A Detection Method for X-Ray Images based on Wavelet Transforms: the Case of the ROSAT PSPC

F. Damiani, A. Maggio, G. Micela, S. Sciortino
Istituto ed Osservatorio Astronomico di Palermo 'G.S.Vaiana', ITALY

ABSTRACT

We have developed a method based on wavelet transforms (WT), evaluated at a set of several scales a, to efficiently detect both point- and extended sources. This method works even when the PSF width varies strongly across the field of view and there are strong gradients in the exposure map, such as for the ROSAT PSPC.

The WT computed at a scale a is mostly sensitive to structures of that size, allowing to estimate source extent, while the height of a WT maximum measures source counts.

The WT probability distribution of the background has been computed through extensive simulations, to set significance thresholds for candidate sources (i.e. local WT maxima). Simulations of both pure background and background+sources realistic PSPC images allow to assess the frequency of spurious detections as a function of detection threshold, and the algorithm sensitivity, respectively.

The use of exposure maps in the source detection process allows us to retain good detection efficiency even near detector ribs and edges. The algorithm may be used also to consistently get upper limits to X-ray count rate of point-like objects, enabling the construction of accurate sensitivity maps.

Applications to real PSPC images of galaxies and stellar clusters show the algorithm good performance even in cases of extended sources and crowded fields.

1 Outline

Wavelet Transforms (WT) are a tool to identify features of interest in a data set. The WT of an image consists of a scale-position representation of the data, which enables to study simultaneously the shape (size) and location of imaged objects; by contrast, classical Fourier analysis gives only shape information, while peak-finding methods (such as sliding-window detect) only yield positions. Similar X-ray source detection techniques have been presented by Rosati et al. (1993), Grebenev et al. (1995) and Vikhlinin et al. (1994).

The multiscale approach of wavelet analysis is very useful to analyze images containing sources of various sizes (both intrinsically, or broadened by the PSF), as is the case for the PSPC: it enables to take full advantage of the good spatial resolution near image center, retaining at the same time efficiency to detect off-axis sources and truly extended sources.

The WT of an image is a function $w(x, y, a)$ depending on both position (x, y) and a scale parameter a, and is defined as:

$$w(x,y,a) = \iint g\left(\frac{x-x_0}{a}, \frac{y-y_0}{a}\right) f(x_0,y_0)\, dx_0 dy_0 \tag{1}$$

where $g\left(\frac{x}{a}, \frac{y}{a}\right)$ is the 'generating wavelet', for which we choose the form:

$$g\left(\frac{r}{a}\right) = \left(2 - \frac{r^2}{a^2}\right) e^{-\frac{r^2}{2a^2}} \qquad (r^2 = x^2 + y^2). \tag{2}$$

Important properties of the WT:

- For $f(x,y) = const.$ or for $\nabla f(x,y) = const.$, $\Longrightarrow w(x,y,a) = 0$.
- $w(x, y, a)$ depends on values of f only within 4-5a from the point (x, y).
- The WT at a scale a brings into evidence image structures of size $\sim a$.
- WT analysis allows the deblending of strong enough sources.

Our method requires the following problems to be solved:

1) Determination of the statistical properties of the WT of background and derivation of detection confidence levels (see section 2).

2) Determination of local, scale-dependent background (see section 3).

3) Putative source recognition in the WT images and assessment of source existence based on local background (see section 3).

4) Determination of source size and intensity in the WT space (see section 3).

The whole method is presented with substantial more detail by Damiani et al. (1995).

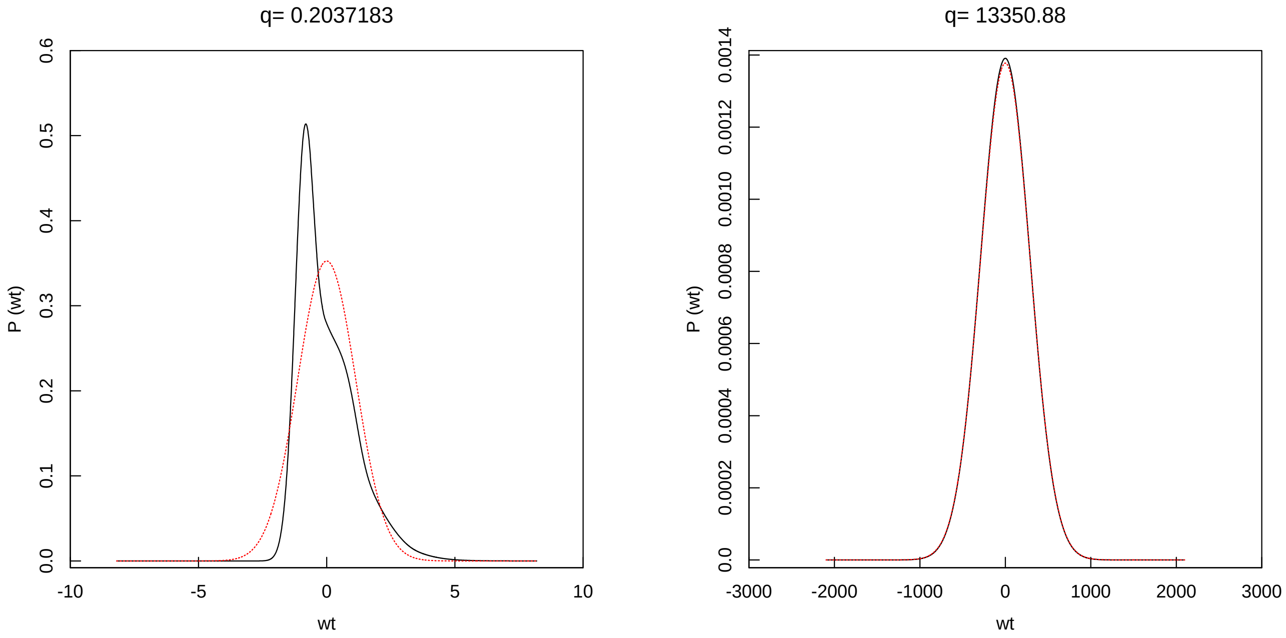

Figure 1: (a, left) Probability distribution $P_q(w)$ for the WT of pure background, for $q \equiv ba^2 = 0.2$; the red curve is a gaussian with zero mean and variance $\sigma_w^2 = 2\pi q$. (b, right) Same as (a), but for $q = 1.3 \cdot 10^4$; for this q, $P_q(w)$ is nearly gaussian, with the predicted variance.

2 Statistical Properties of the Wavelet Transform

The general properties of the WT ensure that $w(x, y, a) = 0$ for a flat background. However, background noise induces fluctuations of the WT value around zero. To discriminate reliably true sources from background fluctuations, it is therefore essential to know the probability distribution of $w(x, y, a)$ generated by background noise.

Such probability distribution has been derived by us using both numerical and analytical techniques. It can be shown that the probability density $P(w)$ of obtaining a value of WT amplitude w is a function of both background density b and scale a of the transforms, but only through the combination $q \equiv b \times a^2$, namely the (adimensional) number of photons per square scale.

It may also be shown that two limits exist for the distribution $P_q(w)$, corresponding to low (Poisson) and high (gaussian) photon statistics, respectively. Two examples of $P_q(w)$ for $q = 0.2$ and $q = 1.3 \cdot 10^4$, computed numerically using FFT techniques, are shown in Figure 1: when the photon statistics is high (panel b), $P_q(w)$ tends to a gaussian distribution with zero mean and width $\sigma_w = \sqrt{2\pi q}$ (as can be shown analytically), while $P_q(w)$ becomes strongly non-gaussian for lower photon statistics (panel a).

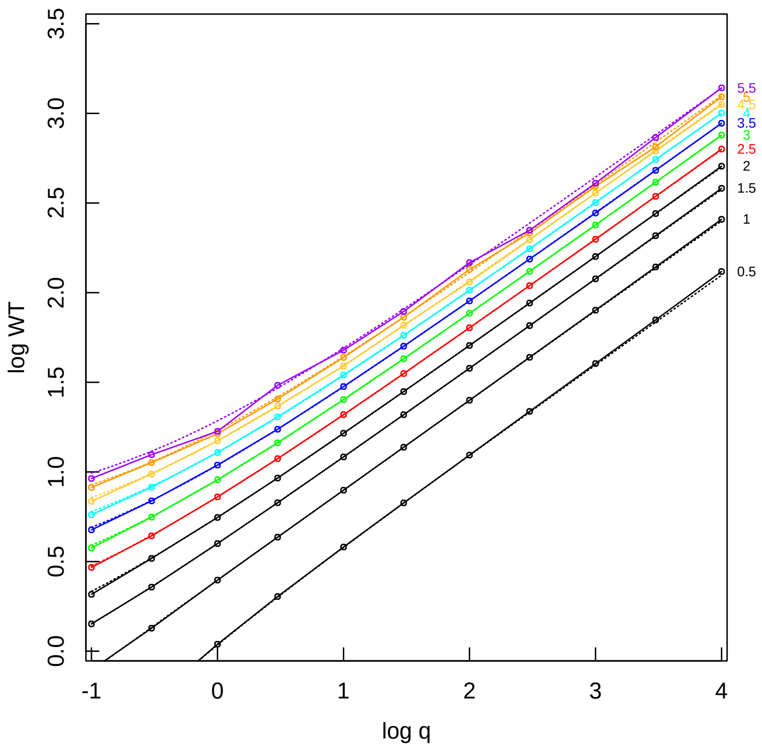

Figure 2: Quantiles of distributions $P_q(w)$ vs. q. We show the quantiles corresponding to probability levels of 0.5σ through 5.5σ, as labeled.

The distributions $P_q(w)$ computed using FFT are in agreement with those computed from numerical simulations. From these we have computed quantiles of $P_q(w)$ corresponding to extreme background fluctuations (namely cumulative probability $P = 0.9999998$), to derive thresholds for source detection above background, at a desired confidence level.

The quantiles of the distributions $P_q(w)$ may be well approximated with the (empirical) formula (see Figure 2):

$$w_0(k, q) = k\sqrt{2\pi q} + (c_1 + c_2 k + c_3 k^2), \tag{3}$$

where k is the significance level equivalent to number of gaussian σ's (e.g. $k = 5$ means a 5σ threshold).

3 Implementation of the Algorithm

- **Background Map:** The reference background used to predict the expected WT fluctuations in the absence of sources is computed by smoothing the image using a gaussian filter. In each point, the width of the gaussian has been taken equal to the local PSF width σ_{PSF}: this ensures that the resulting map is a faithful representation of the image background, describing *real* background modulations on any length scale, and least affected by statistical errors. In order to produce a ribs-free map, we interpolate the background over ribs.

  Since this map includes the effect of sources, a further step is to apply a local median filter that minimizes their effect on the final reference background. The region used to compute medians has a size proportional to the WT scale a, in order to have a fully scale-invariant approach.

- **WT Computation:** The WT of the image is computed at each scale a (by direct integration), over a region of radius such that $a \geq \sigma_{PSF}$ (since no information is present in the data on scales $a \ll \sigma_{PSF}$), and with a grid spacing proportional to a, not to oversample the WT image. In order to avoid the generation of spurious WT maxima near PSPC ribs we compute the WT of the *count-rate* image, not of the *photon* image, dividing the photon image by the observation *exposure map*, pixel by pixel. Of course, this operation alters the original (Poisson) photon statistics per bin, an effect that must be accounted for (see section 3.2).

- **Source Detection:** Candidate sources are identified as local *spatial maxima* of the WT at any scale. A candidate source is retained as a real detection if its WT amplitude w_{peak} is higher than that expected for the WT of the local background (at that scale a), to an assigned probability level.

- **Source Count-Rate and Size Determination:** Sources detected at various scales are cross-identified, to yield the profile of WT amplitude as a function of scale a, for each source. This is used to derive the *source rate* and *size*. At the scale of maximum significance, the WT is recomputed in a small region around the peak, using a finer spatial grid, to best estimate the source position and count rate.

- **Background Map Update:** After having obtained a first list of sources, a second iteration consists in the elimination of all detected point sources from the computation of the reference background map; this step allows us to detect even weak sources close to much stronger ones, which spuriously raise the local background estimate. Having updated the background map, all previous steps are repeated to build a final source list.

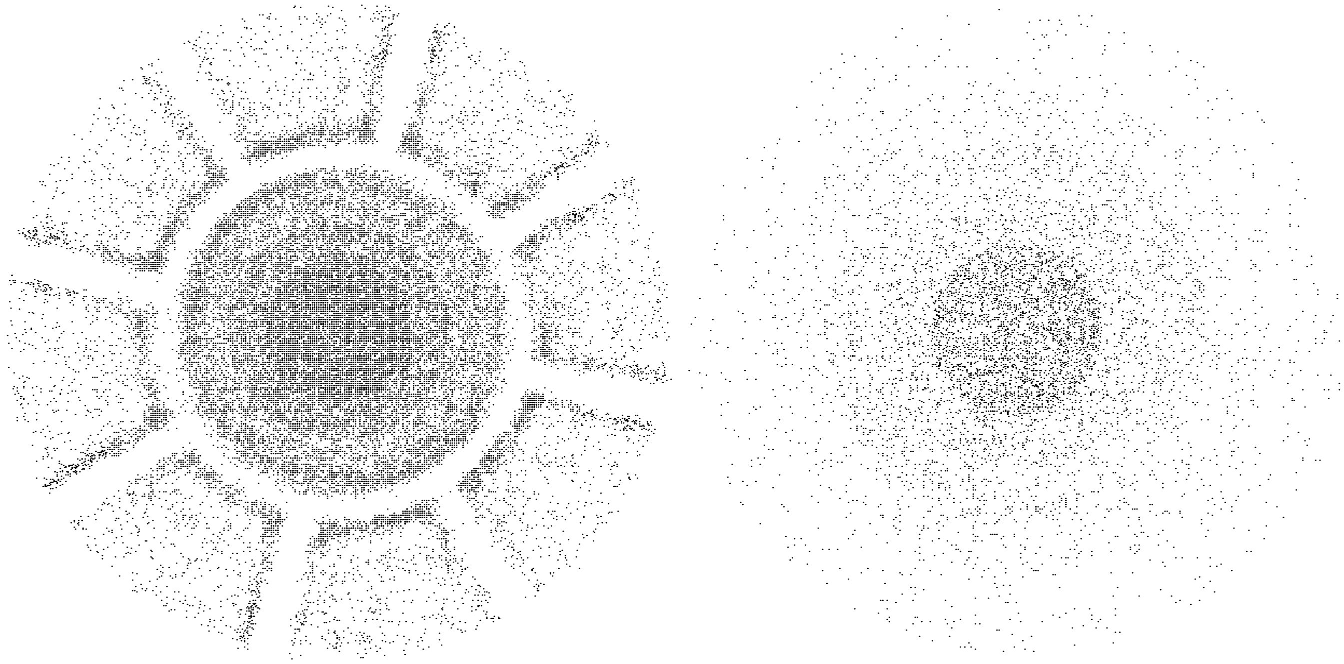

Figure 3: (a, left) Spatial distribution of spurious sources detected above 3σ level in a set of pure background simulations, using the WT of the photon image. (b, right) Same as a, but using the WT of the count-rate image.

3.1 Wavelet Transform Computation

We apply a discrete WT to binned PSPC images (512×512, pixel size=15"). The WT is computed (for a set of scales matching the range of sizes of interesting objects, in steps of $\sqrt{2}$) by convolving the generating wavelet g with the image using direct, elementary integration, rather than FFTs. This largely speeds up calculations, with a modest loss of accuracy compared e.g. to statistical errors.

As mentioned, the background discontinuities produced by ribs obscuration and image edge cause the detection of spurious sources (WT peaks) if the WT is made of the original photon image. For this case, Figure 3a shows the spatial distribution of spurious sources over the field of view in a set of pure-background simulated PSPC images (see section 5): excess sources are evident along ribs and edges. The discontinuity due to ribs is eliminated using rate images; a suitable correction is also applied to the image edge.

To account for the changes of photon statistics in the rate image, we still compute the background map on the photon image, yielding thresholds for the *photon* WT; these thresholds are converted to thresholds for the *rate* WT using a suitable 'effective' exposure time t_{eff}, function of position, that relates the *fluctuations* of photon- and rate WT, respectively; t_{eff} may be derived analytically under certain approximations.

As a result, we obtain the distribution of spurious sources plotted in Figure 3b.

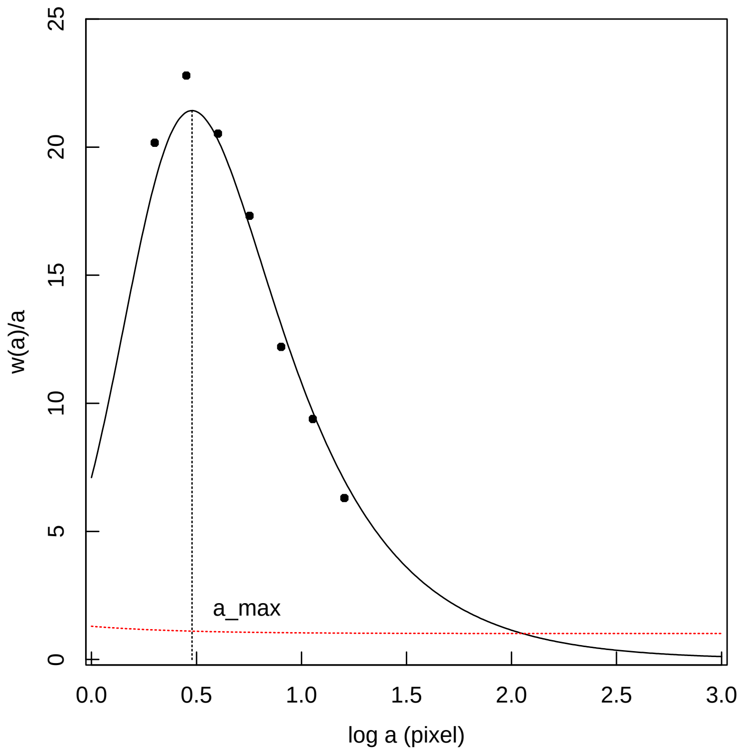

Figure 4: Plot of the function $w_{peak}(a)/a$ vs. a, for a gaussian source with size $\sigma = 1.74$ pixel, and count rate $R_{src} = 57.31$ cts/ksec; the scale a_{max} is indicated. Solid dots are measured values in a simulation with exposure time of 10.0 ksec. Also shown (low dotted line) is the 4σ detection threshold level for a simulated background of 16.32 cts/arcmin^2.

3.2 Source Count-Rate and Size Determination

If $f(x,y)$ is a normalized gaussian (source) of width σ, its WT is

$$w(r,a) = \frac{1}{\left(1+\frac{\sigma^2}{a^2}\right)^2}\left(2 - \frac{r^2}{a^2+\sigma^2}\right) e^{-\frac{r^2}{2(a^2+\sigma^2)}}, \tag{4}$$

having a strong peak for $r = 0$. For a source count rate R_{src}, the WT amplitude $w(r,a)$ at the spatial maximum $r = 0$ is

$$w_{peak}(a) = \frac{2R_{src}}{\left(1+\frac{\sigma^2}{a^2}\right)^2}. \tag{5}$$

The function $w_{peak}(a)/a$ has a maximum for $a_{max} = \sqrt{3}\sigma$ (Figure 4), which allows to estimate the source size σ. Moreover, the amplitude $w_{peak}(a = a_{max}) = (9/8)R_{src}$, from which the source rate is directly computed.

A more accurate estimate of σ, R_{src} may be obtained, in principle, by a fit of the expected $w_{peak}(a)/a$ profile to that derived from the data, rather than considering only its maximum value. However, the data to be fit are not statistically independent, affecting the determination of goodness-of-fit. A better approach to refine the estimates of σ and R_{src} is to consider $w_{peak}(a)$ at two scales close to a_{max}, that identify uniquely σ and R_{src} with no fitting; the errors on σ and R_{src} are then evaluated correctly by studying the covariance of the two values of $w_{peak}(a)$ used.

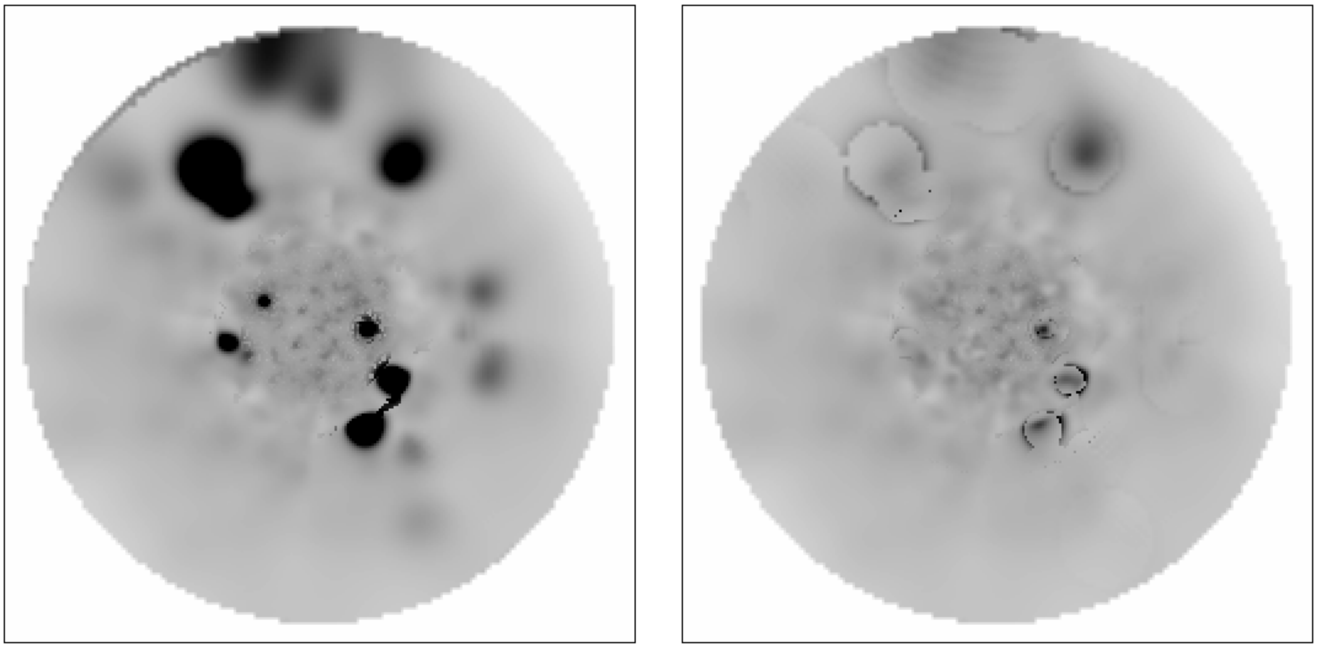

Figure 5: (a, left) Initial background map relative to the Hyades image RP200020; sources are clearly seen in this image, and will be (partially) screened using a median filter to yield the reference background for source detection in the first pass. (b, right) Background map for the same image as (a), but excluding from the computation sources detected in the first pass. Only the extreme wings of the strongest sources are still visible in this image, whose effect will be reduced further by the median filter.

3.3 Background Map Update

It sometimes happens that a real point source close to a much stronger one remains undetected, although it produces a clear peak both in the original image and in the WT image; this is because the effect of sources on the reference background is not completely screened out by the median filter application, yielding a spuriously high detection threshold.

Therefore, after having detected all but the weakest sources in the first pass, we recompute the reference background map interpolating over the positions of detected sources, allowing the detection of weaker ones undetected in the first pass.

Figure $5a, b$ shows the two background maps computed in the first and second pass, respectively.

4 Upper Limits

The upper limit for an undetected source is computed in the same way as the rate R_{src} for detection. A point source (assumed gaussian) would have a known width $\sigma = \sigma_{PSF}$: this fixes the shape of the $w_{peak}(a)$ profile, apart from the normalization R_{src}, which is the sought count-rate upper limit. Once the local background is known, the detection threshold $w_0(k, q)$ at a given level k is simply a function of scale a. Then, the source would be detected if the $w_{peak}(a)$ profile intersects the

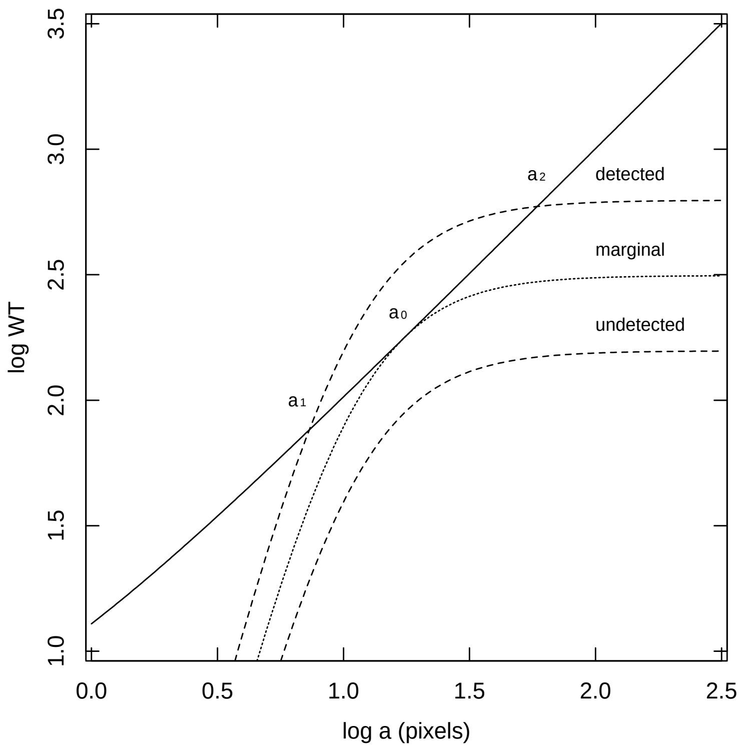

Figure 6: Plot of the $w_{peak}(a)$ profile (dashed) for three values of source intensity R_{src}, and a source size $\sigma = 10$ pixels. These curves are compared with the threshold curve relative to a background of 1.0 cts/pixel^2: the upper curve refers to a source detected for scales $a_1 \leq a \leq a_2$ where it lies above threshold, the lower one refers to an undetected source, and that in the middle to a source marginally detected at $a = a_0$, yielding the desired upper limit to R_{src} in the given position.

threshold curve w_0, and undetected if remains below it, depending on the value of R_{src} (Figure 6). The upper limit to R_{src} is therefore the value that makes the two curves to be tangent to each other.

5 Testing on Simulated PSPC Images

To study in detail the performances of the algorithm we have applied it to sets of simulated PSPC images, containing both pure background, and background plus sources. To predict most closely the true performances on real data, simulations were made as realistic as possible, including the presence of ribs and wobbling by using exposure maps of actual observations. The distribution of source intensities was generated following the $\log N - \log S$ relation of Hasinger et al. (1993a). The source PSF was modeled after Hasinger et al. (1993b). These simulations have been repeated for various exposure times. For illustration, a typical simulated image with sources is shown in Figure 7.

The results of pure-background simulations allow to study the spatial distribution of spurious sources (Figure 3*b*) and their distribution as a function of significance level (Figure 8*a*). As this Figure shows, a detection threshold of $\sim 4.2\sigma$ is expected to yield only one spurious source per image (over the entire field of view).

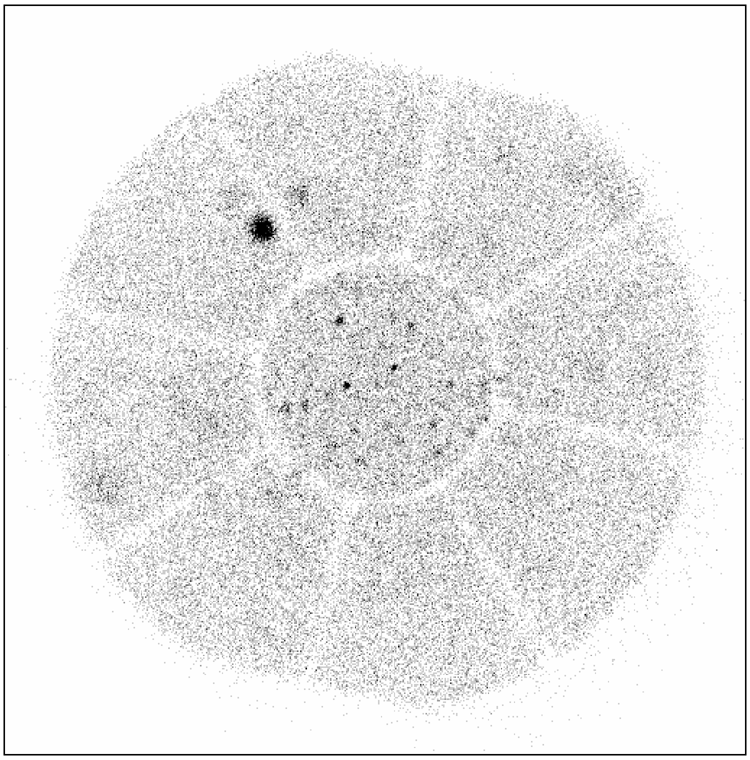

Figure 7: Simulated PSPC image with point sources. The exposure map used in the simulation is that of the sequence RP200068, but scaled to an exposure time of 10.0 ksec; we adopted a background rate of $1.5 \cdot 10^{-3}$ cts/sec/arcmin^2 at image center.

The simulations with sources allow instead to study the source detection efficiency, and the accuracy of the output quantities (source position, counts and size). Figure 8b shows the percentage of sources that are detected by the algorithm out of all put in the simulated images, as a function of input source counts and offaxis angle. Finally, Figures 9a, b show the agreement between input and detected source counts and sizes, respectively.

REFERENCES

Grebenev,S.A., Forman,W., Jones,C., Murray,S. 1995, ApJ, 445, 607.

Hasinger,G., Burg,R., Giacconi,R., Hartner,G., Schmidt,M., Trümper,J., Zamorani,G. 1993a, A&Ap, 275, 1.

Hasinger,G., Boese,G., Predehl,P., Turner,T.J., Yusaf,R., George,I.M., Rohrbach,G. 1993b, MPE/OGIP Calibration Memo CAL/ROS/93-015.

Rosati,P., Burg,R., Giacconi,R. 1994, in *The Soft X-ray Cosmos*, eds. E.M.Schlegel and R.Petre (AIP Conf. Proc. 313), p.260.

Vikhlinin,A., Forman,W., Jones,C., Murray,S. 1995, ApJ, 451, 542.

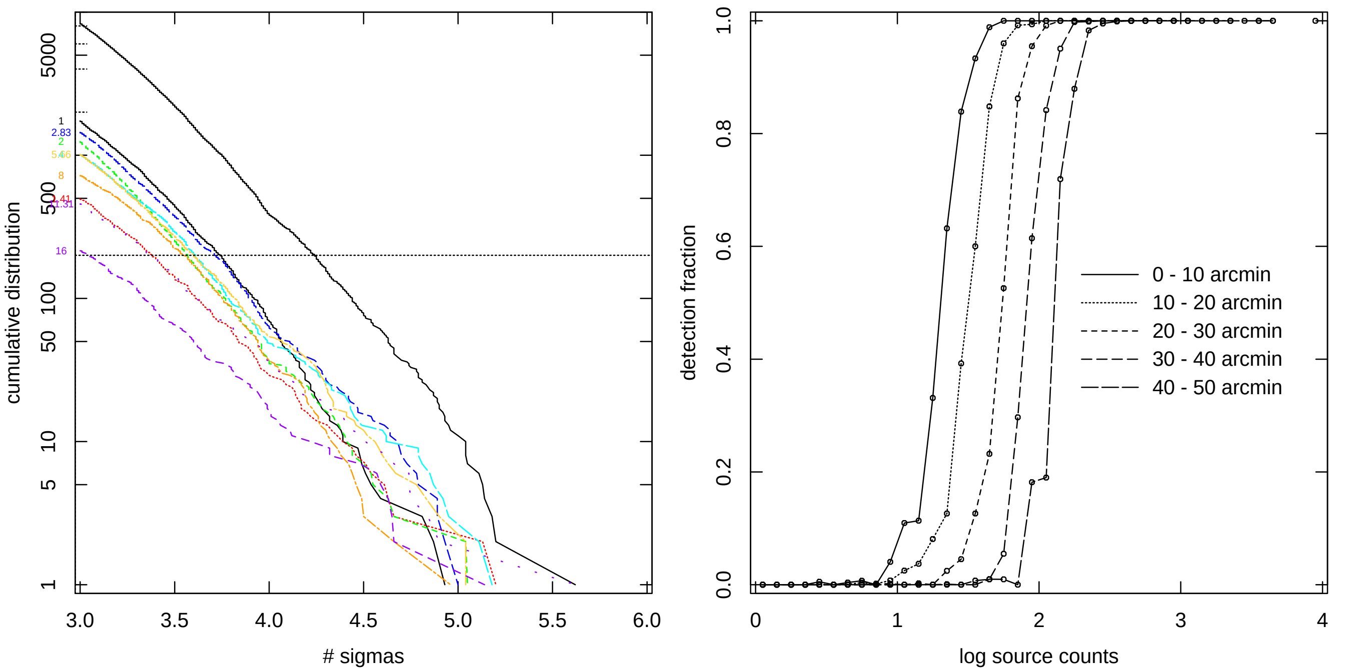

Figure 8: (a, left) Distribution of spurious sources detected in simulated PSPC background images. The upper solid line refers to all sources, regardless of detection scale. The distributions for the individual scales (as labeled) are shown by the lower curves. The horizontal line sets the limit of one spurious source per image. (b, right) Efficiency of real source detection, for various offaxis angles, for a set of simulations with exposure time of 10.0 ksec, and background of 15 cts/arcmin^2.

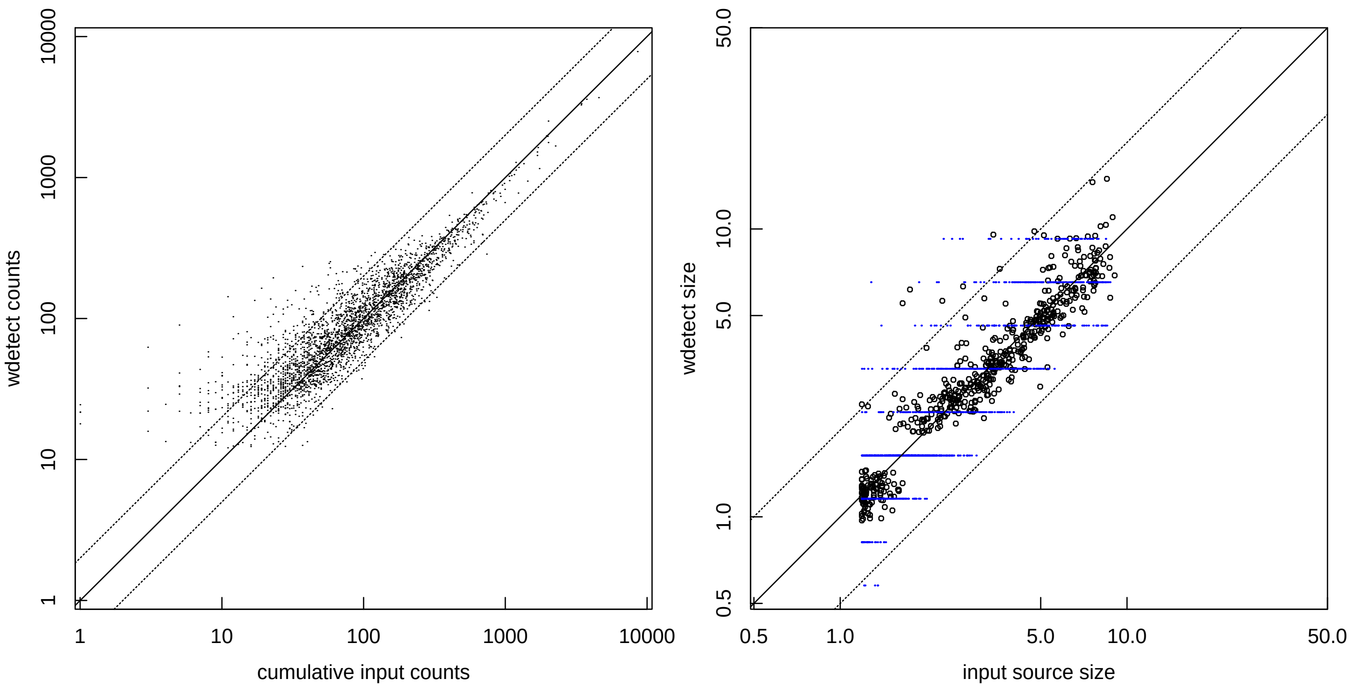

Figure 9: (a, left) Source counts as given by the algorithm, versus counts of the corresponding input simulated sources. Counts of all possible positional counterparts of a given detection are summed. The straight lines indicate identity, and factor of two differences. (b, right) Source sizes as given by the algorithm, versus input ones. Small dots indicate rough size estimates for weak sources, while circles are two-point size estimates for stronger sources. The straight lines indicate identity, and factor of two differences.

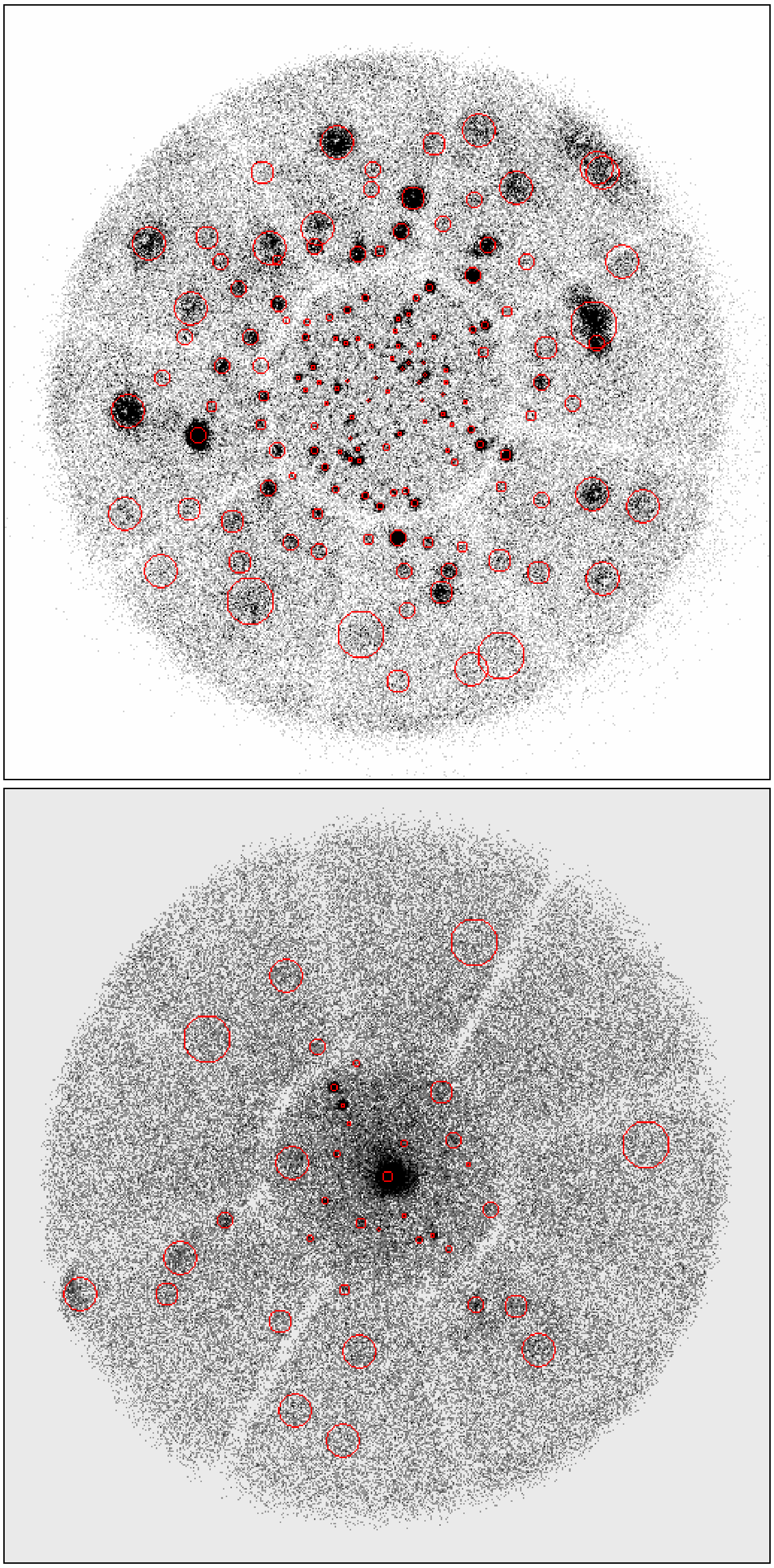

Figure 10: **Examples: Application to Real PSPC Data:** (Upper) The Pleiades open cluster (RP200068, livetime= 40112 sec). Circles are sources detected above 4σ, and their radius is equal to the scale of detection with the highest significance (highest value of $w_{peak}(a)/a$). In this image we detect 154 sources. (Lower) The galaxy NGC4636 (RP600016, livetime= 12735 sec). 41 sources are detected.