

# The number counts of the weakest *Chandra* sources

A. M. Soltan

Nicolaus Copernicus Astronomical Center, Bartycka 18, 00-716 Warsaw, Poland  
e-mail: [soltan@camk.edu.pl](mailto:soltan@camk.edu.pl)

Received 14 February 2011 / Accepted 10 May 2011

## ABSTRACT

**Context.** Most of the X-ray background (XRB) is generated by discrete X-ray sources. It is likely that the population of weak sources below the present detection threshold contributes to the remaining fraction of the XRB. The nature of these weak sources remains a matter of discussion. It is assumed that the unresolved XRB consists of a truly diffuse component and a population of such weak sources. Albeit these sources are not observed directly, their collective nature can be investigated by statistical means.

**Aims.** The goals are: a) to explore the effectiveness of the nearest neighbor statistics (NNST) of the photon distribution in investigating the number counts of the very weak sources, b) to estimate the source counts below the conventional detection limit in the *Chandra* Deep Field-South 2Ms exposure.

**Methods.** All the sources generating at least two counts each induce a nonrandom distribution of counts. This distribution is analyzed by means of the NNST. Using the basic probability equations, the relationships between the source number counts  $N(S)$  and the NNST are derived. The method is tested on the medium deep *Chandra* pointing to assess the source counts  $N(S)$  at flux levels attainable only with the very deep exposures.

**Results.** It is shown that the method yields constraints on the  $N(S)$  relationship below the regular discrete source detection threshold. The *Chandra* 2Ms exposure was used to assess the source counts down to  $3\text{--}4 \times 10^{-18}$  cgs in the soft band (0.5–2 keV) and down to  $2\text{--}3 \times 10^{-17}$  cgs in the hard band (2–8 keV). In the soft band, the source counts appear to steepen substantially below  $\sim 10^{-16}$  cgs. Assuming that the differential slope  $b \approx 1.5\text{--}1.6$  in the range  $10^{-16}\text{--}10^{-14}$  cgs, the number of weaker sources indicates that the slope  $\approx -2.0$ . The steepening is not observed in the hard band.

**Conclusions.** The steepening of counts in the soft band indicates that we detect a new population of sources. A class of normal galaxies at moderate redshifts is a natural candidate.

**Key words.** X-rays: diffuse background – X-rays: general

## 1. Introduction

The X-ray background (XRB) is mostly generated by discrete extragalactic sources (e.g. Luo et al. 2010; Brandt & Hasinger 2005, and references therein). However, many points related to this question are still disputed and await clarification. In particular, we are unable to determine the fractional contribution of various types of sources to the integral XRB. It was established using early imaging X-ray telescopes that quasars play an important role (Tananbaum et al. 1979; Zamorani et al. 1981). However, the exact relationship between various classes of X-ray sources and the XRB requires further study (e.g. Bauer et al. 2004; Worsley et al. 2005; Cappelluti et al. 2009). Identifications of weak sources populating the deep *Chandra* observations show that at low flux levels starburst and normal galaxies rather than AGN become the dominant contributors to the XRB (e.g. Lehmer et al. 2008; Luo et al. 2008). The profound difference between the typical spectrum of an unobscured AGN and the XRB implies that there is a significant contribution from the Compton-thick AGN, which are deficient in soft X-rays (e.g. Gilli et al. 2007; Treister et al. 2009). The diffuse soft X-rays originating in the Local Bubble and other regions of the Milky Way (Galeazzi et al. 2007; Henley et al. 2007) also complicate the investigation of the extragalactic component. In addition, it has not been established what fraction of the XRB arises from a truly diffuse component generated in the intergalactic space via either Thomson scattering (Soltan 2003) or thermal emission by the WHIM (Hickox & Markevitch 2007a; Soltan 2007).

One mean of elucidating some basic questions on the XRB is to analyze the X-ray source counts at all the accessible fluxes.

The individual point-like source is detected if a number of counts within a specified area exceeds the assumed threshold. The size of the detection box is defined by the point spread function (PSF), while the detection threshold is usually selected to minimize the number of false detections and at the same time maximize the number of real sources. The detection threshold is typically set at the level of  $4\text{--}5\sigma$  above the local average count density. The presence of weaker sources, below the formal detection threshold, is manifested by the larger fluctuations in the count distribution compared to the fluctuations expected for the random counts.

A common approach to assessing the counts of sources weaker than the detection limit is based on the count density fluctuation analysis. To quantify the signal generated by the discrete sources, one should determine the intensity distribution  $P(D)$ , i.e. the histogram of the number of pixels as a function of the number of counts. The observed function  $P(D)$  is then compared with the functions obtained from the simulated count distributions (e.g. Hasinger et al. 1993; Miyaji & Griffiths 2002). It is assumed that the simulated source counts represent the actual source distribution if the model  $P(D)$  function mimics the observed histogram. Another procedure for estimating the contribution of point sources to the observed count distribution is based on the auto-correlation function (ACF). Since the integral of the ACF is directly related to the second moment of the  $P(D)$  distribution, (e.g. Soltan 1991) both methods are closely related.

An innovative method for assessing the number of weak sources was proposed by Georgakakis et al. (2008, hereafter GNL). In their approach, the count distribution in the detection cell is explicitly expressed as a sum of the source and background counts. As a result of a rigorous application of the Poisson statistics, a flux probability distribution is derived as a function of the total and background counts observed in the detection cell. This probability distribution combined with the adequately defined sensitivity map of a given observation is then used to estimate the source number counts.

The count fluctuations are proportional to the source intensities. Thus, the observed fluctuation amplitude is dominated by the sources just below the detection threshold set for the individual objects, whereas it is only weakly sensitive to the fainter sources that produce a smaller number of counts. One should note, however, that every source that produces more than one count generates a deviation from the random count distribution. In the present paper, we investigate this feature using nearest neighbor statistics (NNST). We first examine the efficiency of the NNST for the weak source analysis and show that the NNST is a powerful tool for estimating the source counts,  $N(S)$ , down to very low flux levels. We apply this technique to one of the *Chandra* AEGIS<sup>1</sup> fields with the exposure of 465 ks. The NNST allows us to obtain the  $N(S)$  relationship extending down to  $2 \times 10^{-17}$  erg cm<sup>-2</sup> s<sup>-1</sup> and  $7 \times 10^{-17}$  erg cm<sup>-2</sup> s<sup>-1</sup> in the 0.5–2 keV and 2–8 keV energy bands, respectively, i.e. a factor of 5–10 below the standard sensitivity threshold corresponding to this exposure. Since the present count estimates are contained within the flux range covered by the direct source counts derived from the deepest *Chandra* fields, the effectiveness of the NNST method can be directly assessed. The source number counts are then investigated in one of the deep *Chandra* pointing, viz. the CDFS. The NNST enables us to determine the  $N(S)$  relationship for sources that generate merely  $\sim 2$  counts, i.e. roughly an order of magnitude below a standard threshold for a detection of discrete point-like sources. The application of the NNST to the CDFS shifts the sensitivity limit below  $4 \times 10^{-18}$  cgs, in the 0.5–2 keV band, which is a factor of  $\sim 3$  below the statistical estimates for the presently deepest number counts given in GNL.

The organization of the paper is as follows. In the next section, the method and all the relevant formulae are presented. Two sets of data: a) AEGIS field, and b) CDFS are then processed. In Sect. 3, the observational material is described and the computational details including questions related to the PSF are given. Results of the calculations, i.e. estimates of the source counts below the nominal sensitivity limit are presented in Sect. 4. The discussion on the results, including comparison with those of other authors is given in Sect. 5.

## 2. The nearest neighbor statistics

We derive a general formulae based on the theory of probability. All the details related to the actual observations (e.g. relationship between counts and photon energy, instrument sensitivity) are discussed in Sect. 3.

In the present study, we assume that counts in the given *Chandra* ACIS<sup>2</sup> observation are distributed according to the following model. From the point of view of spatial characteristics, the counts collected in the focal plane of the X-ray telescope are arranged into two populations. Some a priori unknown fraction

of counts is randomly distributed, i.e. the positions of the counts are fully described by the Poisson statistics, while all the remaining photons are clustered in the point-like sources. This model corresponds to the observation obtained using the idealized telescope without vignetting and a detector with a perfectly uniform response over the field of view. The real telescope-detector combination introduces numerous deformations to this ideal picture. We address this question below.

We also assume that the positions of sources are randomly distributed. Smoothly distributed counts constitute a physically heterogeneous collection that contains both the particle background and various components of the foreground X-ray emission including scattered solar X-rays, the geocoronal oxygen lines as well as the thermal emission of hot plasma within our Galaxy (e.g. Galeazzi et al. 2007; Henley et al. 2007), and the emission by the WHIM (e.g. Sołtan 2007). Truly diffuse extragalactic background (Sołtan 2003) and weak discrete sources, each producing in the final image exactly one photon, also contribute to these counts. A separation of the extragalactic component from all the counts can be achieved in statistical terms using spectral information, however, individual events cannot be definitely classified as either local or extragalactic.

Discrete sources in the deep X-ray exposures are predominantly extragalactic. Nevertheless, galactic sources are potentially present in the data and are included in the calculations. Photons coming from a discrete sources producing two or more counts are distributed in the image in clumps defined by the PSF of the telescope.

One should expect the distribution of the nearest neighbors for both classes of counts to be different. The counts produced by sources have on average closer neighbors than the non-clustered counts. This feature is conveniently formulated using the NNST. Let  $n_t$  denote the total number of counts in the investigated field,  $n_1$  the number of events distributed randomly, or “single photons”,  $n_2$  the number of photons due to sources producing each exactly two photons,  $n_3$  number of photons due to “three photon” sources and so on. Thus,

$$n_1 + n_2 + \dots + n_k + \dots + n_{k_{\max}} = n_t, \quad (1)$$

where the left-hand-side sum extends over all the sources and  $k_{\max}$  is the number of photons produced by the brightest source in the field. The number of photons  $n_k$  is related to the number of sources in an obvious way of  $k N(k) = n_k$ , where  $N(k)$  denotes the number of “ $k$ -photon sources”.

Using the basic relationships of the probability theory, one can calculate  $P(r)$ , the probability that the distance to the nearest neighbor of a randomly chosen event exceeds  $r$

$$p_1 P(r|1) + p_2 P(r|2) + \dots + p_{k_{\max}} P(r|k_{\max}) = P(r), \quad (2)$$

where  $p_k$  denotes the probability that the randomly chosen photon is produced by the “ $k$ -photon source” ( $k = 1, 2, \dots, k_{\max}$ ), and  $P(r|k)$  is the conditional probability that there are no other counts within  $r$  provided the selected event belongs to the  $k$ -photon source.

The probability  $P(r)$  can be estimated for the given distribution of counts by measuring the distance to the nearest neighbor for each photon in the field. Similarly,  $P(r|1)$  is given by the distribution of distances to the nearest photon from the randomly distributed points. Assuming that the distribution of “single counts” is not correlated with the distribution of photons from  $k \geq 2$  sources, and that sources are distributed randomly, the probability  $P(r|k)$  for  $k \geq 2$  is related to  $P(r|1)$  and to the PSF by the expression

$$P(r|k) = P(r|1) \cdot \mathcal{P}(r|k), \quad (3)$$

<sup>1</sup> All-wavelength Extended Groth strip International Survey, see <http://aegis.ucolick.org/index.html>

<sup>2</sup> In all the analysis we use ACIS-I chips 0–3.

where  $\mathcal{P}(r|k)$  is the probability that the distance from the randomly chosen photon produced by the  $k$ -photon source to its nearest neighbor from the same source exceeds  $r$ . This quantity is fully defined by the PSF.

To estimate the probabilities  $p_k$ , we now use a ratio  $n_k/n_t$  and Eq. (2) then takes the form

$$\frac{n_1}{n_t} P(r|1) + \sum_{k=2}^{k_{\max}} \frac{n_k}{n_t} P(r|1) \mathcal{P}(r|k) = P(r). \quad (4)$$

We note that Eq. (4) is linear in the photon counts  $n_k$ , what makes it particularly suitable for estimating the contribution of weak sources to the total counts. Both the second moment of the count distribution in pixels and the autocorrelation function depend on squares of the photon counts. Substituting successive values of  $r$  into Eq. (4), a set of linear equations is constructed that allows us to estimate the unknown counts  $n_k$ .

In the deep *Chandra* observation, a large number of relatively strong sources is detected and the range of source fluxes is much too wide to apply Eq. (4) in the form given above where the source fluxes are listed consecutively from  $k = 2$  up to some maximum value of  $k_{\max}$  representing the strongest source in the field. Since we are interested in the counts at the faint end (down to  $k = 2$ ), the value of  $k_{\max}$  should be selected at the conventional detection limit, e.g. by applying the  $5\sigma$  threshold. All the sources above this threshold are pinpointed and removed from the data. The subsequent analysis is concentrated on the sources that cannot be individually recognized.

One can express the number of photons originating in  $k$ -photon sources by means of the differential source counts  $N(S)$

$$n_k = k \int_{S_{\min}}^{S_{\max}} dS N(S) \mathfrak{P}(k|S), \quad (5)$$

where  $\mathfrak{P}(k|S)$  is the probability that the source generating the flux  $S$  delivers  $k$ -photons, while the integration limits  $S_{\min}$  and  $S_{\max}$  define the full range of the source fluxes. It is convenient to introduce the instrumental count as a flux unit. The flux  $s$  expressed in the ACIS counts in the definite observation is related to the flux in physical units,  $S$ , by

$$s = S/\eta, \quad (6)$$

where  $\eta$  is the conversion factor that has units of “erg cm<sup>-2</sup> s<sup>-1</sup>/count”. The conversion factor is related to the parameter “exposure map”, EM, defined in a standard processing of the ACIS data<sup>3</sup> by the relationship EM =  $\eta \cdot \langle E \rangle$ , where  $\langle E \rangle$  denotes the average photon energy.

For the power law source counts,  $N(s) = N_0 s^{-b}$ , we have

$$n_k = N_0 \frac{\Gamma(k-b+1, s_{\min}) - \Gamma(k-b+1, s_{\max})}{\Gamma(k)}. \quad (7)$$

If the slope of the counts  $b$  is constant over a sufficiently wide range of fluxes (i.e.  $s_{\min} \ll 2$  and  $s_{\max} > k_{\max}$ ), one might replace the integration limits  $s_{\min}$  and  $s_{\max}$  by 0 and  $\infty$ , respectively, to get

$$n_k = N_0 \frac{\Gamma(k-b+1)}{\Gamma(k)}. \quad (8)$$

<sup>3</sup> See <http://asc.harvard.edu/ciao>. For the real observations, both the  $\eta$  and exposure maps are functions of the position. At this stage, these parameters are assumed to be constant.

Thus, for the source counts represented by a single power law over a sufficiently wide range of fluxes, Eq. (4) takes the form

$$\frac{n_1}{n_t} P(r|1) + \frac{N_0}{n_t} \sum_{k=2}^{k_{\max}} \frac{\Gamma(k-b+1)}{\Gamma(k)} P(r|1) \mathcal{P}(r|k) = P(r). \quad (9)$$

For the different parametrization of the source counts, Eq. (9) can be modified. In particular, for the broken power law

$$N(s) = N_0 \left( \frac{s}{s_0} \right)^{-b}, \quad b = \begin{cases} b_1 & \text{for } s \geq s_0 \\ b_2 & \text{for } s < s_0 \end{cases} \quad (10)$$

the  $\Gamma$  function in the numerator is replaced by the appropriate combination of the incomplete  $\Gamma$  functions. Substituting

$$n_1 = n_t - \sum_{k=2}^{k_{\max}} n_k \quad (11)$$

and using Eq. (8), we finally get

$$\frac{N_0}{n_t} P(r|1) \sum_{k=2}^{k_{\max}} \frac{\Gamma(k-b+1)}{\Gamma(k)} [1 - \mathcal{P}(r|k)] = P(r|1) - P(r). \quad (12)$$

Equation (12) contains two parameters that fully describe the power-law source counts in the investigated flux range, namely the normalization  $N_0$  and the slope  $b$ . Since the normalization at the flux  $s = k_{\max}$  is defined by the actual source counts above this threshold, only the slope remains unknown.

### 3. Observational material

The present method has been applied to two sets of data. The NNST determined for the AEGIS medium depth *Chandra* pointing provided an opportunity to assess the efficiency of the method, while the CDFS 2Ms allowed us to study the  $N(S)$  relationships at previously unattainable low flux levels. I selected 16 close pointings in AEGIS which span a period of six months and the data were processed in a uniform way with the recent pipeline processing versions. The details of 16 observations used in the present paper are given in Table 1. All the exposures were examined for the presence of the background flares and only “good time intervals” (GTI) were used in the subsequent analysis. The data were divided into two energy bands:  $S$  i.e. soft (0.5–2 keV) and  $H$  i.e. hard (2–8 keV).

The CDFS was observed with the ACIS detector in several “sessions”. For the present investigation, a set of 20 observations were used. Although the observations span a period of more than seven years, the data used here were also processed in a relatively uniform way. The selection of the GTI and all the steps in the data handling were performed in a similar way as in the case of the AEGIS. The basic information on the CDFS data is given in Table 2.

#### 3.1. The exposure map

The observations listed in Table 1 were merged to create a single count distribution and a single exposure map for the AEGIS. The same procedure were applied to 20 observations of CDFS placed in Table 2. In the case of AEGIS, a circular area covered by all the pointings with a relatively uniform exposure centered at RA = 14<sup>h</sup>20<sup>m</sup>12<sup>s</sup>, Dec = 53°00′ and of radius 6′0 was selected for further processing. Similarly, for the CDFS pointing, the area

**Table 1.** The *Chandra* AEGIS observations used in the paper.

Obs. ID	Observation and processing dates		Processing version	Exposure time [s]
9450	2007-12-11	2007-12-13	7.6.11.3	29 100
9451	2007-12-16	2008-01-02	7.6.11.4	25 350
9793	2007-12-19	2007-12-21	7.6.11.4	44 750
9725	2008-03-31	2008-04-02	7.6.11.6	28 050
9842	2008-04-02	2008-04-03	7.6.11.6	19 450
9844	2008-04-05	2008-04-06	7.6.11.6	34 600
9866	2008-06-03	2008-06-05	7.6.11.6	31 450
9726	2008-06-05	2008-06-06	7.6.11.6	39 750
9863	2008-06-07	2008-06-07	7.6.11.6	24 100
9873	2008-06-11	2008-06-12	7.6.11.6	30 850
9722	2008-06-13	2008-06-15	7.6.11.6	19 900
9453	2008-06-15	2008-06-17	7.6.11.6	22 150
9720	2008-06-17	2008-06-18	7.6.11.6	26 000
9723	2008-06-18	2008-06-20	7.6.11.6	30 950
9876	2008-06-22	2008-06-23	7.6.11.6	25 050
9875	2008-06-23	2008-06-25	7.6.11.6	33 150
Total exposure				464 650

**Table 2.** Log of the CDFS observations used in the paper.

Obs. ID	Observation and processing dates		Processing version	Exposure time [s]
441	2000-05-27	2007-05-23	7.6.10	56 600
582	2000-06-03	2007-05-23	7.6.10	132 150
2405	2000-12-11	2007-06-20	7.6.10	57 100
2312	2000-12-13	2007-06-20	7.6.10	125 100
1672	2000-12-16	2007-06-20	7.6.10	96 150
2409	2000-12-19	2007-06-21	7.6.10	69 850
2313	2000-12-21	2007-06-22	7.6.10	131 900
2239	2000-12-23	2007-06-22	7.6.10	132 600
8591	2007-09-20	2007-09-21	7.6.11.1	45 800
9593	2007-09-22	2007-09-27	7.6.11.1	46 100
9718	2007-10-03	2007-10-06	7.6.11.1	49 850
8593	2007-10-06	2007-10-08	7.6.11.1	49 250
8597	2007-10-17	2007-10-24	7.6.11.2	59 650
8595	2007-10-19	2007-11-07	7.6.11.2	116 850
8592	2007-10-22	2007-11-01	7.6.11.2	87 750
8596	2007-10-24	2007-11-14	7.6.11.2	116 600
9575	2007-10-27	2007-11-14	7.6.11.2	110 150
9578	2007-10-30	2007-11-30	7.6.11.2	39 000
8594	2007-11-01	2007-11-14	7.6.11.2	143 300
9596	2007-11-04	2007-11-20	7.6.11.2	116 600
Total exposure				1 782 350

centered at RA = 3<sup>h</sup>32<sup>m</sup>, Dec = −27 deg 49′ with radius of 5′0 in the *S* band and 4′0 in the *H* band was chosen.

The exposure map of the individual observation resulting from various instrumental characteristics<sup>4</sup>, has a complex structure. In effect, the exposure map of the merged observation is even less regular and is devoid of any clear symmetries. On the other hand, a rough texture of the individual exposure is reduced and smoothed in a sum of 16 and 20 components. To reduce the variations in the exposure map over the investigated area, a threshold of the minimum exposure was set separately for both energy band. Pixels below this threshold were not used in the calculations.

For the AEGIS data, a threshold was defined to be ~75% of the maximum value of the exposure map for both energy bands.

<sup>4</sup> To name the most obvious: vignetting, gaps between chips, telescope wobbling, and all kinds of chip imperfections.

**Table 3.** Energy bands and conversion factors.

Energy band [keV]	Conversion factors <sup>†</sup>				
	Average	rms	minimum	maximum	
AEGIS					
<i>S</i>	0.5–2	14.69	0.87	12.79	16.87
<i>H</i>	2–8	55.46	3.46	48.62	65.11
CDFS					
<i>S</i>	0.5–2	3.701	0.205	3.232	4.218
<i>H</i>	2–8	14.30	0.85	12.69	16.28

**Notes.** <sup>†</sup> The conversion factor ( $\eta$ ) has units of  $10^{-18}$  (erg cm<sup>-2</sup> s<sup>-1</sup>)/count.

In effect, the maximum deviations of the exposure from the average value do not exceed 15% and 18% in the bands *S* and *H*, respectively, and the corresponding rms of the exposures amount to 5.9% and 6.2%. For CDFS the thresholds were defined at 77% of the maximum value of the exposure map in the *S* band and 78% in the *H* band. Consequently, in both energy bands the maximum deviations of the exposure from the average value do not exceed 15% and the exposure rms over the investigated areas fall below 6%. In Table 3, the conversion factors calculated using the relevant amplitudes of the exposure maps are given for both fields. In the calculations “from counts to flux”, a power spectrum with a photon index  $\Gamma_{\text{ph}} = 1.4$  was assumed (Kim et al. 2007).

Variations in the conversion factor,  $\eta$ , over the investigated area alter the source fluxes via Eq. (6) and, consequently, the source counts  $N(S)$ . For the power law counts, the  $\eta$  uncertainty affects the counts normalization and does not change the slope. It is shown in the Appendix that as long as the  $\eta$  variations remain small, they modify the probability distributions  $P(r|1)$  and  $P(r)$  in such a way that the solution of Eq. (12) is unaffected.

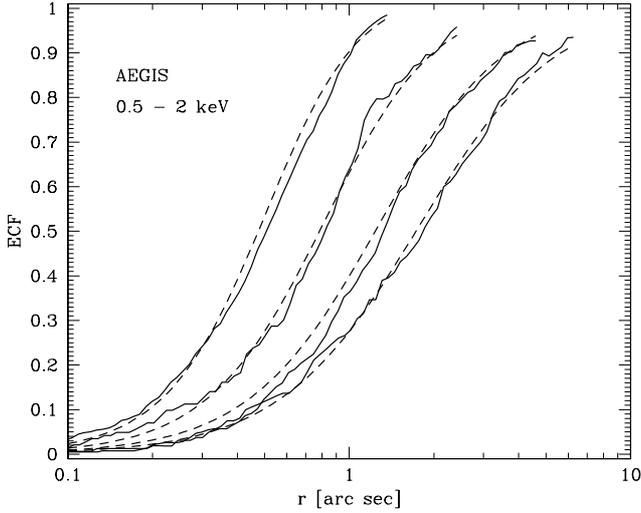
### 3.2. The point spread function

The  $\mathcal{P}(r|k)$  probability was calculated using a Monte Carlo method and a model PSF. The *Chandra* X-ray telescope PSF is a complex function of source position and energy (e.g. Allen et al. 2004). However, to compute the nearest neighbor probability distribution for counts generated by a point-like source,  $\mathcal{P}(r|k)$ , we do not need a full model of the PSF shape. The aim is to find a convenient analytic PSF approximation that adequately reproduces the NNST over the investigated area for each energy band. The model should be applicable to the merged observations processed in a standard way.

We note that the analytic form of the PSF should mimic the radial distribution of counts, while some deviations from the azimuthal symmetry are of lesser importance. A tractable method for obtaining the  $\mathcal{P}(r|k)$  is to find an analytic approximation for the encircled count fraction (ECF) for a point source as a function of the distance from the field center. Several simple analytic models have been tested to fit the observed distribution of counts and the ECF has been found to be adequately represented by the function of the form

$$f(<r) = \frac{r^\alpha}{z + r^\alpha + y \cdot r^{\alpha/2}}, \quad (13)$$

where  $\alpha$ ,  $z$ , and  $y$  are parameters depending on the source position and energy. The shape of the PSF depends strongly on the distance from the optical axis of the telescope. For a single observation, the optical axis is shifted from the geometrical center



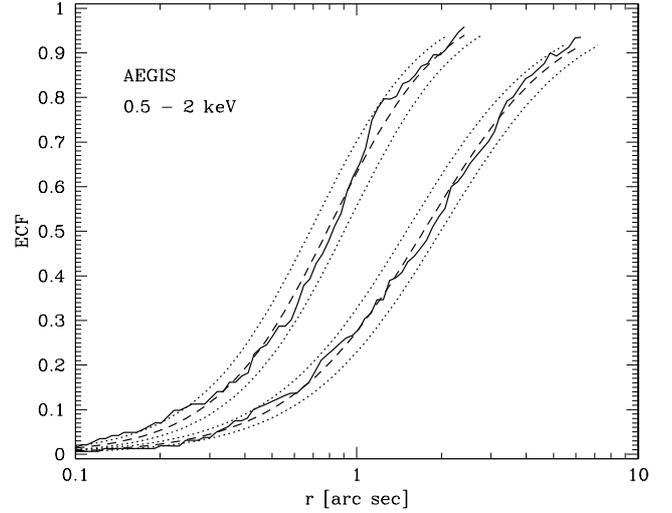
**Fig. 1.** Encircled count fraction (ECF) as a function of distance from the count centroid. Example distributions are shown for four sources in the AEGIS field at 1'0, 3'2, 5'0, and 6'1 from the field center. Solid curves – observed count distributions, dashed curves – fits obtained using Eqs. (13) and (14).

of the ACIS chips 0–3. However, in the merged data of several pointings the axis that is appropriate for the PSF modeling is not well defined, and it was assumed that the variations in the PSF shape are symmetrical with respect to the field of view (fov) center. By fitting the  $\alpha$ ,  $z$ , and  $y$  to the ECF distributions of a number of sources scattered over the entire fov, it was found that variations in these parameters can be conveniently parametrized by the distance from the field center  $\theta$ . The whole procedure looks as follows. We noticed that the PSF parameters  $\alpha$ ,  $y$ , and  $\log z$  are satisfactorily approximated by linear functions of the off-axis angle  $\theta$ :

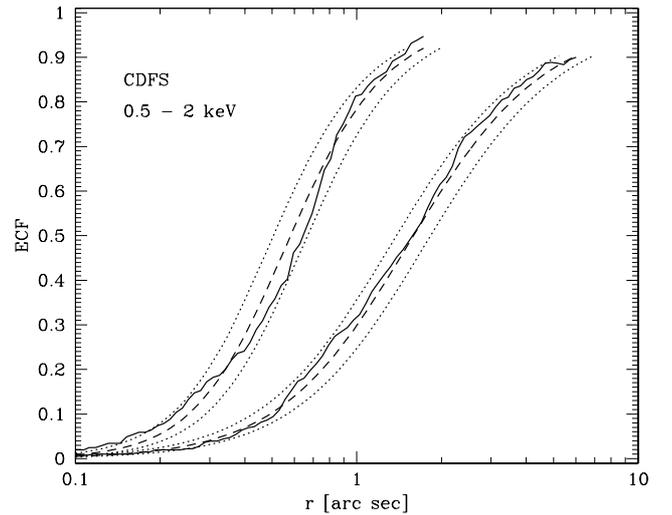
$$\alpha = a_\alpha \cdot \theta + b_\alpha \quad y = a_y \cdot \theta + b_y \quad \log z = a_z \cdot \theta + b_z, \quad (14)$$

where  $a_s$  and  $b_s$  ( $s = \alpha, y, z$ ) are six parameters that are substituted to Eq. (13) and simultaneously fitted to the observed distribution of counts in the several dozen strongest point-like sources. In Fig. 1, a sample of the resultant fits to the observed distributions in the *S* band in the AEGIS field is shown. Since not all the fits are of equal quality, the effects of the  $\mathcal{P}(r|k)$  approximation are carefully examined. Although the fitting procedure provides sensible and functional representation of the PSF over the fov, it is difficult to assess the impact of the potential systematic errors generated by the present approximation on our final results. We find that the deviations between the model and the observed ECF in most cases do not exceed 15% of the distance from the count centroid (see Fig. 1). Thus, an adequate envelope of errors generated by the imperfections of our fitting procedures is given by two model  $\mathcal{P}(r|k)$  distributions based on the ECF functions being systematically wider and narrower by 15% than the best fit. Example results of this procedure in the AEGIS field are illustrated in Fig. 2, where the envelope ECFs for two sources at off axis angles of 3'2 and 6'1 are shown. Although some deviations are quite large, the observed ECFs in fact are predominantly contained within the  $\pm 15\%$  distribution for a wide range of the separations  $r$ .

In the observational material used in the present investigation, the maximum separations found in the NNST very rarely exceed 2''. Thus, our fits appear adequate for the NNST and it is assumed that the  $\pm 15\%$  limits (indicated by the dotted curves



**Fig. 2.** ECF as in Fig. 1 for sources at 3'2 and 6'1 from the field center. Solid curves – observed count distributions, dashed curves – best fits obtained using Eqs. (13) and (14), dotted curves – the ECF distributions with the radius  $r$  scaled by  $\pm 15\%$ .



**Fig. 3.** Same as Fig. 2 for CDFS. Example distributions are shown for two sources at 2'1 and 5'9 from the field center. Solid curves – the observed count distributions, dashed curves – fits obtained using Eqs. (13) and (14), dotted curves – the ECF distributions differing from the best fit by  $\pm 15\%$ .

in Fig. 2) define the maximum systematic errors associated with the PSF fitting procedure, which are used to assess uncertainties in our faint source calculations.

In the Monte Carlo computations of  $\mathcal{P}(r|k)$ , a population of  $10^8$  “sources” of  $k = 2, 3, \dots, k_{\max}$  counts were distributed randomly over the investigated area<sup>5</sup>. The distribution of counts within each source was randomized according to the model ECF. For each source, a distribution of the nearest neighbor separations was then determined and used to obtain the corresponding amplitudes of  $\mathcal{P}(r|k)$ . The procedure was executed for the best fit and  $\pm 15\%$  ECF distributions.

The same procedures were applied to the CDFS data. In Figs. 3 and 4, some examples of the ECF fits are shown for

<sup>5</sup> The value of  $k_{\max}$  is related to the sensitivity threshold and is different for each energy band; see below.

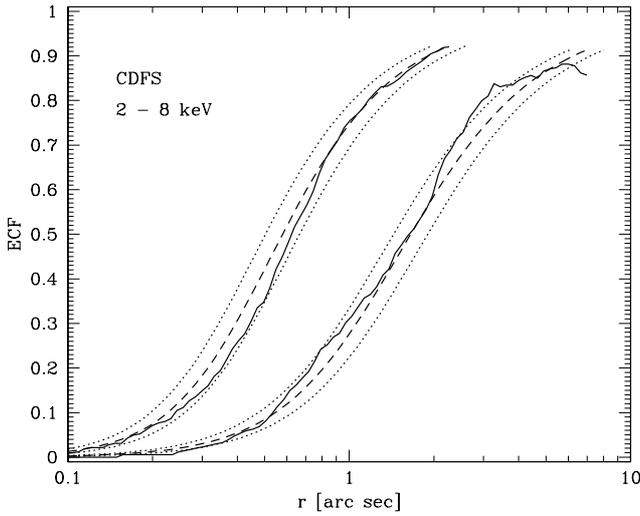


Fig. 4. Same as Fig. 3 for the *H* band.

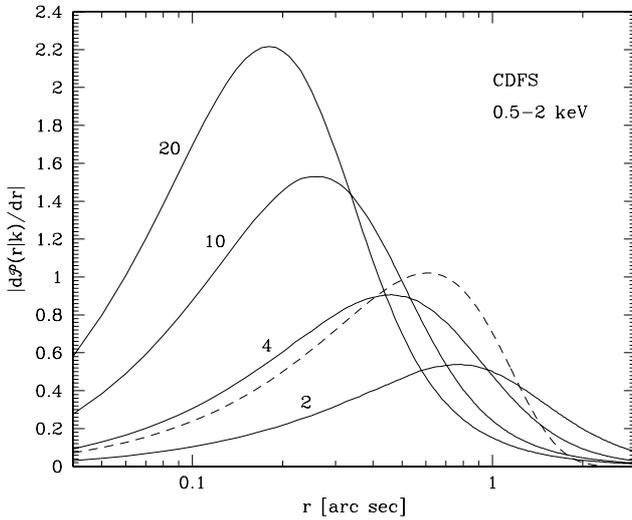


Fig. 5. The probability density of the distance to the nearest neighbor in the *S* band for counts produced by sources generating  $k = 2, 4, 10,$  and  $20$  counts. The dashed curve shows the expected distribution for the pure random count distribution.

sources in the *S* and *H* band. In Fig. 5, the probability densities for the CDFs based on the integral distributions  $\mathcal{P}(r|k)$  in the *S* band are shown for several values of  $k$ . To visualize more clearly details of the relevant distributions, probability densities, i.e.  $|d\mathcal{P}(r|k)/dr|$  rather than  $\mathcal{P}(r|k)$  are plotted. The dashed curve shows the probability density of the nearest neighbors distances for the random distribution.

### 3.3. The “afterglow” correction

A single cosmic ray can induce in the ACIS CCD detector a series of two or more “events”. This well recognized feature<sup>6</sup>, known as an “afterglow”, generates spurious weak sources in the data and could potentially affect the present investigation. The events span typically several seconds and do not need to occur in consecutive frames. As a result, the data contain clumps

<sup>6</sup> See <http://cxc.harvard.edu/ciao/why/afterglow.html> for details.

of counts, which mimic very weak sources. Fortunately, the time sequences of such afterglow counts span short time intervals relative to the exposure times of all the observations. This allows us to unambiguously identify essentially all the afterglows and remove them from the observation.

### 3.4. Strong source removal

To maximize the effect of the weak source population on the NNST, the strong sources should be removed from the data. The threshold source flux is defined by  $k_{\max}$  counts in Eq. (12). Thus, the value of  $k_{\max}$  should be set at a level that is sufficiently large to ensure that all the sources producing more counts than  $k_{\max}$  are found using standard source search criteria. On the other hand,  $k_{\max}$  should be small enough to ensure that the number of sources is adequately represented by the function  $N(s)$ .

A catalog of point sources in the AEGIS field is given by Laird et al. (2009). Around each catalogued source position a radius  $r_{85}$  was calculated using the local PSF. A number of counts within  $r_{85}$  was then obtained, and – by subtracting the background counts – the net counts  $k_{85}$  were assessed for each source. Since the relative variation in the exposure time within the fov are small, the average background was assumed for the entire field. The threshold counts  $k_{\max}$  characterizing the completeness limit of the catalog is not well defined in both data sets. Several trial values of  $k_{\max}$  were applied to assess this threshold in the investigated area. In the AEGIS field, all the sources brighter than  $k_{\max} = 20$  can be clearly recognized, and this value is low enough to allow the statistical analysis of still weaker sources. For a given  $k_{\max}$ , the source was excluded from further processing if  $k_{85} > 0.85 \cdot k_{\max}$ . To assure the removal of the source counts in the PSF wings, the rejection area was a circle with radius  $r_{\text{rem}} = 4 \cdot r_{85} + 4''$ . Strong sources in the CDFs were localized using the Giacconi et al. (2002) catalog based on 1Ms exposure. A range of  $k_{\max}$  between 20 and 45 were applied in the analysis in both energy bands.

### 3.5. Pixel randomization

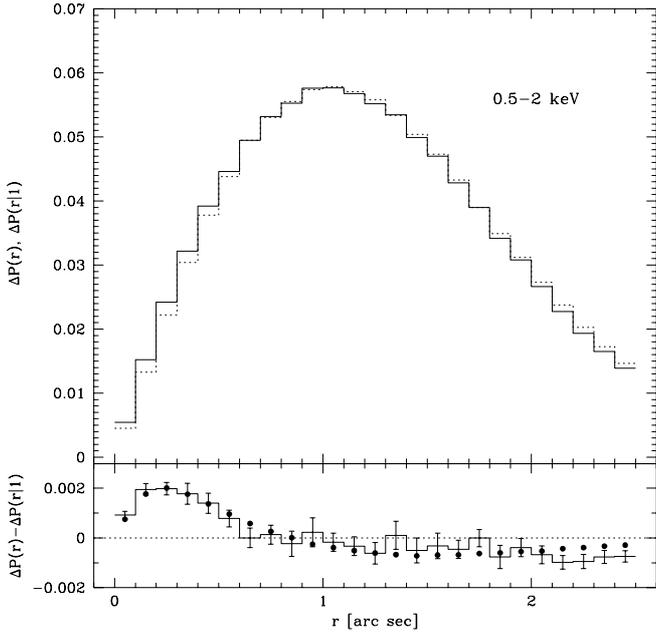
Positions of counts in the standard ACIS processing are randomized within the instrument pixel approximated by a square  $0'.492$  a side. Since the typical nearest neighbor separations are comparable to the pixel size, the NNST on small angular scales is smoothed by the count randomization. The effect is significant and generates most of the statistical noise in our calculations. To assess uncertainties introduced by the count randomization, 12 sets of randomized observations were produced using the original event data with non-randomized (integer) counts positions. The NNST was then determined for each observation and the data were used to obtain the slope  $b$  of the source number counts by means of Eq. (12). A scatter in the 12 count slope estimates represents the statistical error in the present method.

## 4. The source counts

### 4.1. AEGIS

#### 4.1.1. The soft band

Using the selection criteria given in Sect. 3, the area of the field and the total number of counts in the soft band amount to 97.5 square (sq) arcmin and 73711, respectively. After the removal of strong sources according to the procedure described above, the area is reduced to 92.7 sq. arcmin and the number of



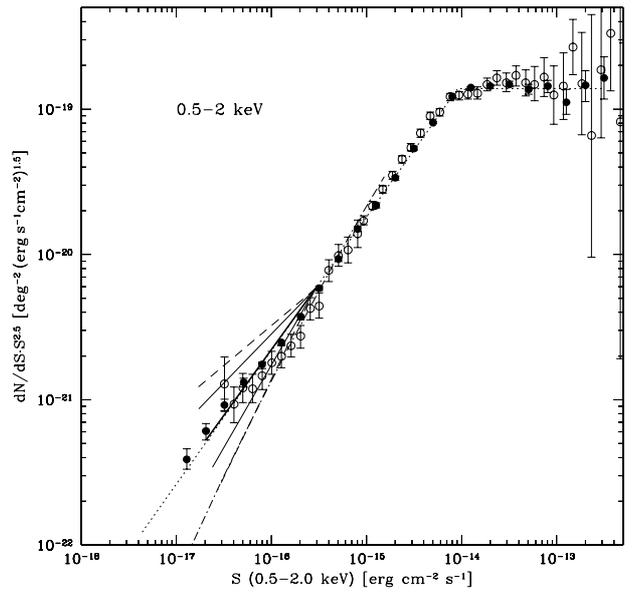
**Fig. 6.** The nearest neighbor probability distributions in the AEGIS data binned with  $\Delta r = 0''.1$  for the observed counts (solid histogram), and between random and observed counts (dotted histogram); error bars in the lower panel show  $1\sigma$  uncertainties.

counts to 48737. The average count density amounts to 0.146 per sq. arcsec and the average distance to the nearest neighbor for the random distribution is equal to  $1''.22$ .

The calculations were performed as follows. First, the count distributions were used to calculate the  $P(r)$  and  $P(r|1)$  probabilities. The distribution of distances to the nearest neighbor for each count defines the  $P(r)$ , while the distribution of the distances between the randomly distributed points and the nearest observed count is used to determine  $P(r|1)$ . The NNST was formulated in Sect. 2 using the cumulative probabilities, and estimates of the count slope  $b$  obtained from Eq. (12) for different separations  $r$  are dependent. To obtain a set of independent equations, we use the differential probability distributions  $\Delta P(r) = P(r) - P(r + \Delta r)$ . In the calculations, the relevant probability distributions were obtained over the entire range of observed separations with  $\Delta r = 0''.1$ . Using the least squares method, the best-fit value of the count slope  $b$  was then determined for each of the randomized distributions. Finally, the average and rms of  $b$  was calculated. The rms obtained in this way represents only the effects of the count randomization.

To facilitate comparison of the present results with those available in the literature, we adopted from GNL the number counts model for bright sources. We fixed the normalization and the number counts slope in the flux range not covered by the present analysis, i.e. at fluxes above  $S = 2.9 \times 10^{-16}$  cgs or  $k_{\max} = 20$ . The present calculations provide the best estimate of the slope below that flux down to the level determined by the sources generating two counts. The average flux of those sources depends on the number counts slope and within the power law approximation amounts to  $S_w \approx (3 - b) \cdot \eta$ , where  $\eta = 1.469 \times 10^{-17}$  cgs is the average conversion factor (see Table 2).

The probability distributions  $\Delta P(r)$  and  $\Delta P(r|1)$  averaged over 12 data sets are presented in the upper panel of Fig. 6. A histogram in the lower panel shows the difference between both



**Fig. 7.** Differential number counts in the 0.5–2 keV band normalized to the Euclidean slope. The data points and the dotted line are taken from Georgakakis et al. (2008); dot-dash curve shows the AGN model by Ueda et al. (2003); solid lines – the number counts estimates based on the NNST in the AEGIS field: thick solid line – the best power-law fit, thin solid lines –  $1\sigma$  uncertainty range due to the counting statistics; dashed lines – maximum total uncertainty including potential errors generated by inaccuracies in the PSF fitting. The count normalization at the bright end of the NNST model is fixed at the amplitude given by the Georgakakis et al. (2008) approximation.

distributions; the error bars indicate the histogram rms uncertainties obtained from the 12 pixel randomization. Dots show the model histogram obtained for the NNST best-fit solution slope of  $b = 1.595$ .

In Fig. 7, the results based on the NNST for the  $S$  band are superimposed on the differential number counts presented by GNL. The counts are normalized to the Euclidean slope of  $-2.5$ . Full dots and dotted line denote the counts and model by GNL, open circles are the counts of Kim et al. (2007), and the dot-dash curve are the predicted AGN counts from Ueda et al. (2003). The data points in Fig. 7 are based on a large number of *Chandra* pointings, including CDFS.

The thick line shows the NNST best-fit solution with the slope  $b = 1.595$  and the normalization fixed at  $S = 2.9 \times 10^{-16}$  cgs. The number counts above that flux are described by the GNL model, i.e. over a wide range of fluxes the counts are approximated by a power law with the differential slope of  $-1.58$  (some deviations at the low flux levels are evident and are discussed below). It is clear that the NNST solution is in full agreement with the direct source counts derived from the deep *Chandra* exposures. One should emphasize that the present NNST solution is based on the relatively shallow exposure relative to the data points. The sources at the low flux end of the present solution generate on the average in our data fewer than two counts each. The NNST is evidently capable of providing a sensitive estimation method for the population of sources which cannot be recognized as individual entities. We note that counts due to the sources generating the signal, i.e. producing  $2 \leq k \leq 20$  photons, constitute a small fraction of all the counts. In the present model, only about 1080 counts or 2.2% comes

from these sources, while the remaining 97.8% is distributed randomly.

#### 4.1.2. Error estimates

The number of counts attributed to the sources that generate  $2 \leq k \leq k_{\max}$  photons is small in comparison to the total number of events in the processed data even for the steep  $N(S)$  relationship. The NN pairs associated with the weak sources constitute a relatively small fraction of all pairs. This causes the rather large statistical uncertainties in the NNST method. In our case, the rms of the slope  $b$  in the  $S$  band amounts to 0.230. The situation is even worse in the  $H$  band, where the rms uncertainty in the slope reaches 0.34 (see below).

The evaluation of the systematic errors is less straightforward. Several potential effects could influence our estimate of the count slope. The first one, already discussed in Sect. 3.2, is associated with our PSF approximation. The complex shape and intricate position dependence of the PSF requires us to adopt an approximate treatment. Our PSF fits inevitably introduce errors. Unfortunately, the amplitude of these errors is difficult to assess. Because of that, we adopted a cautious approach to this problem. In addition to the best-fit PSF model, we considered two ancillary sets of PSFs that delineate the observed count distribution of the strong sources and also confine possible deviations introduced by our simplified model of the PSF variations over the field of view. A visual comparison of the actual count distributions and that of our model PSF shows that the  $\pm 15\%$  modification of the PSF width account for any potential deficiencies in our PSF calculations.

The  $\pm 15\%$  uncertainty of the PSF width introduces a substantial uncertainty in the slope estimate. For the PSF that is wider than the best-fit model by 15% we get  $b = 1.744 \pm 0.207$ , while for the 15% narrower,  $b = 1.421 \pm 0.267$ . One can summarize these results as follows. Statistical  $1\sigma$  limits around the best-fit solution of  $b = 1.595$  are defined as  $1.365 < b < 1.825$ , while the combined statistical and systematic uncertainties are  $1.154 < b < 1.951$ . The statistical uncertainties are shown in Fig. 7 with thin solid lines and the total uncertainties – with dashed lines. One should note that these “total” error estimates are highly conservative. They have been obtained by simple addition of the systematic and statistical errors assuming their highest “possible” values.

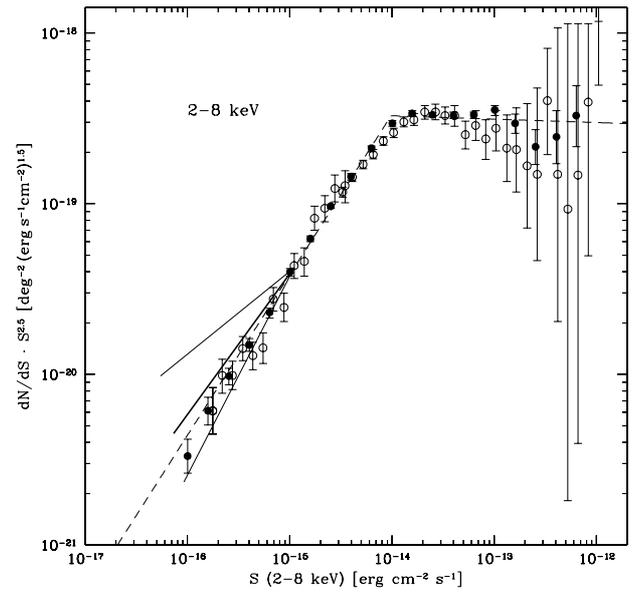
Another source of the slope estimate error is related to the uncertainty in the number counts normalization. Any modification of the number of sources at the strong flux end also affects the count slope. However, in the relevant flux range the realistic normalization uncertainties remain small (see Fig. 7) and do not affect considerably the slope uncertainty range.

Variations in the conversion factor over the field of view also do not play a significant role in our slope estimates. As pointed out in Sect. 3.1, one can incorporate these variations within the investigated area into an uncertainty in the count normalization  $N_0$ . Assuming the differential number counts  $N(S) = \mathcal{N}S^{-b}$  (with flux  $S$  in  $\text{erg cm}^{-2} \text{s}^{-1}$ ), a relationship between the count normalization  $N_0$  and the conversion factor  $\eta$  is

$$N_0 = \mathcal{N} \cdot \eta^{1-b}. \quad (15)$$

Thus, for  $b = 1.595$  and the rms in the conversion factor at the level of  $\sim 5.9\%$  (Table 3), the uncertainty in  $N_0$  amounts to  $\sim 3.5\%$  and is small in comparison with the uncertainty in  $\mathcal{N}$  itself.

Substantial variations in the PSF width with the distance from the telescope axis limit the effectiveness of the NNST



**Fig. 8.** The differential number counts in AEGIS in the 2–8 keV band normalized to the Euclidean slope. The data points and the dashed line are constructed using the 2–10 keV band from Georgakakis et al. (2008), solid lines – the number counts estimates based on the NNST, thick solid line – the best power-law fit, thin solid lines – the  $1\sigma$  uncertainty range due to the counting statistics.

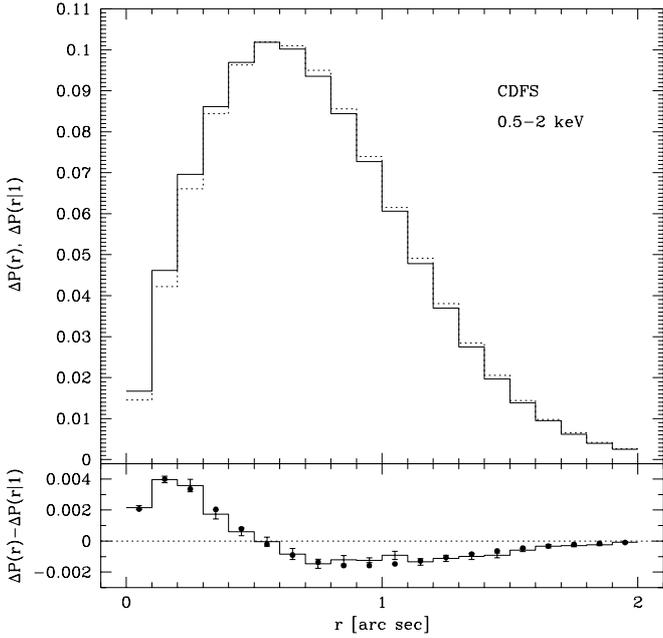
method. This is because the nearest neighbor distribution  $P(r)$  for Eq. (12) is estimated using actual data, i.e. the actual distribution of sources, while the probability  $\mathcal{P}(r|k)$  is determined by averaging the model sources over the field of view. This feature introduces additional uncertainty if the number of sources generating  $k \approx k_{\max}$  is small, but should be of lesser importance in the deep *Chandra* exposures.

#### 4.1.3. The hard band

A fraction of counts produced by discrete sources to all the counts in the 2–8 keV band is substantially smaller than in the 0.5–2 keV band. In effect, the NNST method is less effective in the  $H$  than in the  $S$  band, i.e. the slope estimates in the  $H$  band are subject to larger statistical uncertainties. Here we briefly summarize the results in the  $H$  band.

As a reference data, we used those published by GNL. The counts and the analytic fit in the band 2–10 keV given in that paper were converted to the band 2–8 keV assuming a power spectrum with the photon index  $\Gamma = -1.4$ .

After the removal of strong sources, the area and the number of counts used in the NNST amount to 94.4 sq. arcmin and 101 596, respectively. Applying the same procedure as in the  $S$  band, the best-fit slope  $b = 1.676 \pm 0.340$  was found using the 12 randomized distributions. This is shown in Fig. 8 with a thick line and two thin lines. The sources represented by the NNST solution, i.e. sources producing  $2 \leq k \leq 20$  counts, contribute just  $1214^{+979}_{-427}$  photons or  $1.19^{+0.96}_{-0.42}\%$  of all the counts. Owing to the large statistical uncertainty in this band, the NNST does not provide tight limits on the number counts and we have not plotted here lines representing the range of systematic uncertainties.



**Fig. 9.** *Upper panel:* the nearest neighbor probability distributions in CDFS binned with  $\Delta r = 0'.1$  for the observed counts (solid histogram), and between random points and the observed counts (dotted histogram). *Lower panel:* the difference between both distributions; the error bars show  $1\sigma$  uncertainties; dots – the best fit (see text for details).

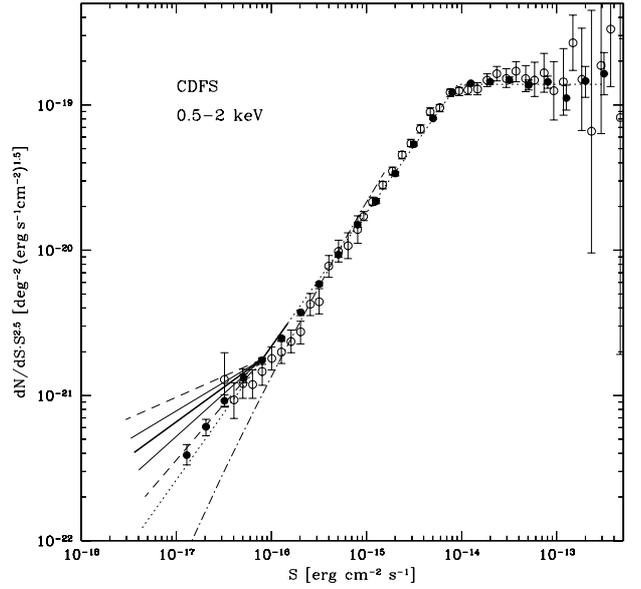
## 4.2. CDFS

### 4.2.1. The soft band

The accepted area and the total number of counts in the soft band amount to 71.8 sq. arcmin and 158 795, respectively. After the removal of strong sources, the area is reduced to 66.5 sq. arcmin and the number of counts to 110 996 for  $k_{\max} = 45$  and to 65.4 sq. arcmin and 108 778 counts for  $k_{\max} = 30$ . The average density of counts amounts to  $\sim 0.463$  per sq. arcsec and the average distance to the nearest neighbor for the random distribution is equal to  $0'.736$ . The subsequent calculations were performed in a similar way as for the AEGIS field. Because of the higher count density, the range of separations was reduced to  $0 < r < 2''$ .

Previous investigations of the deep *Chandra* fields, e.g. Kim et al. (2007) and GNL, provide essentially consistent assessments of the  $N(S)$  counts above the detection threshold for the discrete sources. In particular, in the interesting flux range GNL approximate the number counts by a power law with the slope of  $-1.58$ . To conform the present investigation to the observed counts at the bright end, we assume that the  $N(s)$  relationship defined in Eq. (10) above  $s_0 = 20$  counts matches exactly the GNL model. Thus, the only parameter to be determined using the set of equations generated by Eq. (10) is the slope  $b_2$  at the low flux end.

In the upper panel of Fig. 9, the probability distributions  $\Delta P(r)$  and  $\Delta P(r|1)$  are shown for  $k_{\max} = 40$ . The  $\Delta P(r)$  histogram is the average of 12 realizations of the pixel randomization routine. The difference between both distributions is shown in the lower panel. The error bars represent the rms scatter between 12 randomized observations. The dots show the average of 12 best-fit solutions obtained using the NNST. The analogous distributions constructed for several values of  $k_{\max}$  between 20 and 45 provided qualitatively similar results.



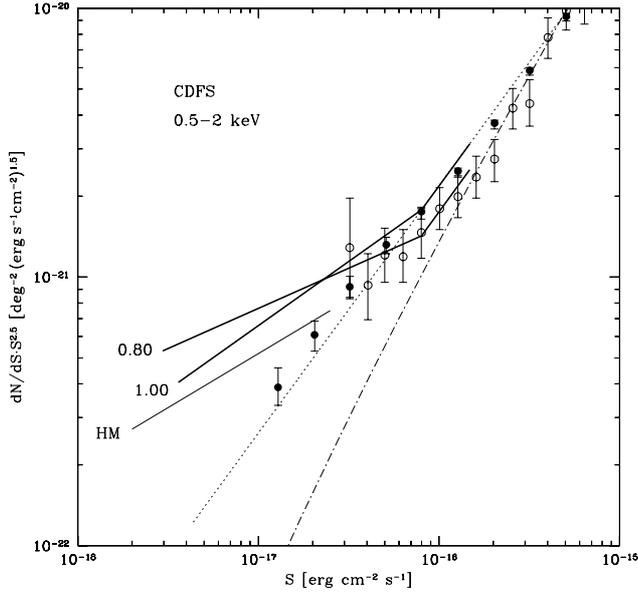
**Fig. 10.** Differential number counts in the 0.5–2 keV band normalized to the Euclidean slope. The data points and the dotted line are taken from Georgakakis et al. (2008); dot-dash curve shows the AGN model by Ueda et al. (2003). The bundle of lines shows the present results based on CDFS; thick line – the best-fit solution, thin lines –  $1\sigma$  statistical uncertainty, dashed lines – maximum total uncertainty range including the potential systematic errors generated by the approximations in the PSF fitting.

Flux  $s_0 = 20$  counts marking the slope change corresponds to  $S = 7.84 \times 10^{-17}$  cgs. The best-fit slope below the power-law break  $b_2 = 2.02^{+0.11}_{-0.08}$ , where the errors represent  $1\sigma$  statistical uncertainties (for the full discussion of uncertainties see below).

The  $N(S)$  counts representing the best-fit model are shown in Fig. 10 together with the existing data (see Fig. 7 for symbol definitions). The present investigation covers the flux range of  $\sim 4 \times 10^{-18} - 1.5 \times 10^{-16}$  cgs. The main results are shown in Fig. 10 with a set of lines indicating the best solution and the uncertainty ranges. As for the AEGIS data, the best-fit estimate of slope is shown with the thick solid line. The uncertainties introduced by the statistical character of the nearest neighbor method are indicated by the thin lines. These uncertainties result from variations in the nearest neighbor probability distributions produced by the randomization of counts within pixels<sup>7</sup>. The systematic errors affecting the investigation are probably dominated by the inaccuracies in the calculations of the  $\mathcal{P}(r|k)$  probabilities. These uncertainties were accounted for using the “extreme” ECF functions described in the Sect. 3.2. A set of 12 solutions was obtained using the  $\mathcal{P}(r|k)$  distribution derived from each of the side ECF. Then, the average values of the slope and the respective rms amplitudes in both sets were calculated. The dashed lines in Fig. 10 show the uncertainty range implied by these calculations, assuming the joint effect of the systematic errors and the rms scatter. As in the case of the AEGIS solution, this estimate of the total error is highly conservative.

At the lowest fluxes, our counts steepen substantially and are incompatible with the GNL model. The counts determined by GNL also seem to be steeper than the slope of 1.58, though the deviation from the power law is rather modest. Our investigation

<sup>7</sup> The minimum and maximum values of  $b$  in 12 data sets are 1.78 and 2.15.



**Fig. 11.** Low flux end of the differential number counts in the 0.5–2 keV band normalized to the Euclidean slope. The data points, the dotted and dot-dash curves as in Fig. 10. Broken solid lines: the NNST solutions for the  $N(S)$  with the bright end normalization according to Georgakakis et al. (2008) – label “1.00”, and the normalization reduced by 20% – label “0.80”. The label “HM” indicates the best estimate of counts obtained by Hickox & Markevitch (2007b).

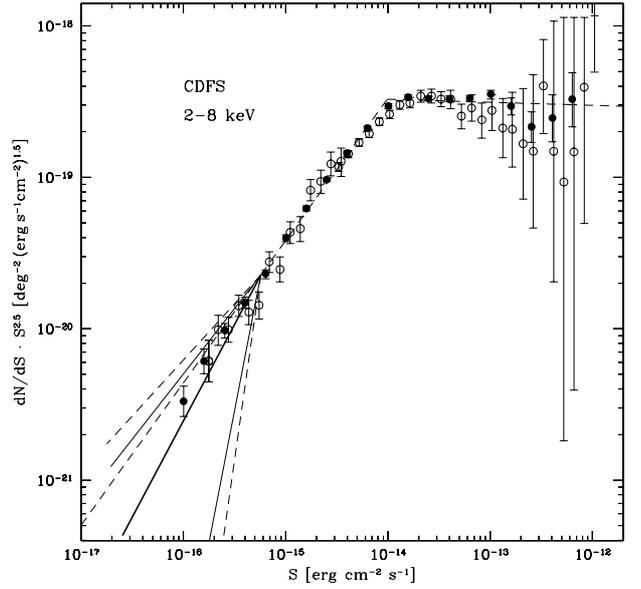
of the AEGIS field has not revealed any significant slope change. However, the AEGIS slope uncertainties are uncomfortably large and do not strongly constrain the  $N(S)$  relationship. The present results, which are based on the CDFS, also do not provide tight constraints. Nevertheless, the acceptable slopes seem to be distinctly steeper than those by GNL.

Equation (12) explicitly involves both parameters that define the source counts: the slope  $b$  and the normalization  $N_0$ . In our calculations, only the slope was estimated while the normalization was fixed. Nevertheless, we are able to obtain a formal solution for  $b$  and  $N_0$ . Unfortunately, the estimates of  $b$  and  $N_0$  found by the simultaneous fitting are highly correlated. This is because the NNST is affected predominantly by the total number of close pairs. Thus, a quality of the fit depends on the proper combination of  $b$  and  $N_0$ , rather than on each parameter separately. In Fig. 11, two best-fit solutions are shown for two different bright end normalization  $N_0$ . The broken line labeled “1.00” is the same as in Fig. 10, while the line labeled “0.80” shows the counts with the  $N_0$  reduced by 20%.

#### 4.3. The hard band

The NNST is sensitive to the excess of the close photon pairs relative to the number of pairs expected for the random distribution. A high overall count density in the 2–8 keV band significantly limits the efficiency of the NNST method for the weak source investigation. After the removal of strong sources, the average distance to the nearest neighbor for the random distribution amounts to just  $0'.51$  and the NNST applied to the  $5'$  radius fov has not produced any meaningful constraints on the  $N(S)$  slope.

To improve the signal-to-noise ratio, we confined our calculations to the central area of  $4'$  radius where the PSF is relatively narrow. The total number of counts within this limited field amounts to 179 371. After the removal of sources generating



**Fig. 12.** Differential number counts in the 2–8 keV band normalized to the Euclidean slope. The data points and the dashed line are constructed using the 2–10 keV band from Georgakakis et al. (2008); the thick solid line – the best fit solution, thin lines –  $1\sigma$  statistical uncertainty range, dashed lines – maximum total uncertainty range including the potential systematic errors generated in the PSF fitting.

more than  $k_{\max} = 40$ , the number of counts is reduced to 144632. Our slope estimate and its statistical uncertainty  $b = 1.23^{+0.53}_{-2.28}$  are consistent with the available source counts. Unfortunately, the constraints imposed on  $N(S)$  are still not very tight, the lower slope limit in particular being insignificant. Nevertheless, the NNST rather strongly excludes any substantial steepening of the counts, in contrast to the result obtained for the  $S$  band. This is shown in Fig. 12, where the present results are compared with the available observational material. The points with the error bars are drawn using the 2–10 keV data presented by GNL. The power spectrum with a photon index of  $-1.4$  was assumed to convert fluxes from the 2–10 keV band to our  $H$  band. The NNST solution and  $1\sigma$  statistical uncertainties are indicated by the solid thick line and two thin lines, respectively. The dashed lines indicate the combined effect of the maximum potential systematic errors and the statistical noise. The systematic errors were assessed in a similar way to those in the  $S$  band. The width of the best-fit PSF was altered by  $\pm 15\%$  to accommodate for the conceivable systematic deviations in the analytic fits from the actual PSF (see Fig. 4). On the basis of the modified PSFs, the new  $\mathcal{P}(r|k)$  probabilities were then derived and used in the subsequent calculations.

## 5. Conclusions

We have developed and applied the new method to assess the population of sources substantially below the standard detection thresholds. It is demonstrated that the NNST provides constraints on the number of sources that generate just a couple of counts. The comparison of our count estimates based on the medium depth exposure with the actual counts in the CDFS has allowed us to evaluate effectiveness as well as the limitation of the method.

Figure 11 clearly shows that our slope estimate in the soft band below  $S \approx 7 \times 10^{-17}$  cgs is substantially steeper than

the estimates by GNL. Although these authors also observe the steepening of counts at the lowest attainable flux levels, their slope change is markedly smaller and the discrepancy between our results remains unexplained. Both the NNST and the GNL approaches make full use of the Poissonian character of the count distribution produced by the individual source. Nevertheless, both methods are distinctly different. GNL assess the source presence by counting the events within the detection cell, while here we have analyzed the NN distances between the events. We have tested the NNST method on the AEGIS observations, but it still should be considered as a new tool, and one cannot exclude that unrecognized systematic errors have influenced the present result. Hopefully, the recent *Chandra* 4Ms observation of the CDFS will help clarify this uncertainty.

A novel approach to the problem of the extremely weak sources in the *Chandra* deep pointings was tested by [Hickox & Markevitch \(2007b\)](#). They investigated the possible association of counts in the CDFN with the HST objects detected in the  $z_{850}$  band. The regions around the detected X-ray sources were excluded. Approximately 2200 HST objects distributed within an area of 12.2 sq. arcmin at the center of CDFN were used in the analysis. The distribution of excess counts correlated with the HST objects indicates that in the flux range  $\sim 2 \times 10^{-18} - 2 \times 10^{-17}$  cgs the data support quite steep counts with  $b = 2.1^{+0.5}_{-0.3}$ . It is shown in Fig. 11 with a solid line labeled “HM”. The [Hickox & Markevitch \(2007b\)](#) estimate represents the lower limit to the counts, since some X-ray sources might not have been detected by HST.

The integral source counts are constrained by the amplitude of the extragalactic XRB component. A varying galactic contribution to the total signal makes the estimates of the extragalactic part in the  $S$  band somewhat uncertain. For the purpose of the present discussion, as a reference figure we adopt the XRB assessment by [Moretti et al. \(2003\)](#) of  $f_s = (7.53 \pm 0.35) \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$ .

One should note, however, that measurements of the total extragalactic XRB flux are subject to systematic uncertainties related to the absolute calibration errors and cosmic variance. It is usually assumed that the extragalactic component of the XRB is acceptably approximated by the power law with the photon spectral index  $\Gamma = 1.4 - 1.5$  over a wide energy band above 0.5 keV.

The XRB spectrum between 1 keV and  $\sim 20$  keV adequately resembles a power law with a photon index of 1.4 ([De Luca & Molendi 2004](#), and references therein), but below 1 keV a conspicuous XRB softening is observed ([Hasinger 1993](#); [Chen et al. 1997](#)). The soft excess varies from field to field and evidently exhibits some local and Galactic contribution (e.g. [Markevitch et al. 2002](#)). However, the fraction of the soft XRB generated within the Galaxy is not well established ([Soltan 2007](#)). Consequently, the exact spectral characteristics of the extragalactic XRB are not satisfactorily determined. To assess the amplitude of the extragalactic component, the whole soft excess is commonly attributed to the galactic emission (e.g. [Hickox & Markevitch 2006](#)). Applying this assumption and using the absolute measurements of the XRB in the CDFS above 1 keV by [Hickox & Markevitch \(2006\)](#), we infer that the total extragalactic XRB in the 0.5–2 keV band  $f_s = 7.32 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$  for  $\Gamma = 1.4$  and  $7.53 \times 10^{-12} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ deg}^{-2}$  for  $\Gamma = 1.5$ . The assumption about the power law shape of the XRB over the whole considered energy band results in a close coincidence of these figures with the [Moretti et al. \(2003\)](#) estimate. Nevertheless, it should be stressed that spectral properties of the weak sources generating the unresolved fraction of the XRB are not well constrained (see

below) and the power-law character of the extragalactic XRB below 1 keV is not determined observationally.

The counts described by the GNL model integrated above  $S = 7.5 \times 10^{-17}$  cgs ( $\equiv 20$  counts in the present investigation) generate  $\sim 78\%$  of the XRB. Using our slope best estimate of  $b = 2.02$ , sources producing  $2 \leq k \leq 20$  counts contribute a further 10%. Assuming that the point-like sources generate the whole extragalactic XRB, the  $N(S)$  counts should flatten for  $S \lesssim 4 \times 10^{-20}$  cgs. If some fraction of the soft XRB is attributed to the diffuse component, such as the WHIM, the counts flattening has to occur at higher flux levels. In particular, if the WHIM generates 10% of the XRB ([Soltan 2007](#)), the source counts cannot continue with the same slope below  $\sim 2 \times 10^{-18}$  cgs. Apparently, even the modest extension of the  $N(S)$  relationship supplemented with the precise measurements of the integrated XRB would provide a valuable data set to investigate of the diffuse component.

In the  $S$  band, neither the GNL nor our results are consistent with the predicted counts of AGNs based on a wide class of evolutionary models ([Miyaji et al. 2000](#); [Gilli et al. 2001](#); [Ueda et al. 2003](#); [Bauer et al. 2004](#)). Consequently, a new population of objects emerging below  $\sim 10^{-16}$  cgs is required.

Young and/or starburst galaxies appear as natural candidates for these sources. Owing to the well-established correlation of the X-ray luminosity with the star formation rate ([Ranalli et al. 2003](#)), the contribution of those galaxies to the XRB is expected to rise at the low flux levels. [Bauer et al. \(2004\)](#) analyzed the sources in the *Chandra* deep exposures and found that at  $\approx 2.5 \times 10^{-17}$  cgs in the  $S$  band, AGNs still outnumber the star-forming galaxies (SFG) by a factor of  $\sim 4$ . However, extrapolating the SFG steep count slope, they show that the SFGs reach the AGN densities at  $\sim 1 \times 10^{-17}$  cgs and start to dominate the number counts at lower fluxes. Owing to the large uncertainties involved in the process of identification of various classes of X-ray sources (see [Bauer et al. 2004](#), for details), a fair conclusion that one can draw is that models based on the extrapolation of the AGN and SFG number counts are consistent with the measurements obtained by means of the NNST.

A question of the discrete source contribution to the diffuse background in the radio domain is also present in the literature. It is interesting that the extragalactic radio source counts exhibit pronounced slope variations of a character resembling those observed in the soft X-rays (see [Vernstrom et al. 2011](#), for the compilation of the radio data). The counts derived from the VLA-COSMOS survey at 1.4 GHz [Bondi et al. \(2008\)](#) above  $\sim 0.5$  mJy indicate the slope of 1.6, while just below that flux the slope is equal to  $\sim 2.3$ . The counts decline again below  $\sim 0.1$  mJy.

In the  $H$  band, the NNST provides consistent results with the previous investigations and the source counts do not exhibit any measurable steepening. This indicates that the weak sources generating the count rise in the  $S$  band have soft spectra and their contribution to the XRB above 2 keV is low.

*Acknowledgements.* I thank all the people who created the Chandra Interactive Analysis of Observations software for producing a really user-friendly environment. This work has been partially supported by the Polish MNiSW grant N N203 395934.

## Appendix A: Impact of the exposure fluctuations on the NNST

We investigate the effects of the small variations in the exposure over the field of view on the probability distributions  $P(r)$  and  $P(r|1)$  in Eq. (12).

The observed distribution of counts is considered here as the realization of a Poisson process. According to this premise, the probability of getting an event in the  $x - y$  plane of the field of view (fov) is described by a smooth function  $\rho(x, y)$ . The  $\rho(x, y)$  amplitude is proportional to the exposure map (Sect. 3.1). One can define the “true” probability density  $\rho_o(x, y)$  that would describe the distribution of counts expected for the perfect instrument with the flat exposure map and – consequently – a constant conversion factor. Thus, the  $\rho$  distribution is related to  $\rho_o$  and the fluctuations of the exposure map, EM

$$\rho(x, y) = \rho_o(x, y) \cdot \frac{\text{EM}(x, y)}{\text{EM}_o}, \quad (\text{A.1})$$

where  $\text{EM}_o$  is the exposure map of the perfect instrument. A natural normalization of the exposure map is assumed  $\langle \text{EM}(x, y) \rangle = \text{EM}_o$ , where the brackets  $\langle \dots \rangle$  denote the averaging over the fov. Since the distributions  $\rho_o$  and EM are independent,  $\langle \rho \rangle = \langle \rho_o \rangle$ . To quantify the amplitude of the EM fluctuations with a single parameter  $\varepsilon$ , we define a function  $f(x, y)$

$$\varkappa(x, y) = \frac{\text{EM}(x, y)}{\text{EM}_o} = 1 + \varepsilon f(x, y), \quad (\text{A.2})$$

in such a way that  $\langle f \rangle = 0$  and  $\sigma_f^2 = \langle f^2 \rangle = 1$ . We also have  $\langle \varkappa \rangle = 1$  and  $\sigma_\varkappa = \varepsilon$ , where  $\sigma_\varkappa$  is the rms of  $\varkappa$ . Since  $\sigma_\varkappa = \sigma_{\text{EM}}/\text{EM}_o$ , the data in the Table 3 show that the  $\varepsilon$  parameter characterizing the present observational material remains small; in particular, it is equal to 0.059 and 0.062 for energy bands  $S$  and  $H$ , respectively.

The expected nearest neighbor probability distributions  $P(r|1)$  and  $P(r)$  are related to the count distribution  $\rho$  in the way:

$$P(r|1) = \langle e^{-\pi r^2 \rho(x, y)} \rangle, \quad (\text{A.3})$$

$$P(r) = \frac{1}{\langle \rho \rangle} \langle \rho(x, y) \cdot e^{-\pi r^2 \rho(x, y)} \rangle. \quad (\text{A.4})$$

Substituting Eqs. (A.1) and (A.2) into Eqs. (A.3) and (A.4), one can expand the distributions  $P(r|1)$  and  $P(r)$  as powers

of the parameter  $\varepsilon$ . Since the distributions of  $f(x, y)$  and  $\rho_o(x, y)$  are uncorrelated, one finds that the linear term in  $\varepsilon$  vanishes.

## References

- Allen, C., Jerius, D. H., & Gaetz, T. J. 2004, Proc. SPIE, 5165, 423  
 Bauer, F. E., Alexander, D. M., Brandt, W. N., et al. 2004, AJ, 128, 2048  
 Bondi, M., Ciliegi, P., Schinnere, E., et al. 2008, ApJ, 681, 1129  
 Brandt, W. N., & Hasinger, G. 2005, ARA&A, 43, 827  
 Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, A&A, 497, 635  
 Chen, L.-W., Fabian, A. C., & Gendreau, K. C. 1997, MNRAS, 285, 449  
 De Luca, A., & Molendi, S. 2004, A&A, 419, 837  
 Galeazzi, M., Gupta, A., Covey, K., & Ursino, E. 2007, ApJ, 658, 1081  
 Georgakakis, A., Nandra, K., Laird, E. S., Aird, J., & Trichas, M. 2008, MNRAS, 388, 1205 (GNL)  
 Giacconi, R., Zirm, A., JunXian, W., et al. 2002, ApJS, 139, 369  
 Gilli, R., Salvati, M., & Hasinger, G. 2001, A&A, 366, 407  
 Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79  
 Hasinger, G. 1993, Observational Cosmology, ASP Conf. Ser., 51, 439  
 Hasinger, G., Burg, R., Giacconi, R., et al. 1993, A&A, 275, 1  
 Henley, D. B., Shelton, R. L., & Kuntz, K. D. 2007, ApJ, 661, 304  
 Hickox, R. C., & Markevitch, M. 2006, ApJ, 645, 95  
 Hickox, R. C., & Markevitch, M. 2007a, ApJ, 661, L117  
 Hickox, R. C., & Markevitch, M. 2007b, ApJ, 671, 1523  
 Kim, M., Wilkes, B. J., Kim, D.-W., et al. 2007, ApJ, 659, 29  
 Laird, E. S., Nandra, A., Georgakakis, A., et al. 2009, ApJS, 180, 102  
 Lehmer, B. D., Brandt, W. N., Alexander, D. M., et al. 2008, ApJ, 681, 1163  
 Lehmer, B. D., Alexander, D. M., Bauer, F. E., et al. 2010, ApJ, 724, 559  
 Luo, B., Bauer, F. E., Brandt, W. N., et al. 2008, ApJS, 179, 19  
 Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, ApJS, 187, 560  
 Markevitch, M., Bautz, M. W., Biller, B., et al. 2003, ApJ, 583, 70  
 Miyaji, T., & Griffiths, R. E. 2002, ApJ, 564, L5  
 Miyaji, T., Hasinger, G., & Schmidt, M. E. 2002, A&A, 353, 25  
 Ranalli, P., Comastri, A., & Setti, G. 2003, A&A, 399, 39  
 Moretti, A., Campana, S., Lazzati, D., & Tagliaferri, G. 2003, ApJ, 588, 696  
 Softan, A. M. 1991, MNRAS, 250, 241  
 Softan, A. M. 2003, A&A, 408, 39  
 Softan, A. M. 2007, A&A, 475, 837  
 Tananbaum, H., Avni, Y., Branduardi, G., et al. 1979, ApJ, 234, L15  
 Treister, M., Urry, C. M., & Virani, S. 2009, ApJ, 696, 110  
 Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886  
 Vernstrom, T., Scott, D., & Wall, J. V. 2011, MNRAS, accepted  
 [arXiv:astro-ph/1102.0814]  
 Worsley, M. A., Fabian, A. C., Bauer, F. E., et al. 2005, MNRAS, 357, 1281  
 Zamorani, G., Henry, J. P., Maccacaro, T., et al. 1981, ApJ, 245, 357