Astron. Astrophys. 354, 847–852 (2000)

ASTRONOMY AND ASTROPHYSICS

Extraction of a deterministic component from ROSAT X-ray data using a wavelet transform and the principal component analysis

II. The data analysis

L. Liszka[1], **A.G. Pacholczyk**[2], **and W.R. Stoeger S.J.**[3]

[1] Swedish Institute of Space Physics, Umeå Division, Sörfors 634, 905 88 Umeå, Sweden
[2] Steward Observatory, The University of Arizona, Tucson, AZ 85721, USA
[3] Vatican Observatory Research Group, Steward Observatory, The University of Arizona, Tucson, AZ 85721, USA

Received 27 September 1999 / Accepted 9 December 1999

Abstract. A deterministic component in the X-ray photon series from Seyfert 1 galaxies and quasars, observed by ROSAT, is studied with the wavelet spectra method. A semi-regular deterministic modulation of the photon series is stronger and occurs more frequently in Seyfert 1 nuclei than in quasars indicating that by studying statistical properties of an X-ray photon train it is possible to identify unambiguously the character of its source. An interpretation of these differences is suggested within a scenario provided by a black hole nuclear cluster paradigm.

Key words: methods: data analysis – methods: statistical – galaxies: quasars: individual: 3C 273 – galaxies: individual: NGC 5548 – galaxies: Seyfert – X-rays: galaxies

1. Introduction

There is some evidence that X-ray photons from astronomical sources can not be fully described by a Poisson process. A deterministic modulation of the photon series is detected, which is reflected in observed wavelet spectra of photon counts (Liszka & Holmström 1999; hereafter referred to as Paper I). A part of the short-term variations of photon counts recorded by the ROSAT most likely contains deterministic information corresponding to the properties of the source. Wavelet spectra are a useful tool in studying these short-term temporal variations.

In particular, for active galactic nuclei (AGN) there is an indication of a physical source of deterministic variations of the X-ray flux. Pacholczyk & Stoeger (1994) propose "building blocks" in the X-ray photon flux from AGN resulting from discrete events in the source such as accretion disk flares or smaller black holes passing through the accretion disks of larger black holes. The present study indicates evidence for such building blocks.

The latest results from the Rossi X-ray Timing Explorer (RXTE) also indicate the presence of quasi-regular oscillations of the X-ray flux from some sources, both galactic and extragalactic. The oscillations may be generated by relativistic effects (Nowak et al. 1997) or by oscillations of the accretion disk.

Send offprint requests to: L. Liszka (ludwik@irf.se)

2. Data selection

In the present study photon-event files from the Position Sensitive Proportional Counter (PSPC) instrument onboard the ROSAT satellite were used. Data corrected for instrumental effects were obtained from HEASARC archives.

Since the spacial resolution of the instrument decreases towards the lowest energies, only photons with energies above 0.5 keV were used. Furthermore, there is always a problem of interference from nearby sources with completely different statistical properties. One example is 3C 345 which is surrounded by a number of close quasars (Arp 1997). Other example is NGC 6251 where a X-ray jet and halo have been identified close to the central source (Mack et al. 1997). For that reason, for each source, a centrum of gravity of the image was determined and then a circular section of the image, limited by 1/e of the centrum intensity was selected.

In the present paper we are studying 19 AGN of Seyfert 1 (S1) type and 9 quasars (QSOs). Table 1 lists those sources together with the ROSAT files we analyze.

In order to carry on the present study, it was necessary to find a proper method of extracting the information from the photon series. A combined approach using both wavelet spectra and probability density distributions of time intervals between photons was chosen.

We derive wavelet spectra of temporal variations of X-ray counts from the photon event files using the Morlet wavelet transform (see Paper I). As the wavelet spectrum does not change its character when the length of sampling bins is changed, photon counts were sampled in 1-second bins, even for weak sources. An example of wavelet spectrum for NGC5548, ROSAT observation request number (ROR) 701242, bin lengths of 1, 2, 4 and 8 seconds, is shown in Fig. 1.

With long sampling bins it is obvious that high frequency information is lost. For that reason 1 second bins were used for all data in the present study. For weak sources it results, of course, in

Table 1. Objects analyzed and their ROSAT files

Object	ROSAT observation request number (ROR) & references
Seyfert 1	Schartel et al. (1997); Nandra et al. (1997)
NGC3227	700996
NGC3783	701357
3C120	700913, 700916, 700919, 700945
NGC4051	700143, 700557, 700558
NGC4151	700911, 700914, 700917
IC 4329 A	700907
MKN 509	700105, 701360, 701361, 701362
NGC5506	700332
NGC5548	150071, 701239, 701241, 701243, 701244, 701246, 701247, 701248, 701249, 701270, 701271, 701280, 701281, 701282, 701290, 701291, 701292, 701293
NGC6251	700098
NGC6814	700923, 701090, 701460, 701472, 701478
NGC7213	700131
NGC7314	700873
NGC7469	700106
MKN 766	700970, 701353
MKN 841	700102, 700257
Fairall 9	700132
MCG-2-58-22	700998, 701250, 701364
MCG-6-30-15	700293
QSO	Sambruna (1997)
0208-512	701159, 701160, 701161, 701163, 701164, 701167, 701168, 701158, 701157, 701156
0403-132	700271
0521-365	701179
836+710	700493, 701061
3C273	141509, 700191, 701255, 701256, 701258, 701259, 701260, 701261, 701262, 701263, 701264, 701265, 701266, 701267, 701268, 701269
3C345	201538, 700869, 700870
1921-293	900338
1928+738	700142
3C446	900337

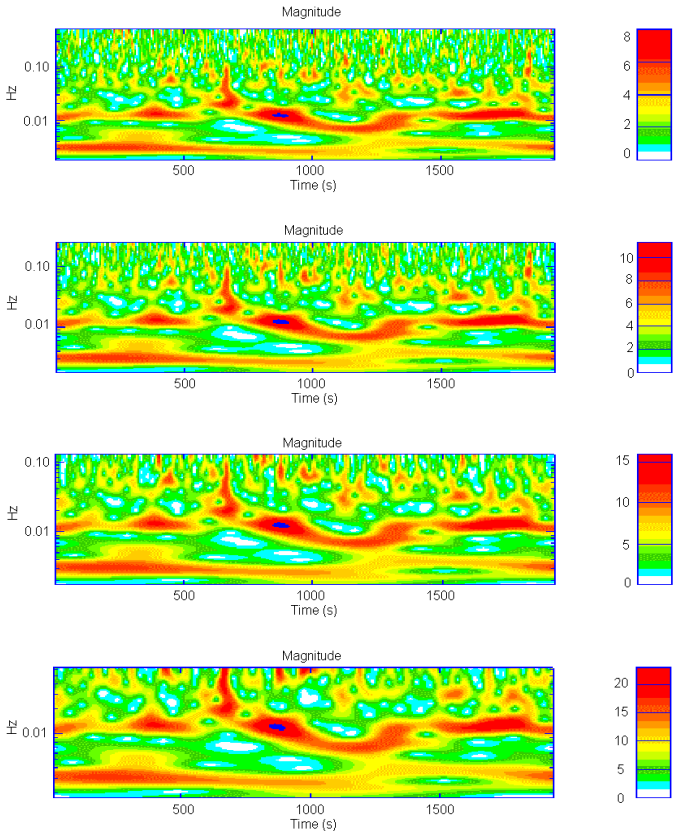

Fig. 1. An example of wavelet spectrum for NGC5548, ROR 701242, bin lengths of 1, 2, 4 and 8 seconds. The frequency scale in Hz.

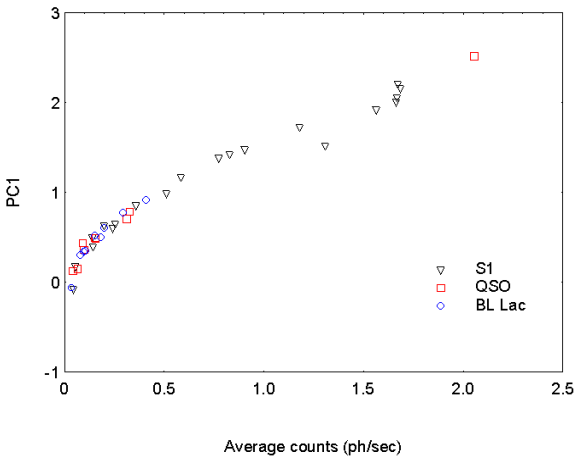

Fig. 2. The PC1 of the survey spectra as a function of the average counting rate.

very low counting rates. In order to perform the non-linear filtering of the data, as many as 128 dilations (equivalent to frequency steps) in the Morlet wavelet transform were needed. Only observation periods covering at least 1024 seconds, without interruption, were used. Observation periods covering a multiple of 1024 seconds were divided into several samples. A uniform sample length for all analyzed data was used so the same dilation (frequency) scale could be used for all sources and all observations.

A survey of the data has been performed using low-resolution wavelet spectra, with only 15 dilations. The principal component analysis of the data shows that the dominant component in the wavelet spectra is related to the apparent brightness of the source described here by the average counting rate during all analyzed observation periods (cf. Fig. 2).

Thus, the apparent brightness of the sources is a factor dominating the variability properties of the analyzed data. It appears to be responsible for about 25% of the total variance. The influence of apparent brightness therefore obscures the high order, source specific effects. As the first step in the analysis, that influence may be removed from the data matrix. It may be done using a Principal Component Analysis (PCA)-based decomposition technique described in Paper I. After the removal of the

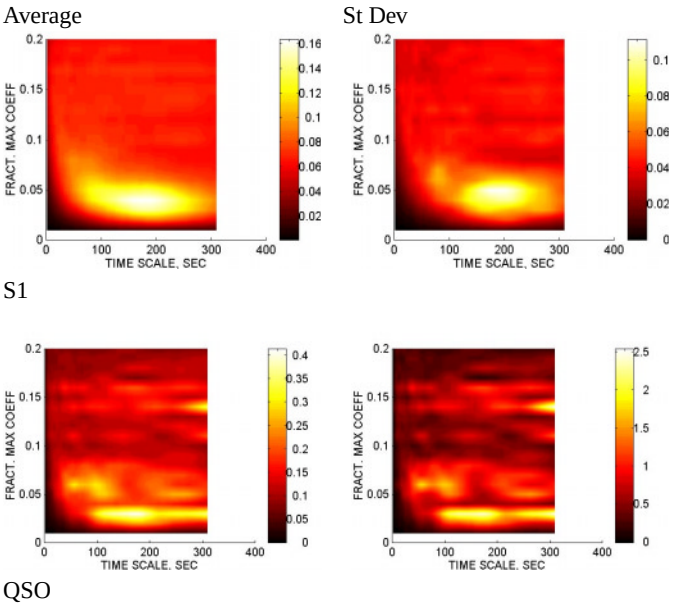

Fig. 3. A comparison of time-scale average spectra and their standard deviations for both categories of objects

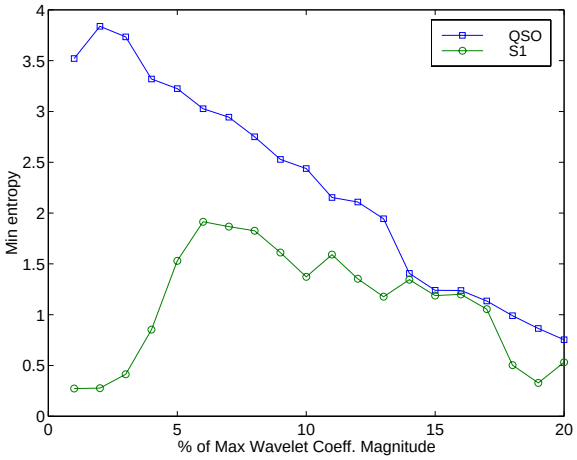

Fig. 4. Minimum entropy of distributions for each level of wavelet coefficient magnitude.

Table 2. Subgroup of objects with low X-ray counting rates

Source	Number of objects	Number of analyzed samples	Min counts ph/sec	Max counts ph/sec	Average counts ph/sec
S1	8	94	0.043	0.359	0.179
QSO	8	75	0.042	0.327	0.156

most significant principal component (PC1) a new data matrix, hopefully corrected for effects of the apparent brightness, is obtained. However, it has been found that the principal component analysis alone is not able to eliminate all the effects due to the apparent brightness.

3. The analysis of ampligrams and time scale spectra

The present study was therefore aimed at isolating the weak components of the X-ray photon trains using the technique described in Paper I and then searching for the differences between the two groups of sources. In order to ensure negligible influence from the apparent brightness, the brightest sources were removed from the data. Thus 3C 273 was removed from the QSO group and that only 8 S1 type sources were used: NGC 6814, NGC 6251, NGC 5506, NGC 7314, NGC 4151, NGC 4051, NGC 3227 and MCG-2-58. The subgroup selected according to the above limitations is described in Table 2.

All the data in the photon-count time series were converted into low-20 ampligrams, showing components corresponding to the lowest 20% of the maximum wavelet coefficient amplitude (see Paper I for detailed description of the method). The ampligrams were again wavelet-transformed to obtain a time-scale spectrum for each individual 1024-second sample. Time-scale spectra were then used to construct both the average time-scale spectrum for each category of sources and the corresponding graph of standard deviations, these are presented in Fig. 3.

There are obvious differences between the categories, especially for S1 and QSO. As can be seen, the structures in the S1 time-scale spectrum have larger extent in the vertical direction than those for the QSO time-scale spectrum. That means more deterministic structure in S1 variability data. Another difference is that the levels in the time-scale spectra and their standard deviations are about 2.5 times larger for QSO than for S1, which indicates a significantly larger variability for QSOs. As we have removed apparent brightness effects, this is an inherent characteristic of QSOs.

4. The entropy of amplitude distributions of ampligrams

There is another method for studying differences in dynamical properties between sources and groups of sources. It has been described in Sect. 7 of Paper I. For each level of wavelet coefficient magnitude in the ampligram a normalized amplitude distribution is created. For levels with pure stochastic components, Gaussian amplitude distributions will be obtained. For levels with deterministic components, broader amplitude distributions, often revealing structures, will be found.

It is useful to quantify the character of the distribution by calculating the distribution entropy for each level of wavelet coefficient magnitude. The distribution entropy, E_k, is given by Shannon & Weaver (1949):

$$E_k = -\sum_i p_{ik} \ln p_{ik}, \quad i = 1, \ldots, B;\ k = 1, \ldots, M \qquad (1)$$

Here p_{ik} is the measured probability density, B is number of distribution bins (100 in our case) and M = 20 for a low-20 ampligram. The results of the analysis performed for the 2 groups of objects with lower apparent luminosities, listed in Table 2 are shown in Fig. 4.

It may be seen that the S1 objects show lower minimum values of entropy for low levels of wavelet coefficient magni-

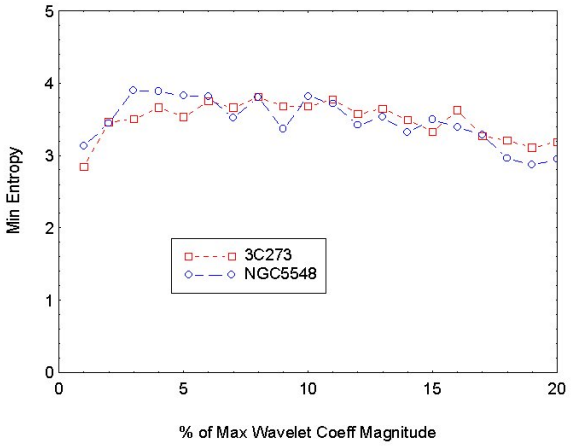

Fig. 5. A comparison of lowest measured entropy for two strong sources: NGC 5548 and 3C 273.

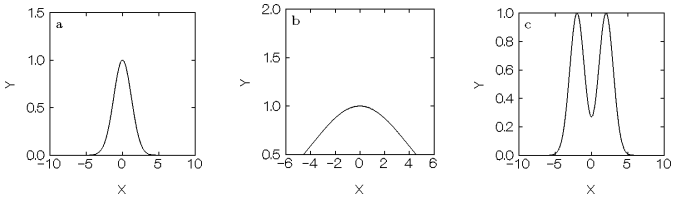

Fig. 6a – c. The distribution shape and the value of entropy: **a** pure Gaussian noise, narrow distribution, low entropy; **b** the distribution broadened due to presence of semi-regular components with different periods, increased entropy; **c** strong harmonic component in the data, decreased entropy.

tudes (<14%) where most of deterministic structures is known to occur in the ampligrams. Thus S1 ampligrams show more pronounced deterministic structures (lower entropy) than the ampligrams for QSOs.

It must be remembered, since the entropy is a logarithmic variable, that calculation of averages and standard deviations are really not possible. It may be seen from Fig. 4 that the difference between the entropy for S1 and QSO is undoubtly significant up to 6% of the wavelet coefficient magnitude. Between 6% and 14% of the wavelet coefficient magnitude there might still be significant differences.

For strong sources, like NGC5548 (S1 type source) and 3C 273 (QSO type source) the situation is different (see Fig. 5). The reasons for the difference are given below.

The entropy may be used as a measure of the distribution properties, but it must be remembered that the relation between the entropy and the character of the distribution is not simple. As it has been pointed out in Paper I, the entropy is relatively low for a narrow distribution of pure Gaussian noise, it increases when the distribution broadens due to presence of semi-regular components with different periods and finally it decreases when deterministic components become stronger. See Fig. 6.

It is interesting to study more in detail the shape of the distributions for different groups of sources and how it varies with the apparent luminosity of the source. Since the entropy is a logarithmic quantity a conventional statistics can not be used. For that reason neural network models for both types of sources (S1 and QSO) were constructed, where the entropy of the ampligram distributions was modeled as a function of the percentage of the maximum wavelet coefficient magnitude and of the apparent luminosity of the source (average photon counts during the observation periods). A simple model architecture with a back-propagation neural network having two input processing elements (wavelet coefficient magnitude and average counts), seven processing elements in the hidden layer and one output processing element for the entropy was employed. The number of processing elements in the hidden layer should be related to the number of expected subpopulations in the data. A measure of the model's accuracy may be obtained using a test subset of data. Scatter plots of entropy retrieved from the respective model versus the measured entropy are shown in Fig. 7.

The general dependence of the distribution entropy on the apparent luminosity (average counting rate) for both groups of sources, as retrieved from respective models is shown in graphs of Fig. 8.

Both these figures illustrate properties of the entropy curves for different sources. A vertical section of the graph corresponds to curves shown in Fig. 5. For the QSO objects there is a monotonous increase of entropy towards small wavelet coefficients due to broadening of the distribution. It is valid for all apparent luminosities in the investigated range. For the S1 objects the entropy increases with decreasing wavelet coefficients up to a certain apparent luminosity, after which the deterministic components become so distinct that the entropy starts to decrease again towards small wavelet coefficients. The latter situation corresponds to case (c) in Fig. 10.

The curves of Fig. 5 are in agreement with the above graphs remembering that that the average counting rate for analyzed observation periods is for NGC 5548 is 1.8 photon/sec and for 3C 273 2.3 photons/sec. It may be seen from graphs of Fig. 8 that there would be a larger difference between the curves of Fig. 5 if NGC5548 would have the same apparent luminosity as 3C273.

There is another interesting difference in the structure of entropy variations. If the upper limit of the ampligram will be increased from standard 20% without changing the number of intervals of wavelet coefficient magnitude, the width of a single interval/bandwidth will be increased. A test has been done for NGC 5548 and 3C 273 to see the influence of the upper limit and of the bandwidth on the entropy variations. The curves of Fig. 9 correspond to the upper limit of the ampligram of 20% and correspondingly to a bandwidth of 1%. In the left graph of Fig. 9 the upper limit of the ampligram has been increased to 30% and thus the bandwidth to 1.5%. Respective numbers for the right graph of Fig. 9 are 40% and 2%. It may be seen that there is no significant change in entropy values between Fig. 5 and the left graph of Fig. 9. However, there is a significant change in the right graph. The change is not just a transformation of the horizontal scale due to increase of the upper limit. The decrease of entropy for NGC 5548 must be due to increased bandwidth. It means that the deterministic structures/building blocks in S1

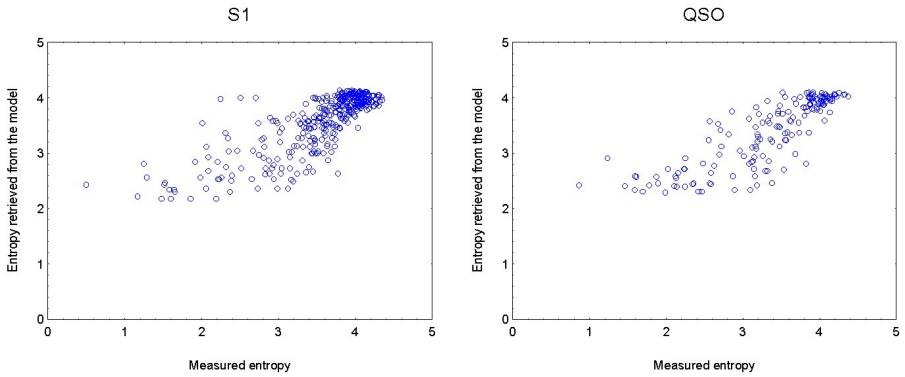

Fig. 7. Test of the entropy models for both groups of sources: S1 and QSO.

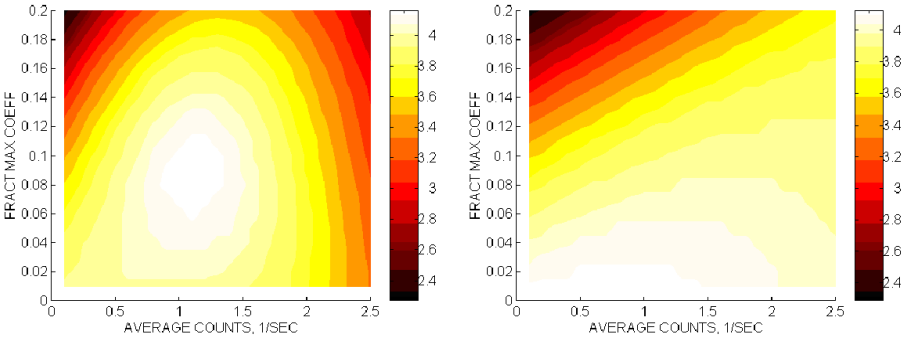

Fig. 8. Changes of entropy (gray scale) with the average counting rate (horizontal axis) and the wavelet coefficient magnitude (vertical axis) for both types of sources: the left graph for S1 and the right graph for QSO. The decreased entropy corresponds to dark areas of the graph.

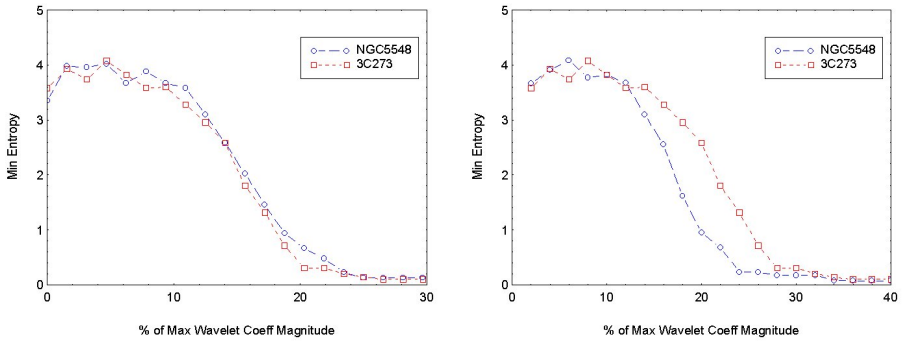

Fig. 9. A comparison of minimum entropy curves for NGC 5548 and 3C 273 for increased upper limit of the ampligram (30% in the left graph, 40% in the right) and increased bandwidth (1.5% in the left graph, 2% in the right).

has a finite width in the spectral density domain and that doubled bandwidth is needed to reveal their presence.

5. Summary and conclusions

The present study clearly proof that there are significant differences in the structure of the X-ray photon trains from different types of sources. A comparison of time scale spectra (Fig. 3), calculated for low apparent luminosity objects, shows twice as large variability for the QSO group as for the S1 group. Larger vertical extent of the average features in the spectra of S1 indicates more frequent occurrence - relative dominance of semi-regular deterministic components.

A study of ampligram amplitude distributions shows that there is another significant difference in the occurrence of deterministic structures in both types of sources. As a measure of information in the distributions the entropy was used. We have seen that there is a lower minimum entropy for lower level wavelet coefficient magnitudes in S1 variability spectra than in QSO spectra, indicating less random and more deterministic or regular contributions in S1 objects. This is also seen in different change in entropy with apparent source luminosity in the two categories (Fig. 10). These results indicate that by studying statistical properties of an X-ray photon train it is possible to identify unambiguously the character of its source.

These differences have a natural interpretation. The luminosity "building blocks" or events in Seyfert 1 galaxies, which are responsible for the intrinsic X-ray variability of these sources show more correlations and regularities than those in QSOs. This means that the accretion disk flares or the cluster black hole collisions with accretion disks around the larger holes ("ballistic events", see Pacholczyk & Stoeger 1994) - or whatever causes these discrete events - are themselves more strongly correlated in Seyfert 1 galaxies than in QSOs. If we assume that the black hole cluster scenario, for example, this would mean that subclustering in Seyfert 1 galaxies would be stronger than in QSOs. That might be either because there are more smaller and intermediate size black holes in Seyfert 1 nuclei (QSO nuclei have perhaps evolved much further towards a configuration

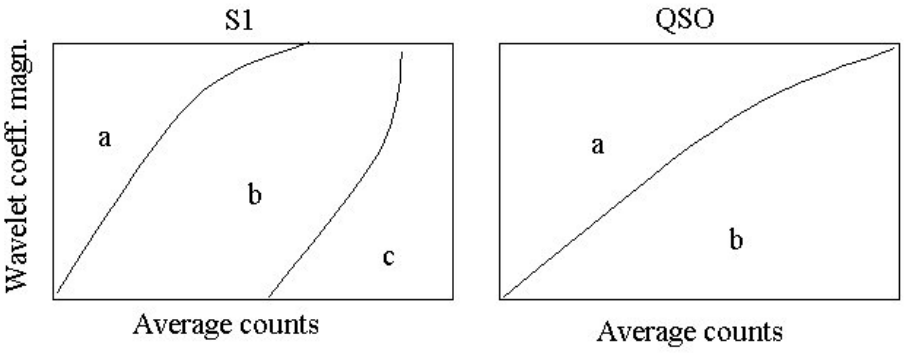

Fig. 10. Occurrence of different distribution types (see Fig. 6) for both groups of sources and for different apparent luminosities.

of just one or a few supermassive black holes surrounded with relatively few smaller satellite black holes, whereas Seyfert 1 nuclei are possibly at a less-evolved black hole nuclear cluster stage, cf. Pacholczyk & Stoeger (1994), and/or because the cluster in QSOs exhibit less tendency towards subclustering for some other reason (they may be more massive and more compact, making subclustering less pronounced or less possible). If subclustering of black holes are common in Seyfert 1 nuclei, their ballistic encounters with accretion disks around other black holes, or with intracluster clouds will be correlated. A spherical subcluster penetrating an accretion disk will exhibit a definite signature - a swarm of little black holes collisions with the disk causing a swarm of small luminosity building blocks followed by fewer larger events from the collisions of the larger black holes in the subcluster, and then finally another swarm of smaller building blocks from the smaller black holes on the back side of the subcluster. More than likely, however, most subclusters will have been tidally deformed into elongated trains of black holes, due to close encounters with larger black holes in the cluster, or with other subclusters. As the black holes in these subcluster trains hit accretion disks in correlated, fairly rapid succession they will generate a detectable series of quasi-periodic series of luminosity events, much like a fragmented asteroid or a comet hitting a planet.

Acknowledgements. The software used in this work has been developed by Pär-Ola Nilsson, Jan Karlsson and Fredrik Rutqvist of the Umeå Division of the Swedish Institute of Space Physics. The data used in the present analysis are ROSAT PSPC photon event history files obtained through the High Energy Astrophysics Science Archive Research Center Online Service, provided by the NASA/Goddard Space Flight Center.

References

Arp H., 1997, A&A 327, 479
Liszka L., Holmström M., 1999, A&AS 140, 125
Mack K.H., Kerp J., Klein U., 1997, A&A 324, 870
Nandra K., George I.M., Mushotzky R.F., Turner T.J., Yaqoob T., 1997, ApJ 476, 70
Nowak M.A., Wagoner R.V., Begelman M.C., Lehr D.E., 1997, ApJ 477, L91
Pacholczyk A.G., Stoeger W.R., 1994, ApJ 434, 435
Sambruna R.M., 1997, ApJ 487, 536
Schartel N., Schmidt M., Fink H.H., Hasinger G., Truemper J., 1997, A&A 320, 696
Shannon C.E., Weaver W., 1949, The Mathematical Theory of Communication. University of Illinois Press