

We have investigated the potential of large samples of X-ray-selected AGN to probe the large-scale structure of the Universe, in particular, to detect the BAO. For the X-ray-selected AGN, most of the BAO signal comes from redshifts $z \sim 1$, where the X-ray AGN population peaks. These redshifts are currently uncovered by any of the existing dedicated BAO surveys. However, although X-ray surveys are very efficient in producing large samples of AGN, they do not provide any redshift information, and so have to be complemented with an optical follow-up.

The main goals of this study were (i) to find out necessary quality criteria for the optical follow-up and (ii) to formulate the optimal strategy of the X-ray survey in order to facilitate accurate measurements of the clustering two-point function and detection of the BAO.

Our main results are presented in Figs. 7 and 8 where the confidence levels for the BAO detection are shown as a function of X-ray survey parameters $(F_{\rm lim}, f_{\rm sky})$ and the parameters describing the quality of the optical follow-up $(\sigma_0, f_{\rm fail})$. In particular we demonstrated that redshift accuracy of $\sigma_0 = 10^{-2}$ at $z = 1$ and the catastrophic failure rate of $f_{\rm fail} \lesssim 30\%$ are sufficient for a reliable detection of BAO in the future X-ray surveys. If spectroscopic quality redshifts ($\sigma_0 = 10^{-3}$ and $f_{\rm fail} \sim 0$) are available, the confidence level of the BAO detection will be boosted by a factor of ~ 2.

For the meaningful detection of BAO, X-ray surveys of moderate depth of $F_{\rm lim} \sim$ few 10^{-15} erg/s/cm^2 covering sky area from a $\sim$few hundred to a $\sim$ten thousand square degrees are required. For the fixed survey duration, the optimal strategy for the BAO detection does not necessarily require full sky coverage. For example, in a $T = 1000$ days survey by an eROSITA type telescope, an optimal strategy for BAO detection requires a survey of ~ 9000 deg^2 and would yield a $\sim 16\sigma$ BAO detection. A similar detection will be achieved by ATHENA+ or WFXT type telescopes in a survey with a duration of 100 days, covering similar sky area. XMM-Newton can achieve a marginal BAO detection in a 100 days survey covering ~ 400 deg^2.

These surveys would impose moderate to high demands on the optical follow-ups requiring determination of redshifts of $\sim 10^5$ objects (XMM-Newton) to $\sim 3 \times 10^6$ objects (eROSITA, ATHENA+ and WFXT) in the above mentioned sky areas. Given the progress in the instrumentation for multi-object spectroscopy, these demands appear to be within the reach of the modern and future ground based optical facilities.
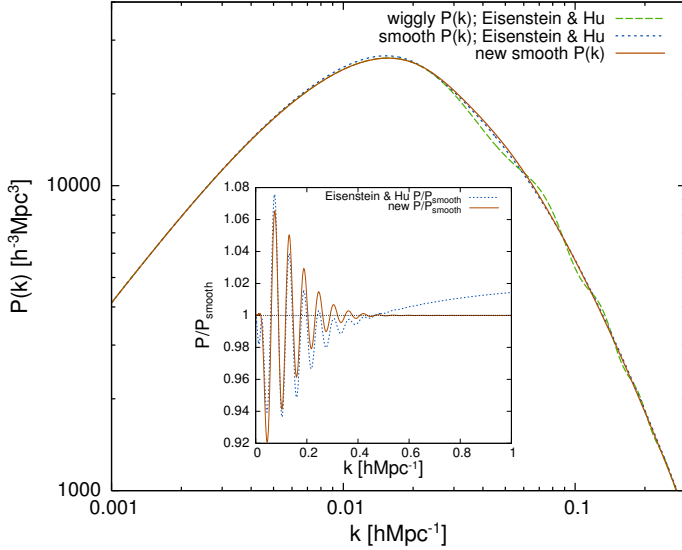
Fig. A.1. The smooth spectral template used in this paper. The blue dashed line shows a similar smooth template proposed by Eisenstein & Hu (1998)

Since the BAO is $\sim 5\%$ modulation on top of a smooth broad-band spectral shape the amplitude of the power spectrum, and thus the AGN clustering bias, can be determined with an order of magnitude larger signal-to-noise, and so can be done with much poorer quality photometric redshifts.

Appendix A: 'No wiggle' spectral template

In order to isolate the BAO signal, we do not use the usual 'no wiggle' spectral form of Eisenstein & Hu (1998). Instead, we divide the power spectrum with a smooth model that is calculated essentially the same way as the 'wiggly' case, with only the following replacement in Eq. (21) of Eisenstein & Hu (1998)

$$ j_0(x) = \frac{\sin(x)}{x} \longrightarrow \exp\left(-\frac{x^2}{6}\right). \qquad (A.1) $$

With this replacement the peak region of $j_0(x)$ at small x is well matched (up to $\mathcal{O}(x^4)$) and for larger arguments the exponent efficiently smoothes out the oscillatory features.

The resulting smooth spectrum and extracted BAO in comparison to the original Eisenstein & Hu (1998) form are shown in Fig. A.1. Since in this study we apply an effective wavenumber cutoff of $k_{max} \sim 0.1 \ h\,\mathrm{Mpc}^{-1}$, both smooth spectral forms lead to almost indistinguishable results.

References

Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539
Allevato, V., Finoguenov, A., Cappelluti, N., et al. 2011, ApJ, 736, 99
Anderson, L., Aubourg, E., Bailey, S., et al. 2012, MNRAS, 427, 3435
Asorey, J., Crocce, M., Gaztañaga, E., & Lewis, A. 2012, MNRAS, 427, 1891
Blake, C., Kazin, E. A., Beutler, F., et al. 2011, MNRAS, 418, 1707
Brandt, W. N. & Hasinger, G. 2005, ARA&A, 43, 827
Brusa, M., Comastri, A., Gilli, R., et al. 2009, ApJ, 693, 8
Budavári, T., Connolly, A. J., Szalay, A. S., et al. 2003, ApJ, 595, 59
Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96
Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, MNRAS, 362, 505
de Bernardis, P., Ade, P. A. R., Bock, J. J., et al. 2000, Nature, 404, 955
Eisenstein, D. J. & Hu, W. 1998, ApJ, 496, 605
Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560
Feldman, H. A., Kaiser, N., & Peacock, J. A. 1994, ApJ, 426, 23
Fisher, K. B., Scharf, C. A., & Lahav, O. 1994, MNRAS, 266, 219
Hanany, S., Ade, P., Balbi, A., et al. 2000, ApJ, 545, L5
Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, ApJS, 172, 29
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Huterer, D., Knox, L., & Nichol, R. C. 2001, ApJ, 555, 547
Kim, M., Wilkes, B. J., Kim, D.-W., et al. 2007, ApJ, 659, 29
Kolodzig, A., Gilfanov, M., Hütsi, G., & Sunyaev, R. 2013a, arXiv:1305.0819
Kolodzig, A., Gilfanov, M., Sunyaev, R., Sazonov, S., & Brusa, M. 2013b, A&A, 558, A89
Krumpe, M., Miyaji, T., Coil, A. L., & Aceves, H. 2012, ApJ, 746, 1
Matsubara, T. 2008, Phys. Rev. D, 78, 083519
Meiksin, A. & White, M. 1999, MNRAS, 308, 1179
Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv:1209.3114
Mountrichas, G., Georgakakis, A., Finoguenov, A., et al. 2013, MNRAS, 430, 661
Padmanabhan, N., Schlegel, D. J., Seljak, U., et al. 2007, MNRAS, 378, 852
Peebles, P. J. E. & Yu, J. T. 1970, ApJ, 162, 815
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013a, arXiv:1303.5076
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013b, arXiv:1303.5084
Predehl, P., Andritschke, R., Böhringer, H., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7732, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, ApJ, 742, 61
Scoccimarro, R., Zaldarriaga, M., & Hui, L. 1999, ApJ, 527, 1
Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1
Slosar, A., Iršič, V., Kirkby, D., et al. 2013, J. Cosmology Astropart. Phys., 4, 26
Smith, R. E. 2009, MNRAS, 400, 851
Sunyaev, R. A. & Zeldovich, Y. B. 1970, Ap&SS, 7, 3
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, Phys. Rep., 530, 87
Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, ApJS, 195, 10